

Assemblée MaDICS

20 - 21/11/2018

Action GRAMINEES

Responsables Romain DAVID - Luc HOGIE































GRAphe data Mining In Natural, Ecological and Environnemental Sciences





<u> MOYENS :</u>

L'indexation, l'identification pérenne des données avec le consortium IndexMeed

Rappel: Consortium IndexMeed: cadre de l'action GRAMINEES

Objectif principal d'IndexMeed:

développer la culture des bases de données et leur accessibilité

<u>Mais surtout de nouvelles méthodes d'utilisation efficace des données</u>

<u>dans le milieu de la recherche en écologie et biodiversité</u>.

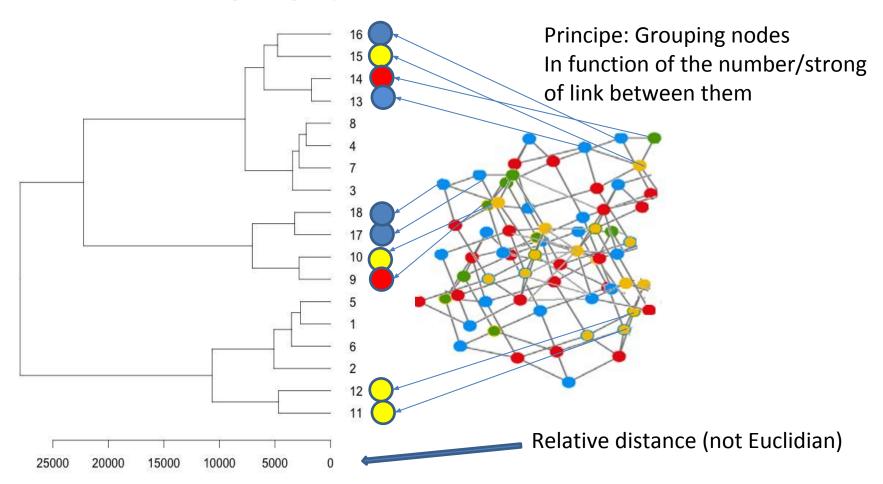
Conditions:

Des bases de données <u>ouvertes</u>, <u>interopérables</u>, <u>stables</u> (et donc recensées) et des données <u>traçables</u> et bien qualifiées,

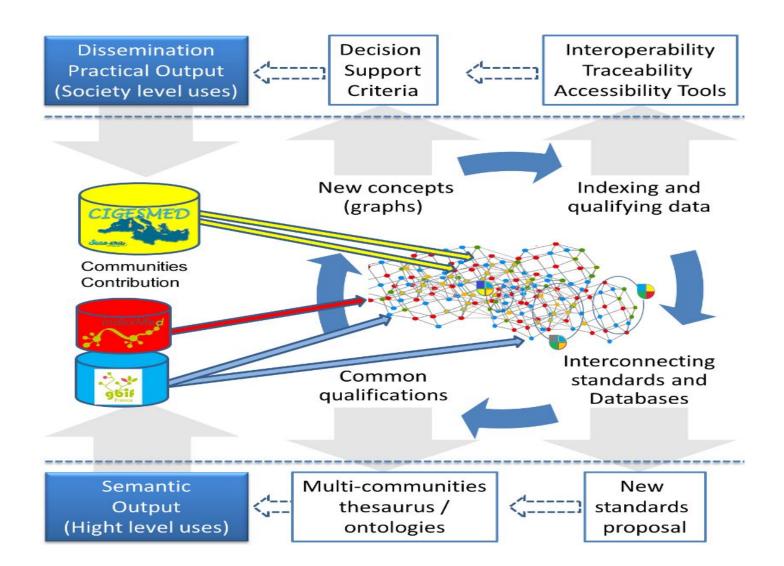
en adaptant les approches issues d'autres disciplines Astronomie, Physique, Mathématiques, Economie, Informatique, Fouille de données...

