

Stages 2024

Jalon 30 avril 2024

Tristan Hillairet

Sommaire

- ▶ Avancée du stage
- ▶ Présentation de SRA
- ▶ Exploration de GeoSRA : évaluation du géoréférencement
- ▶ Exploration de GeoSRA : évaluation de organism

Avancée du stage

Avancée du stage

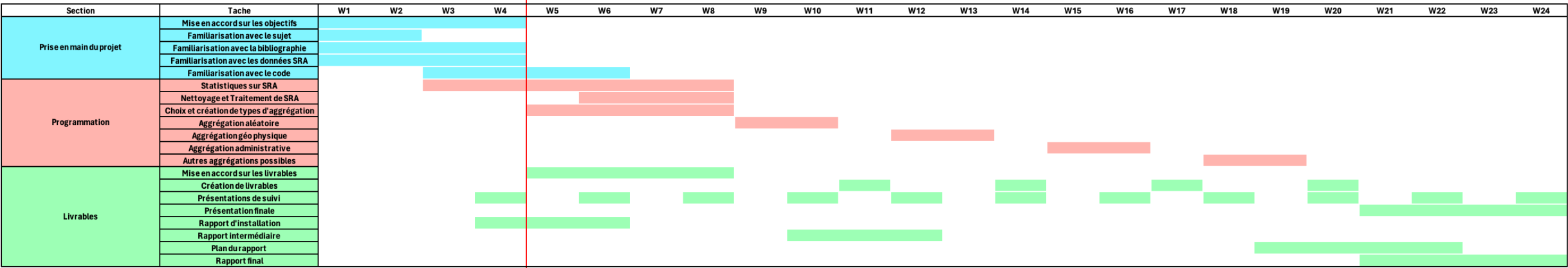


Diagramme de Gantt au 30/04/2024

Présentation de SRA

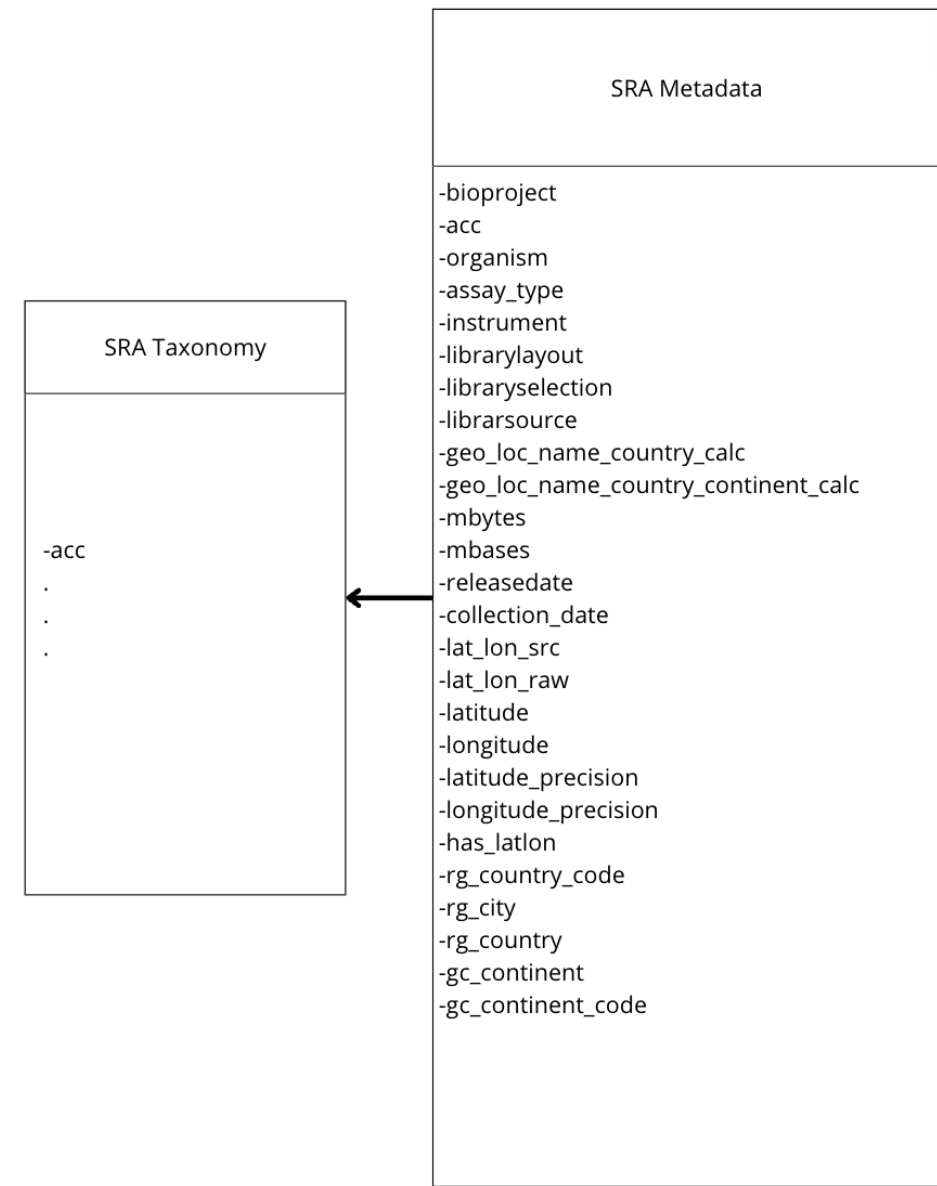
Présentation de SRA

- Combinaison des données de séquençage de 3 organisations (NCBI, EBI, DDBJ)
- Dépôt accessible au public. Données provenant de tous les domaines de la vie ainsi que des études métagénomiques et environnementales.
- Accessibles via un cloud ou par téléchargement.
- Une table des données taxonomique et une table associée des métadonnées.



