

Stage de Master et d'Ingénieur 2016-2017 : Intégration et visualisation de données issues du projet Patrimoine Numérique Scientifique du Cirad

Sandrine Auzoux, Sophie Fortuno, Mathieu Roche
Cirad – Campus de Lavalette
sandrine.auzoux@cirad.fr, sophie.fortuno@cirad.fr, mathieu.roche@cirad.fr

1 Contexte

Le projet Patrimoine Numérique Scientifique (PNS) du Cirad¹ est un chantier d'Établissement lancé en 2013, qui vise à gérer, conserver et valoriser les données scientifiques ou données de la recherche produites par l'établissement et ses partenaires. Dans ce contexte, de nombreux groupes de travail ont permis de contribuer à l'identification des données et d'experts pouvant porter/constituer des cas d'étude thématiques très prometteurs (Roche *et al.*, 2015).

De manière concrète, les unités de recherche du Cirad² se sont fortement mobilisés pour constituer un inventaire précis de données importantes du Cirad (cf. Figure 1). Les jeux de données inventoriés contiennent un certain nombre d'informations (meta-données), par exemple, type de données, pays d'exécution, couverture temporelle, thématiques Cirad, auteurs, etc.

The screenshot displays the 'Annuaire des données scientifiques' (Prototype) interface. At the top, there is a search bar with the text '(mot clé : pays, lieu, thème, filière, format, ...)' and a search button labeled 'Rechercher'. Below the search bar is a world map and a link for 'Index géographique'. The main content area shows search results for 'AWARE, Atlas Cartographique Web du Cirad à La Réunion'. The results include a summary, authors (Mickaël MEZINO and Agnès TENDERO), temporal coverage (2015-), geographic interest (Réunion), and keywords (information spatiale, SIG, cartographie, couches, cartes, documents, WMS, WFS, WCS, métadonnées, océan indien, interopérabilité, atlas, web, INSPIRE). The interface also features a navigation sidebar on the left with categories like 'Famille', 'Thématique', and 'Taxonomie/Ontologie/Référentiel'. The Cirad logo is visible in the top right corner.

FIGURE 1: Interface de l'inventaire des données du Cirad.

1. Centre de coopération internationale en recherche agronomique pour le développement – Établissement public à caractère industriel et commercial (EPIC) spécialisé dans la recherche agronomique appliquée aux Pays du Sud.

2. UPR AIDA (146), UMR SELMET (28), UMR TETIS (23), UMR MOISA (13), UMR AMAP (11), UPR AGIRS (8), UMR Innovation (5), UPR HORTSYS (5), UMR ART-DEV (2), UPR GREEN (2), Dgdrs-dist (1), UMR CMAEE (1)

2 Travail à réaliser

Le travail demandé dans le cadre de ce stage, détaillé en section 2, consiste à (a) intégrer et normaliser les données structurées issues de l'inventaire et de fournir des visualisations adaptées (Liu *et al.*, 2014), (b) mettre en relation les données de l'inventaire avec les publications scientifiques issues d'Agritrop³ via plusieurs entrées : informations thématiques (mots-clés), auteurs, informations spatiales, informations temporelles.

Dans le cadre de ce stage, quatre tâches principales devront être réalisées :

- Analyse et pré-traitement des données issues de l'inventaire Cirad. Le pré-traitement sera essentiellement dédié à la normalisation de certaines données et/ou meta-données (par exemple, les mots-clés).
- Mise en relation des données de l'inventaire avec les publications d'Agritrop (cf. Figure 2).
- Visualisation des données via les bibliothèques javascript *Ext JS*⁴ et *D3.js* (<https://d3js.org/> – cf. Figure 3).
- Rédaction d'un rapport incluant la description détaillée du protocole reproductible (workflow) sur d'autres ensembles de données et métadonnées.



FIGURE 2: Exemple de publication issue d'Agritrop (archive ouverte des publications scientifiques du Cirad).