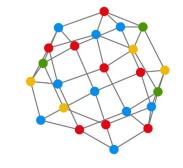
-> INTEROPERABLES et MULTIDISCIPLINAIRES

Reminder: Clustering of graphs



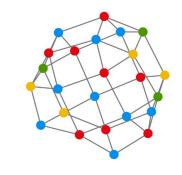
Reminder





Objectifs de GRAMINEES

1/ Développer la dynamique d'échange entre des experts en écologie/biodiversité et des experts du domaine des STIC concernant la fouille de graphes ; (GRAMINEES)



Objectifs de GRAMINEES

1/ Développer la dynamique d'échange entre des experts en écologie/biodiversité et des experts du domaine des STIC concernant la fouille de graphes ; (GRAMINEES)

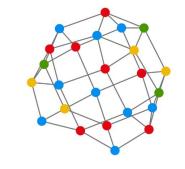
2/ Réaliser une première version de carte des compétences de laboratoires qui pourront apporter des outils méthodologiques ou des techniques algorithmiques adéquates pour l'analyse des données sur la biodiversité et l'environnement. (GRAMINEES)

Themes pour GRAMINEES 2018

Organisation d'ateliers sur des compétences complémentaires s'appuyant sur les cas d'études développés dans le cadre des ateliers de 2017.

Les thèmes proposés

- « la sémantique et les ontologies interdisciplinaires dans le domaine environnemental »
- « Algorithmes et calculs sur les graphes basés sur des données en sciences de l'environnement »



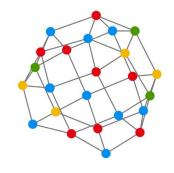
Moments marquants

Décembre 2017 - Janvier 2018 :

finalisation des cartes de compétences et graphes sur cas d'étude

production et publications des « how to do »

Ces documents reprenant étape par étape s'appuient tous les éléments des différents ateliers pour les deux périmètres de compétences STIC « Curation de données pour la construction de graphes» et « Visualisation de données sous forme de graphes»



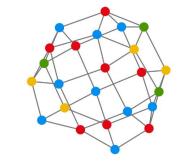
Moments marquants

2017-2018 JMaxGraph

21 - 23 mars : 10ème plénière de la Research Data Alliance

7 Juillet : Ateliers de réflexion sur les compétences autour des graphes

12 - 15 novembre : APSEM 2018 (!*)



Moments marquants

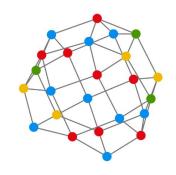
À venir

4-6 décembre : 2 posters Journées nationales de la science ouverte à Paris ((!*) JNSO - 400 participants annoncés, posters soumis)

10 - 12 décembre : préparation des ateliers algo à Nice (10 personnes prévues) (Animation Luc HOGIE(*), et Romain DAVID (!).

Décembre - Janvier : production et publications d'un « how to do » V1.0 sur le choix des algos selon le type de graphe (!*) Ce document présentant étape par étape s'appuie tous les éléments des différents ateliers pour les deux périmètres de compétences STIC. Ils reprennent le même schéma que les how to do « Curation de données pour la construction de graphes» et « Visualisation de données sous forme de graphes» produits en début d'année 2018.

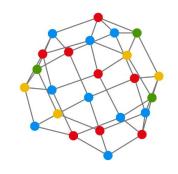




Mobilisation de données

Investigation d'une autre piste qui consiste, au sein d'un seul et même ordinateur (on ne parle donc plus de distribution), à charger par parties des données qui ne peuvent résider de manière complète en mémoire vive. Le graphe est donc stocké sur disque dur et séquentiellement chargé morceau par morceau;

Verrous 2018



Mobilisation de données

élaboration de grilles d'évaluation des chercheurs afin de mesurer leur niveau de prise en compte des principes FAIR dans la gestion de leurs données. L'objectif est d'inter-opérer les fournisseurs de données de contexte pour construire ces grands graphes environnementaux.

Atelier GRAMINEES pendant APSEM 2018



Contacts



Connexion Inscription au rés		Recherch
Vous êtes ici: DevLOG » ANF APSEM	ANF APSEM2018 : Apprentissage et sémantique l'apport croisé et les nouvelles perspectives des technologies du web des données et de la recherche par les données. Lieu: - Toulouse à ENSEEIHT Date: - du 12-15 novembre (Semaine 46) - 4 jours Nb personne: 30-40 Date limite d'inscription : 12 octobre	Table des matières
La Communauté Manifestations		•ANF APSEM2018 : Apprentissage et sémantique • Objectifs et enjeux • Description • Comité d'organisation • Comité de programme • Pré-programme • Jour1 • Jour2 • Planning • Jour3 • Jour4
Groupes de travail		
Formations en ligne Liste de discussion		
Charte du réseau		
Comptes rendus Membres du réseau		

Appel à retour d'expérience pour le jour 4

Inscription et appel à contribution

• Appel à panélistes pour les tables rondes du jour 2 et du jour 3 (présentation de 10 minutes)

Appel à questions pour les tables rondes du jour 2 et du jour 3

\$\rightarrow\$ ===== Appel \(\text{à contribution} ======

L'organisation se fait en collaboration avec l'Inra avec le soutien financier d'Ingenum pour la prise en charge des frais d'inscription pour les agents Inra.