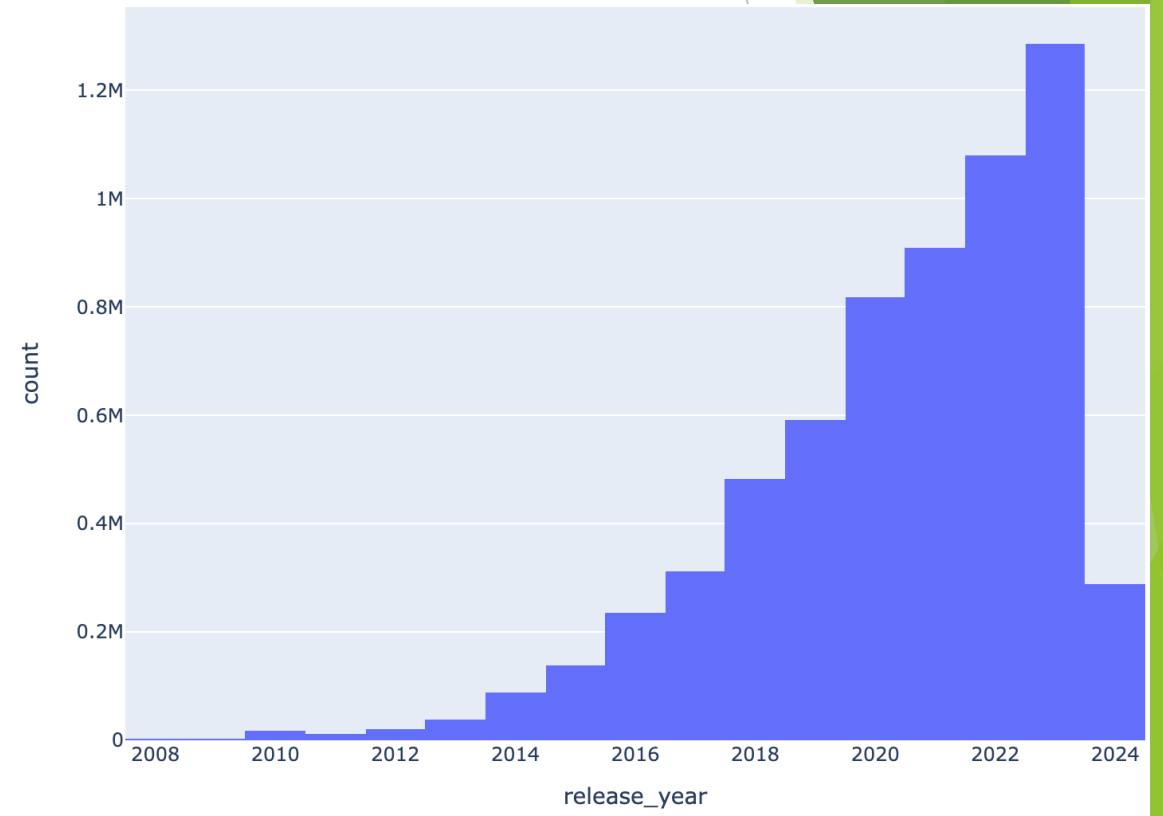
Présentation de SRA

- Différents types de métadonnées (type d'étude, temporelles, spatiales, autres...)
- Champs de métadonnées ambigus
- Simplification en agrégeant les champs
- Simplification en séparant les métadonnées en plusieurs tables.



Présentation de SRA

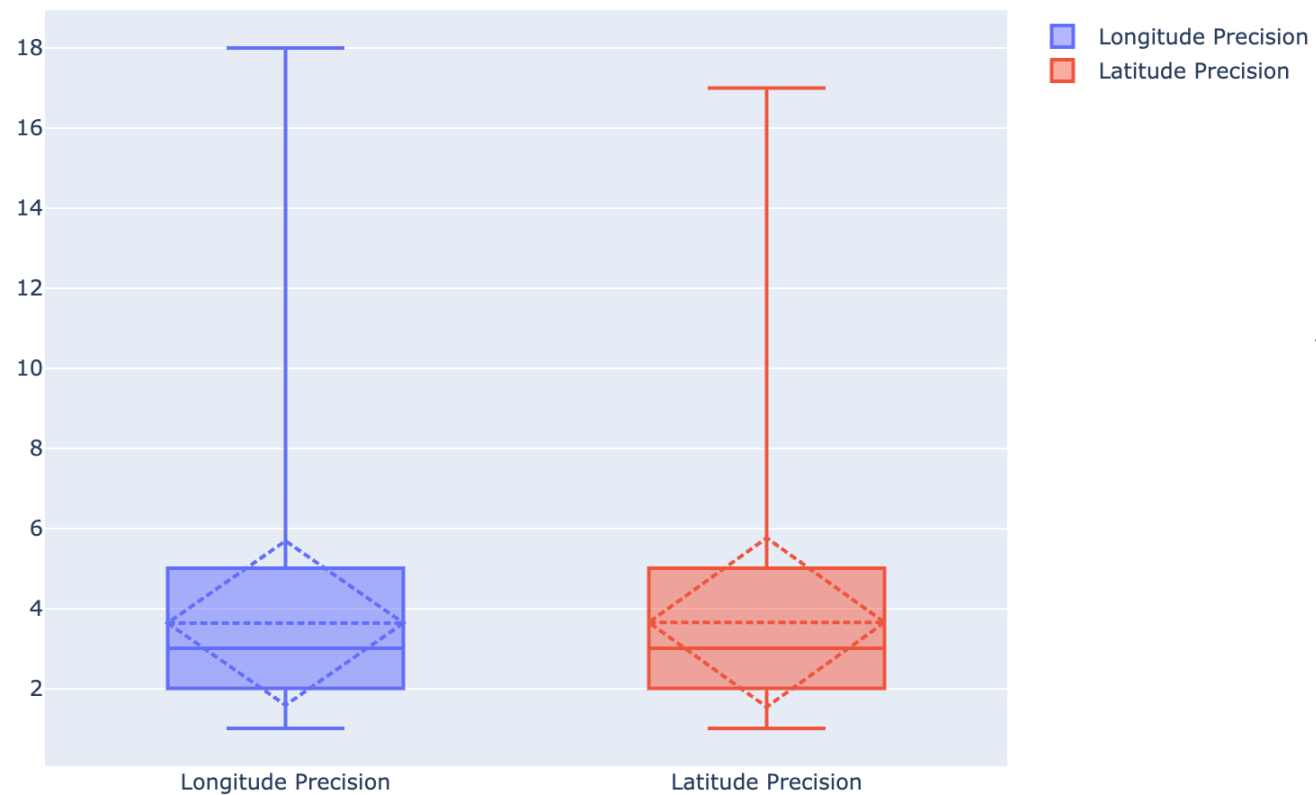
- Calcul d'un champ release_year
- Premières acquisitions en 2008.
- Augmentation exponentielle du nombre d'acquisitions.
- 6,3 M d'acquisitions renseignées dans le répertoire début avril 2024.