L'application sera développée à partir des données de l'inventaire, en particulier les données de UPR AIDA (Agroécologie et intensification durable des cultures annuelles)⁵ qui a recensé 146 jeux de données. La généralisation aux autres unités de recherche sera également effectuée. Une réflexion pour intégrer ces propositions dans le cadre du projet étandard STRADIV (System approach for the TRANSition to bio-DIVersified agroecosystems) sera également menée.

Références

- LIU S., CUI W., WU Y. & LIU M. (2014). A survey on information visualization : recent advances and challenges. *The Visual Computer*, **30**(12), 1373–1393.
- ROCHE M., FORTUNO S., LOSSIO-VENTURA J. A., AKLI A., BELKEBIR S., LOUNIS T. & TOURE S. (2015). Extraction automatique des mots-clés à partir de publications scientifiques pour l'indexation et l'ouverture des données en agronomie. *Cahiers Agricultures*, **24**(5), 313–320.

3. archive ouverte des publications scientifiques du Cirad : <https://agritrop.cirad.fr/>

4. <https://www.sencha.com/products/extjs/>

5. <http://ur-aida.cirad.fr/>

Visual Index

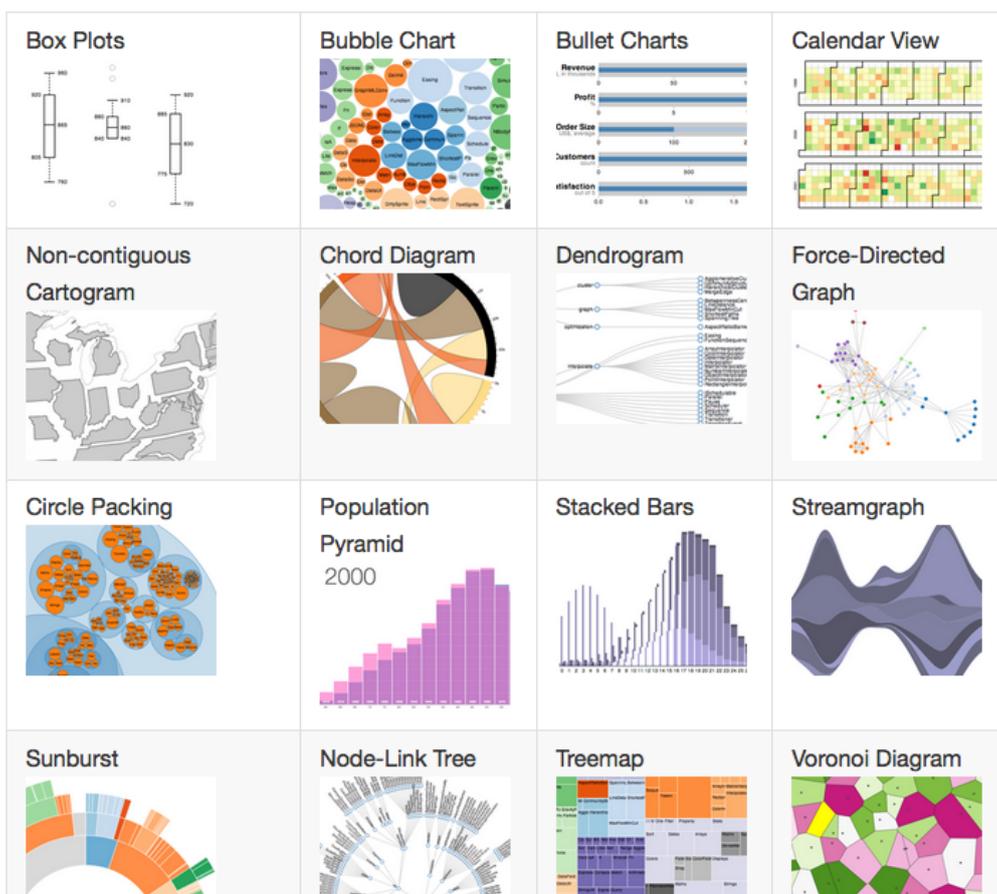


FIGURE 3: Librairie javascript D3.js.