Objectifs et enjeux

Atelier GRAMINEES pendant APSEM 2018

45 participants dont 30 à l'atelier





Atelier GRAMINEES pendant APSEM 2018

45 participants dont 30 à l'atelier



Participation active à l'ig SHARC De RDA

Partage et mobilisation des données de contextes stratégique pour GRAMINEES (utilisation de données hétérogènes et multi sources)



Operationalizing and evaluating the FAIRness concept for a good quality of data sharing in Research: RDA-SHARC-IG (SHAring Rewards and Credit Interest Group)



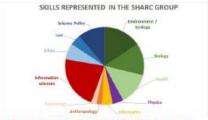
COMMUNICATION DESIGNATION OF THE CONTROL TO THE CON

Contacts: Romain David romain david@imbe fr. Laurence Mabile: Jaurence mabile@unit-tise3 fr. Anne Cambon-Thomsen: anne cambon-thomsen@unit-tise3

DEVELOPING CREDITING PROCESSES & ACTIONS TO FOSTER DATA SHARING

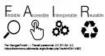
Data sharing statements and promotion is a strong reality but challenging, especially when considering the many obstacles that remain on several fronts. Among these obstacles is the lack of relevant and recognized rewarding mechanisms for the very specific efforts required to share organized distances and physical resources (Cambon-Thomsen et al., 2011; Mabile et al., 2016). The DAR-GHARD interest group is an interdisciplinary volunteer member-based group set up to unpack and improve crediting and rewarding mechanisms in the distancesories sharing process. The obsectives are

- 1/ To review the existing rewarding mechanisms and their limits in various scientific communities, and to identify key factors to improve the process and optimize the sharing of datasets and bioresources; i.e. data and physical samples (ex: tools, incentives, requirements...).
- 2/ To develop processes for stepwise adoption of principles and implementation measures tuned to national, local and institutional contexts.
- 3/ To rely upon this analysis to encourage the inclusion of physical biorecourses sharing, related oriteria in the research evaluation process at the European and national institutional levels;
- 4/ To disseminate information and findings to diverse communities of stakeholders;



ASSESSING DATA SHARING ACTIONS TO BUILD CREDITING/REWARDING PROCESSES

As a way to foster data sharing, the RDA-SHARC-IG is developing two assessment grides using criteria to establish if data are compliant to the FALIR principles (indicable (accessible) interoperable / reusable) from FORCE 11°, and on the Open Science Career Assessment Matrix designed by the EC Working group on Rewards under Open solence.



BUILDING FAIR- BASED ASSESSMENT GRIDS

To be generic and trans-disciplinary, assessment grids should be understandable by all scientist including ones who are not expert in data science.

The two grids are based on previous works on FAIR data management (Reymonet et al., 2018; Wilkinson et al., 2018; and E.U.Guidelines*):

- If The self-assessment grid to be used by a scientist as a 'checklist' to identify heribis own activities and to pinpoint the hurdes that hinder efficient sharing and reuse of higher data by all potential users.
- 2) The two-level grid (quicklentersive) to be used by the evaluator to assess the quality of the researcheriscientist sharing practice, over a given period, taking into account the means & support available over that period. Assessment criteria are classified according to their level of stringency for FAIRness (mandatory / recommended / optomail.