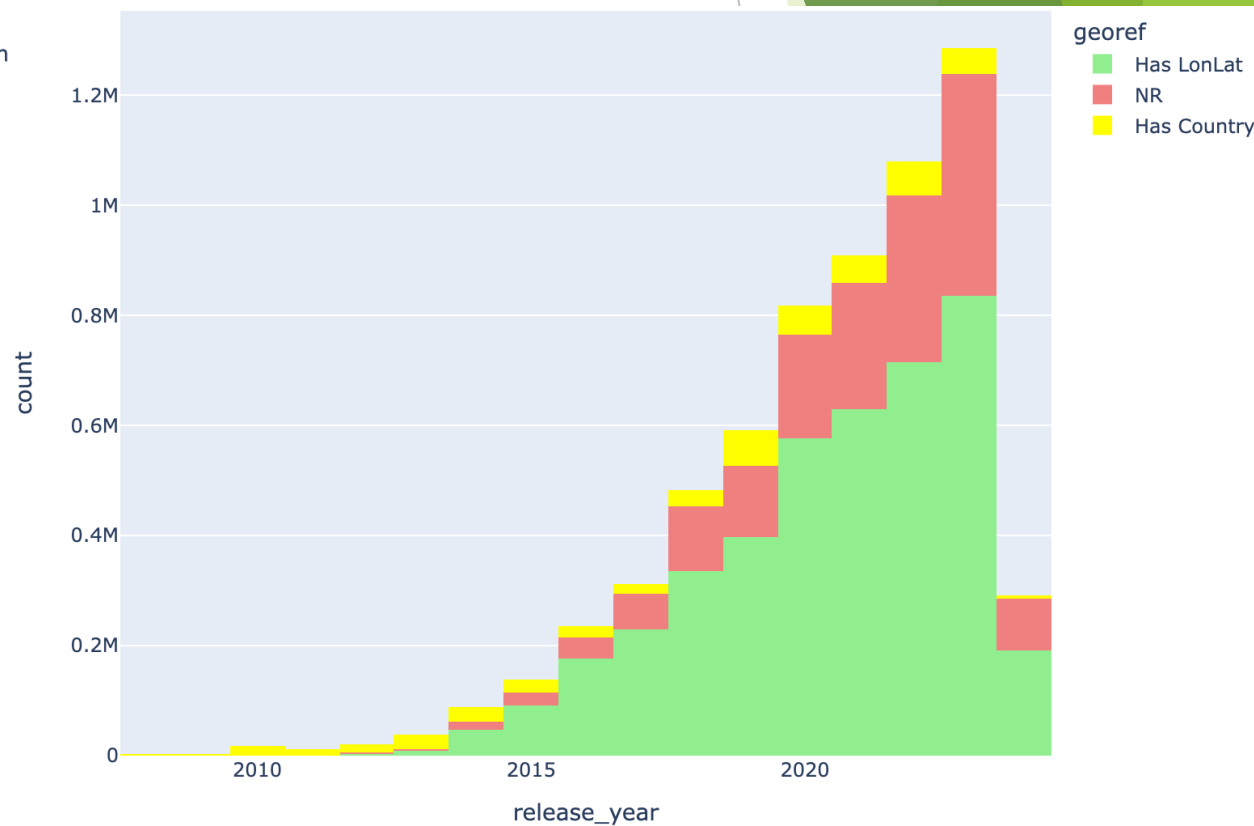
Nombre d'acquisition renseignées par années

Exploration de GeoSRA : évaluation du géoréférencement

Evaluation du géoréférencement



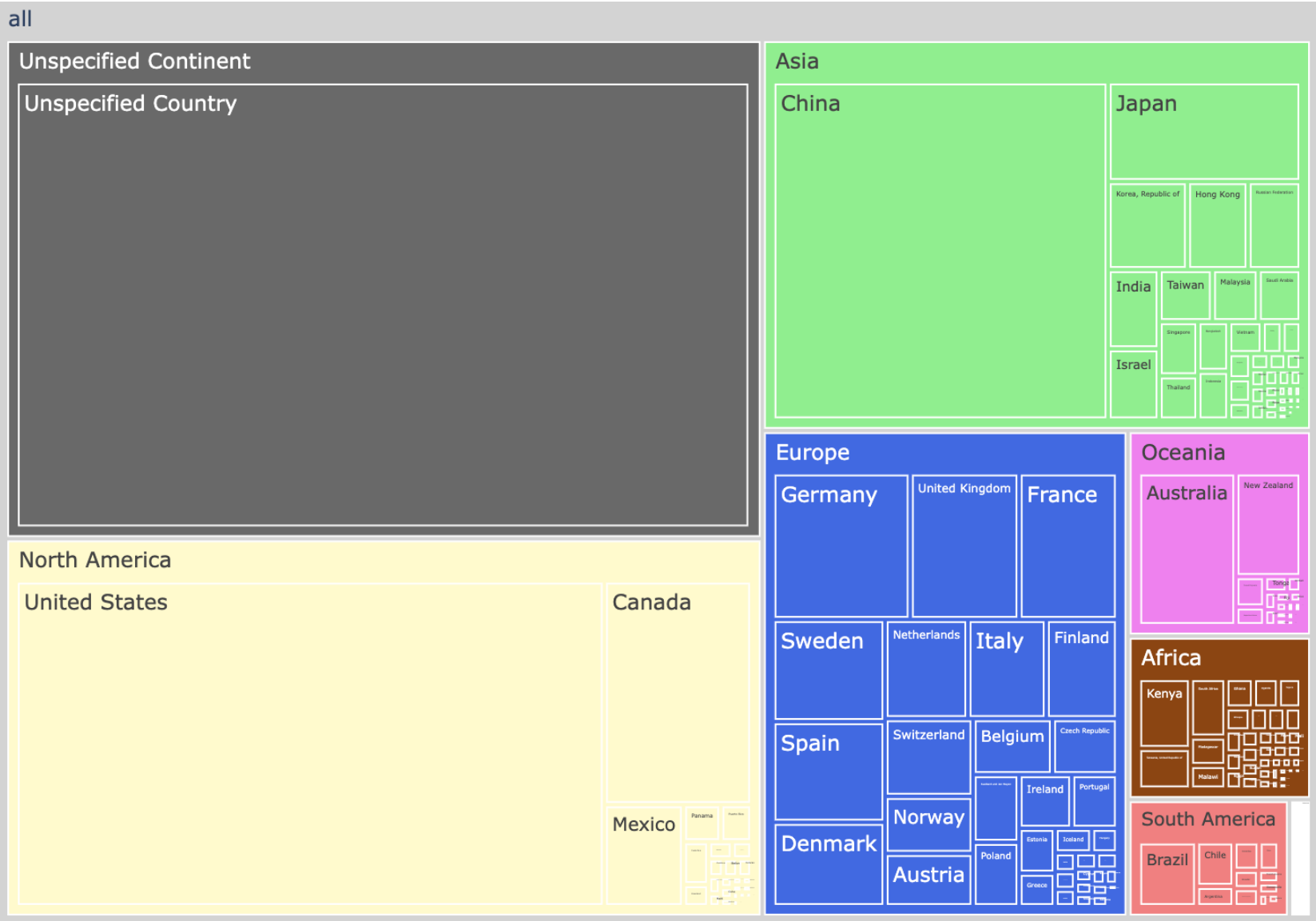
Précision des coordonnées en nb de décimales



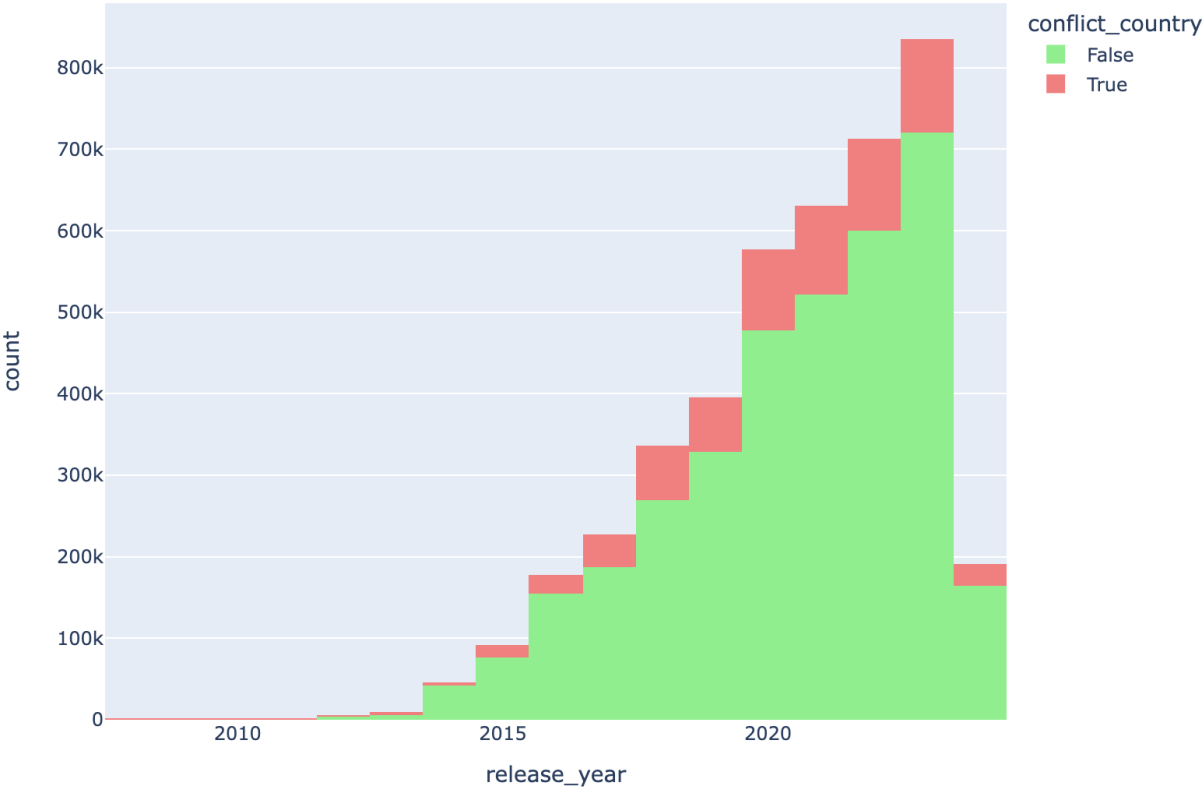
Types de géoréférencements par années

Evaluation du géoréférencement

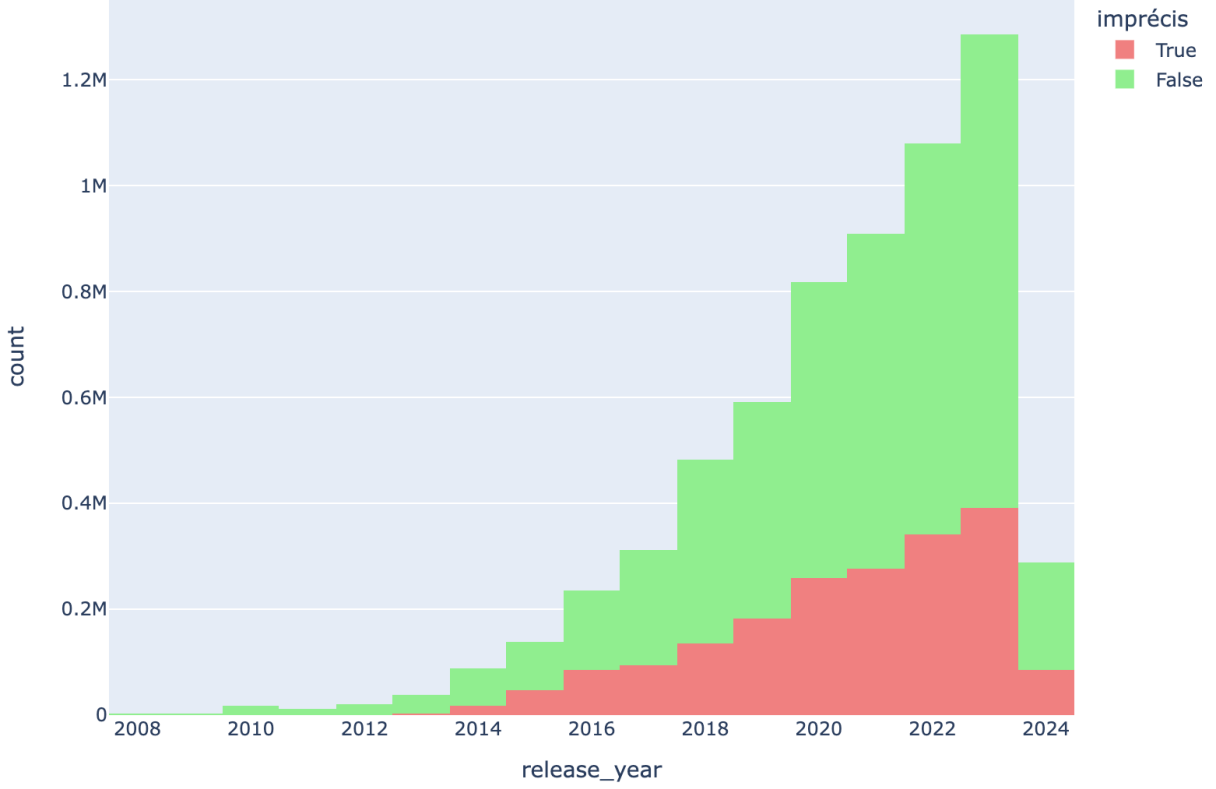
TreeMap du nombre d'acquisitions
par pays