SHARC-IG TARGETED OUTCOMES

- Background paper on 'rewarding' mechanisms for sharing datasets & bioresources and their link to research institutional evaluation -To be published as an RDA endorsed paper in an open access highly visible science journal with a science poliny section.
- Set of generic recommendations towards relevant stakeholders meant at helping the implementation of crediting/rewarding processes for FAIR sharing
- Dissemination of the ongoing work in relevant spaces: RDA P9, Apr 2016, Barcelone, SP, RDA P11 Mar 2018, Berlin, GE; European Science Open Forum, ESOF July 2018, Toulouse, FR; RDA-P12, Nov 2018, Gaborone, BW; MaDICS Nov 2018, FR; Journées nationales Science Ouverte, Dec 2018, Paris FR.
- Connecting SHARC with RDA working groups to address specific issues such as developing a grid for assessing the FAIR sharing activity

GRAMINEES members have been participating actively in the SHARC Interest Group from the beginning. As the aim of GRAMINEES action is to develop analyses based on the graph theory with an interdisciplinary approach, FAIRness of data sharing is a sine qua non condition to permit transdisciplinary work without misunderstanding as well as preserving consistency of aggregated data.

SHARC'S MECHANISM

SHARC-IO is a recognised and endorsed interest group (40 pers., oct 2018) within RDA (Research Data Allamoer, IRCA is a community-driven organisation that aims to enable open sharing of data worldwide. RDA provides a nekating size withers like members can come logicitier through focused global Working and interest Groups. SHARC-IG works through virtual web meetings & tools as often as necessary, with a minimum of conce month for a regular update and face to face meetings when possible at each RDA plenary conference every 6 months. When needed, saxt sub-groups are formed to work on specific issues. Regular information and feedback is relayed vis the RDA SHARC's mailing list and webpage towards RDA members.

Ongoing background paper available on the RDA SHARC's webpage

PUT NEEDED FROM RESEARCH COMMUNITIES

To implement a highly fair assessment of the sharing process, appropriate criteria must be selected in order to design optimal generic assessment grids. This process requires participation, time and input from volunteer scientifics data producersuleers from various fields. Since the beginning of this bask group in RDA-6HARCHG on sept 4th, 2019, about two reliconferences per week have been held as a working subgroup, to enable exchanges between participants on this specific issue. The aim was to get feedback from a larger community regarding the validity of the criteria over various fields. The assessment grids will oliroutate within the RDA community as an online questionnative before and of 2019.

Are you a scientist producing or using data? Please participate to the development of the FAIRness assessment grids by completing the questionnaire. It will help you get credit back for your efforts!

HOW?

Join the SHARC RDA community (free) at https://www.rd-alliance.org/get-involved.html and there join the SHARC interest group at

https://www.rd-alliance.org/groups/sharing-rewards-and-credit-sharc-ig You will then be informed in real time.

Please JOIN US!
FOLLOW our work!
and if interested,
PARTICIPATE!



https://www.rd-alliance.org/groups/sharing-rewards-and-oredit-sharo-ig

* MORE INFORMATION AT:

- FAIR principles: FORCE 11:
- . SHARC-IG at RDA Plenary 9.
- auditiestatocia/libriatati
- SHARC-ID at RDA Plenary 11:
- SHARC-IQ at ESOF 2018.
- SHARC-IG at ESQF 2018:
- European Commission Directorate-General for Research and Innovation report. Evaluation of Research Censers fully acknowledging Open Science Practices, Rewards, incentives andlor recognition for researchers practicing Open Science, 2017.
- https://ec.europa.eu/research/ppenacieroe/pdf/pa.rewards.worecort final.pdf

Carlon Frances A. Green A. Carlon S. Carlon S.

Williams M.S. Security N. Market C. Court P. Section & Market C. S. Court Section (2014). A simple fearer and comprise media to Priferent School (2014). A simple fearer (2014). A simple fearer of the Court School (2014). A simple fearer (2014). A

Autor different frame (OSD) in branch (OSD). A SUM (in the Contract of Contrac



Ambitions pour 2019 - 2020

Optimisation des accès aux données (notamment de contexte c'est à dire produites par des acteurs tiers)

Préfiguration d'un "environmental data hub"?

Ambitions pour 2019 - 2020

Construction d'un écosystème (E.N.T.) pour la recherche basé sur les graphes et une science reproductible

Thèmes:

- Approches et méthodes orientées graphes spatio-temporels
- Mining de graphe, pourquoi et comment (orientés aide à la décision)?
- Intégration de données par la construction de graphes
- curation des graphes/assimilation de données/valeurs manquantes / désambiguïsation
- formalisation / sémantisation
- algorithmique, implémentation et calcul
- méthode d'analyse statistiques
- mining sémantique

Questions?