Evaluation du géoréférencement



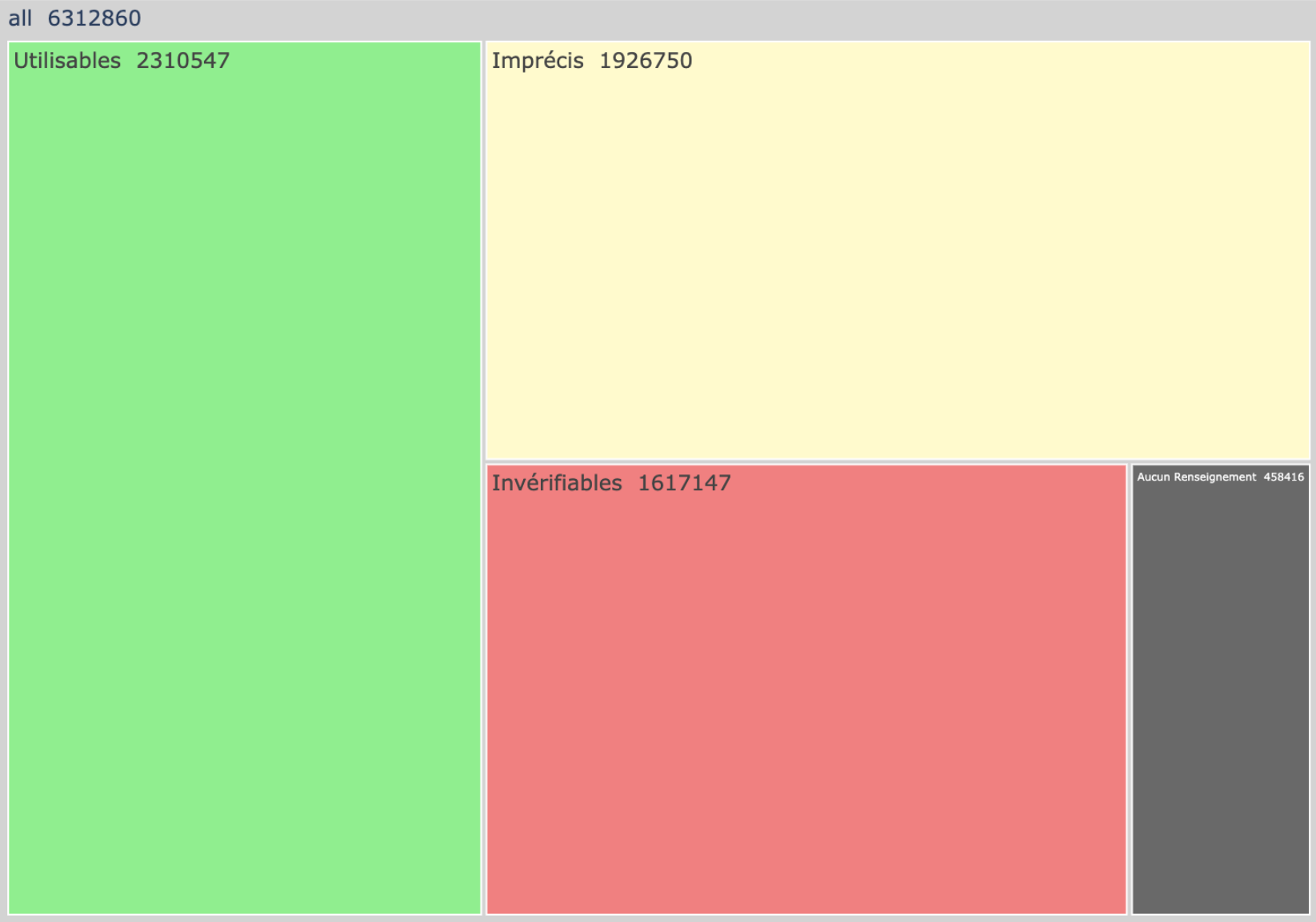
Conflits de pays renseignés par années



Nombre de coordonnées imprécises par années

Evaluation du géoréférencement

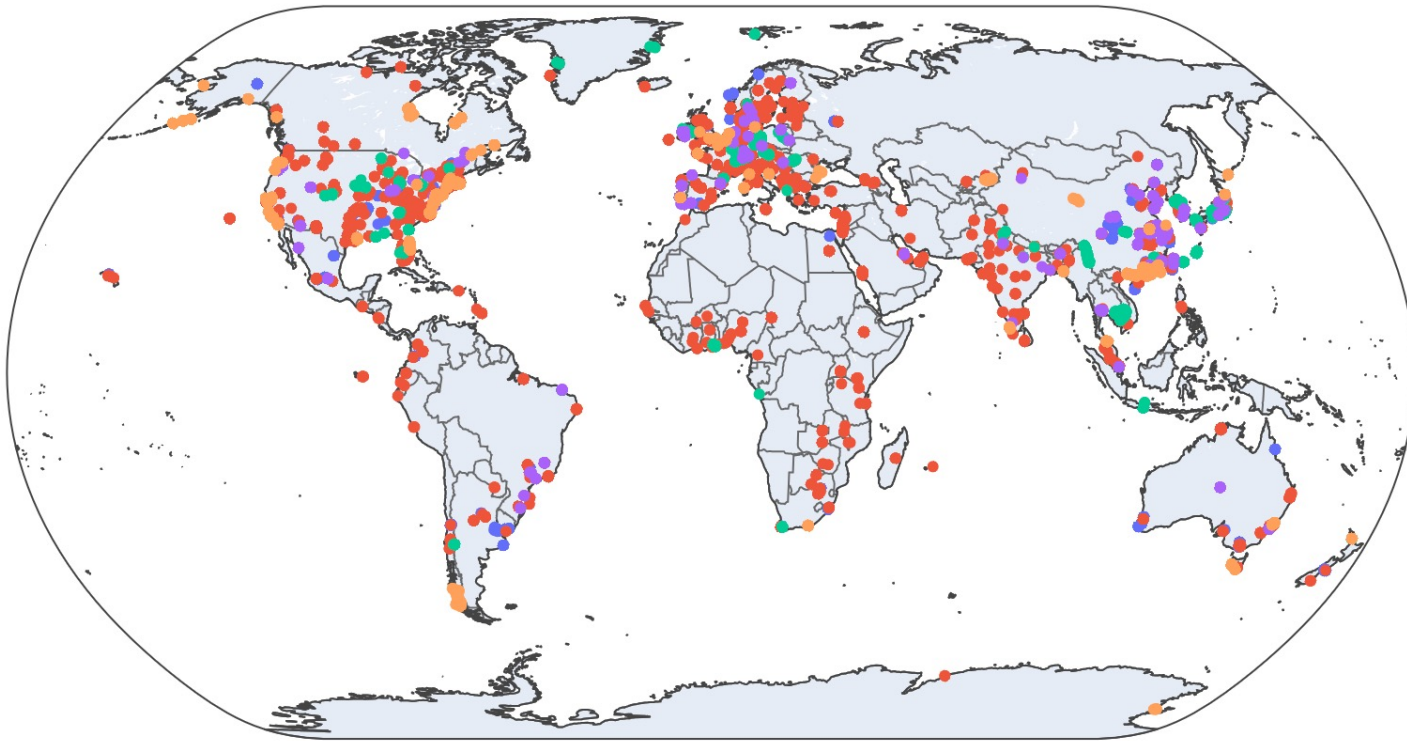
TreeMap du nombre d'acquisitions
par évaluations



Evaluation du géoréférencement

organism

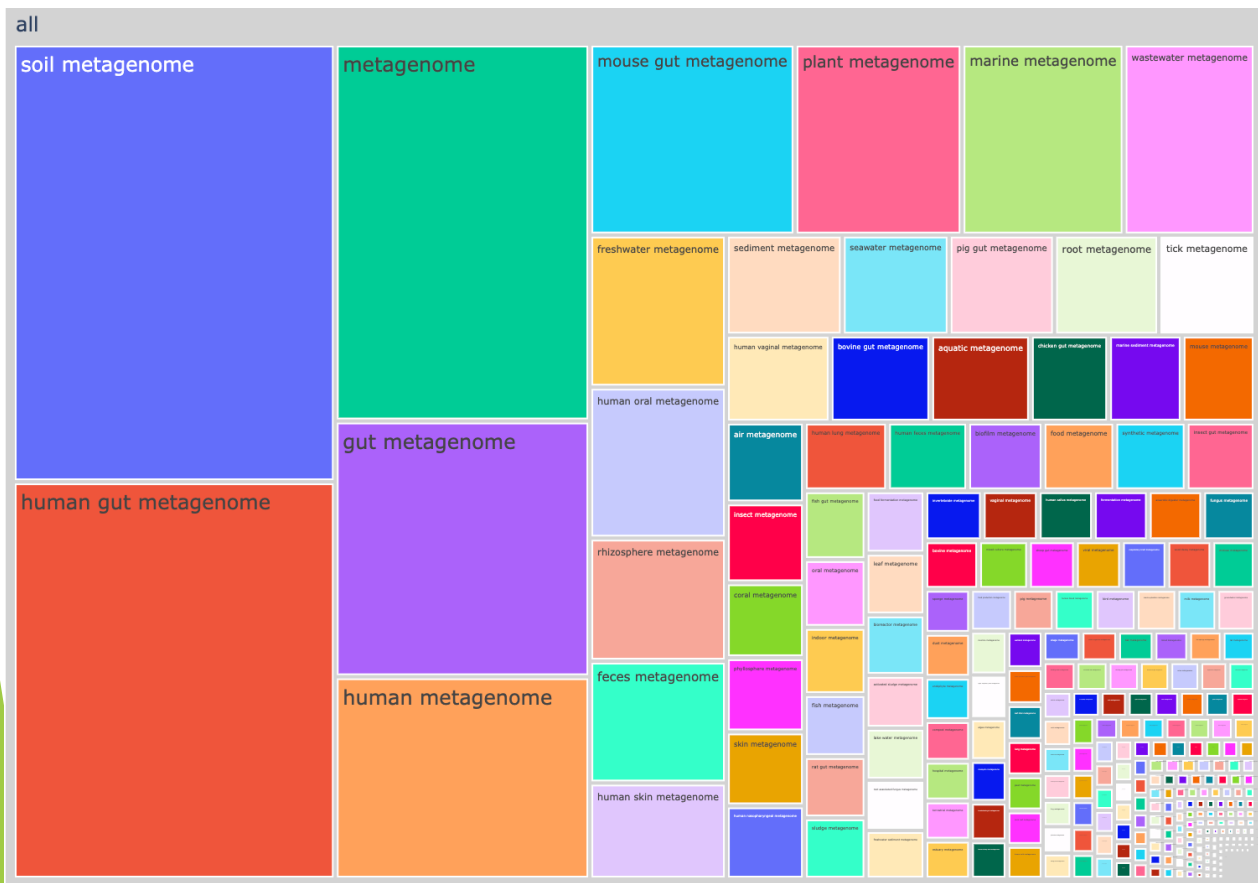
- activated sludge metagenome
- wastewater metagenome
- riverine metagenome
- sludge metagenome
- estuary metagenome



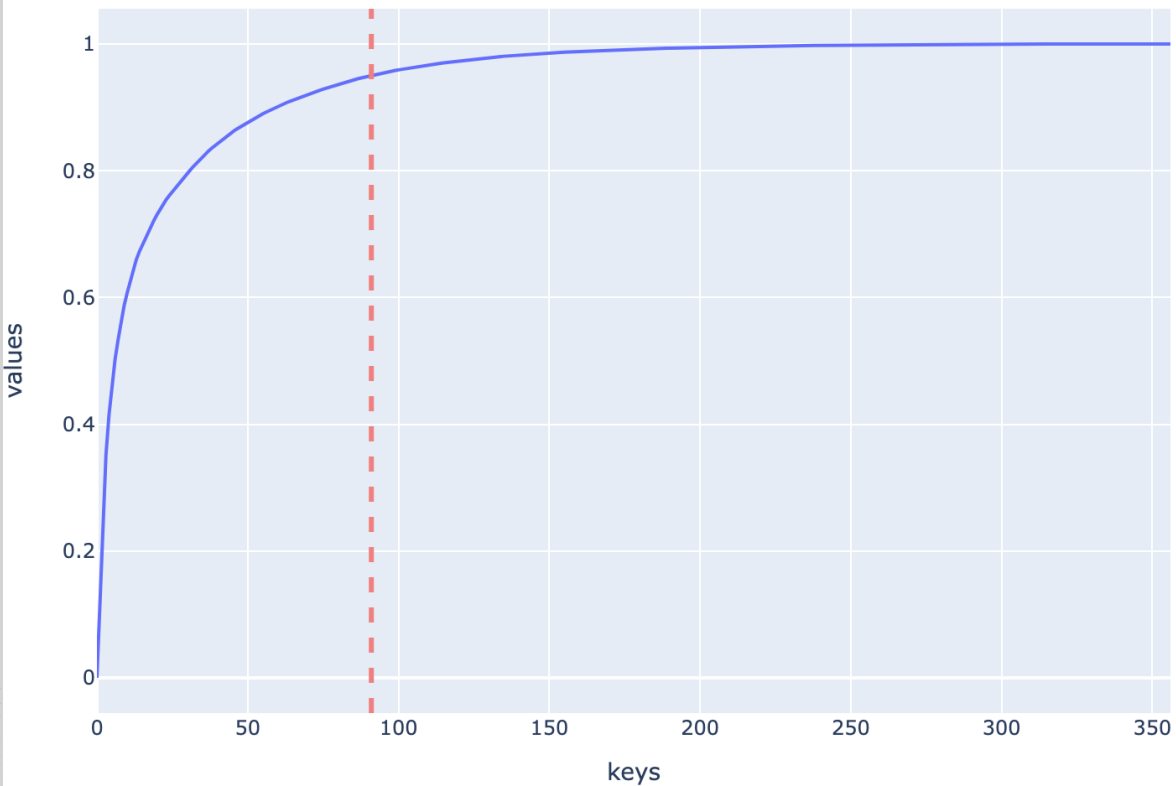
*Points d'acquisition liés aux réseaux
d'assainissement*

Exploration de GeoSRA : évaluation de organism

Evaluation de organisme

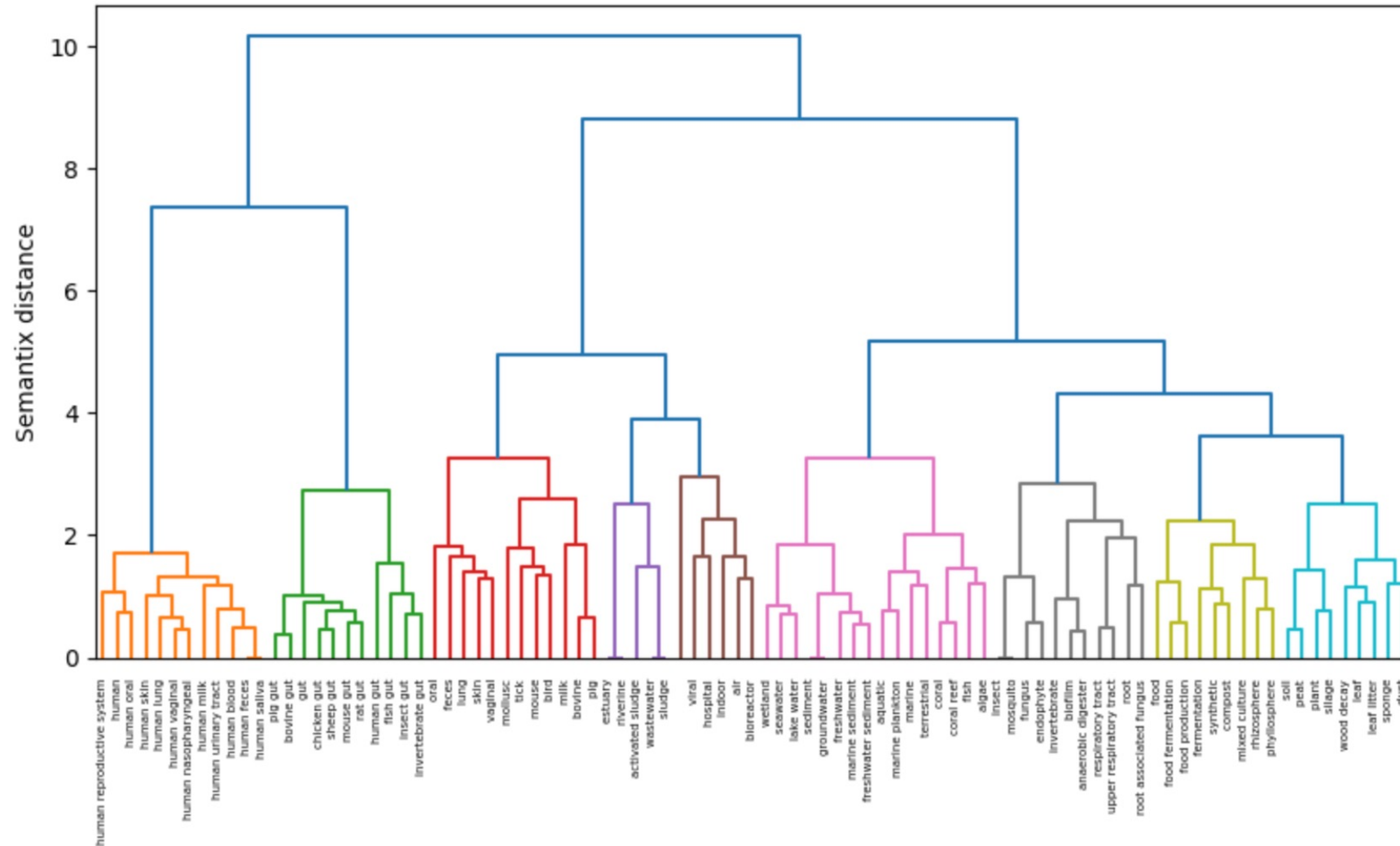


TreeMap du nombre d'acquisitions
par organisme



Fonction de répartition de la colonne organisme

Evaluation de organisme

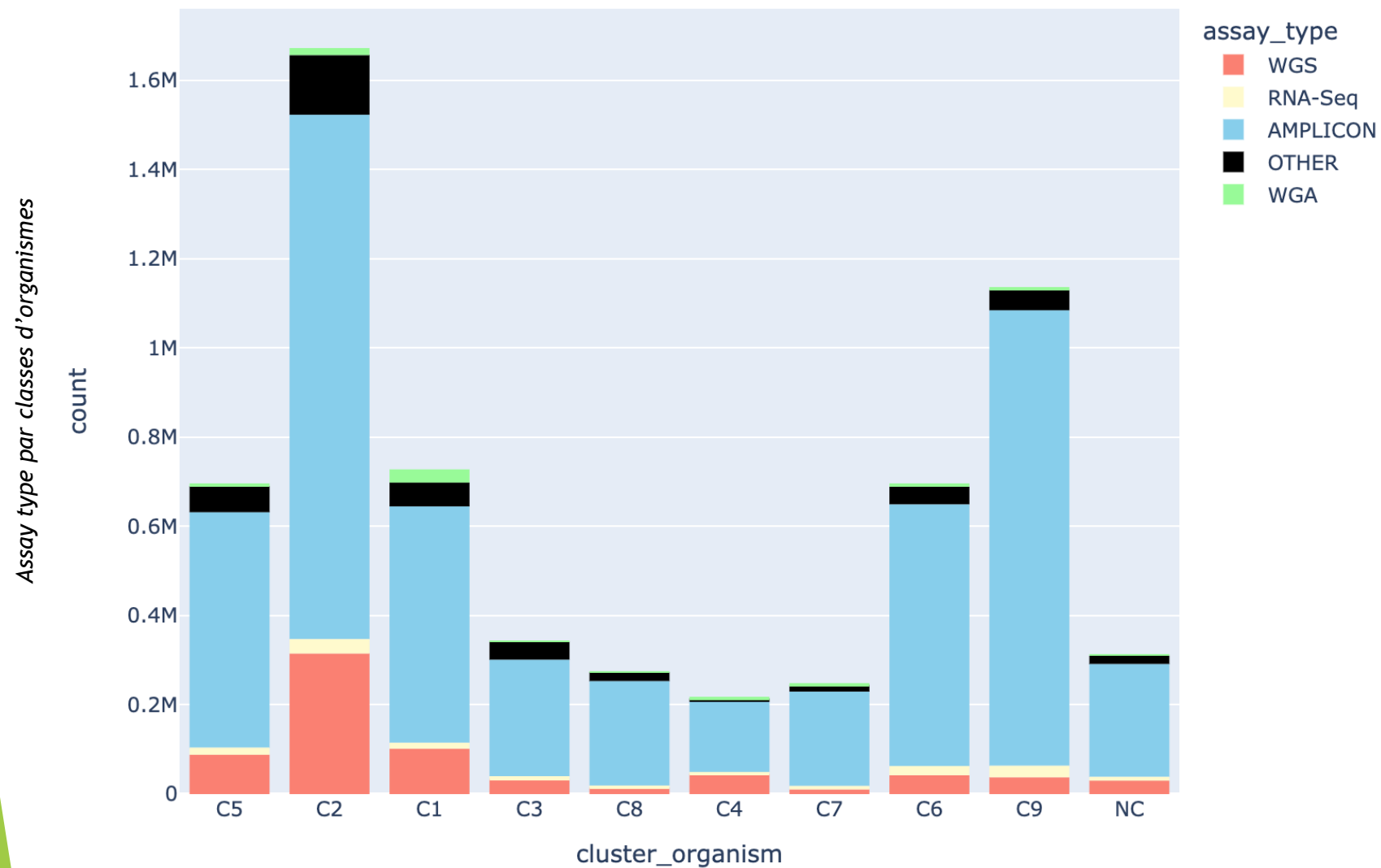


Dendrogramme sémantique des principales classes d'organisme

TreeMap du nombre d'acquisitions par classes d'organismes



Evaluation de organism



Stages 2024

Jalon 30 avril 2024

Tristan Hillairet