

Lettre d'information GT Atelier Données n° 5 - Novembre 2020

Journées Calcul Données : Rencontres scientifiques et techniques du calcul et des données. - Sciencesconf.org JCAD 2020 auront lieu à distance du 2 au 4 décembre 2020 sur un format horaire allégé. Le programme sera publié sous peu. <https://jcad2020.sciencesconf.org/>

Colloque virtuel “Science des données, IA et biologie” Mercredi 2 décembre 2020 <https://databio.sciencesconf.org/> La biologie est en pleine mutation technologique avec la production de données massives permettant d’appréhender de manière nouvelle les molécules, les génomes, les cellules, les organismes et les environnements. Cette mutation appelle une mobilisation renforcée de la science des données et de l’intelligence artificielle (SD-IA) pour la biologie.

GDR MaDICS_Atelier Humanités numériques : “Quels problèmes posent l’hétérogénéité des données en informatique et en Humanités numériques ?” Mercredi 25 novembre 2020 de 14h00-16h00 (en visioconférence <https://bigbluebutton.univ-lyon2.fr/gl/lou-e6q-wrp>) 14h00-14h45 - Pour gérer l’hétérogénéité des données : une alternative avec les lacs de données. Jérôme DARMONT, Cécile FAVRE, Sabine LOUDCHER (Université Lyon 2)

International FAIR Convergence Symposium 2020 The International FAIR Convergence Symposium will now take place as a fully virtual event in the week 27 November - 4 December. <https://conference.codata.org/FAIRconvergence2020/> The International FAIR Convergence Symposium is organised by CODATA and GO FAIR from November 27 till December 4. This event will provide a unique forum for advancing international and cross-domain convergence around FAIR. The two internationally recognized organisations will convene a global community with an interest in combining data across domains for a host of research issues - including major global challenges, such as those relating to the SDGs or the COVID-19 pandemic. Outcomes will directly inform the CODATA Decadal Programme ‘Making data work for cross-domain grand challenges’ and the GO FAIR community’s convergence towards bottom-up development of the Internet of FAIR data and services globally.

A des fins pédagogiques, Le CEFE a rédigé un PGD sur un projet fictif de “Suivi de population de poissons dans le lac du Bourget”. Il correspond à la version qui devra être transmise dans les 6 mois qui suivent le démarrage scientifique du projet financé par l’ANR. Il est disponible sur le site de DMP opidor (<https://dmp.opidor.fr/plans/5848/export.pdf>)

video Table ronde sur le plan de gestion de données pendant le Colloque Sciences Ouvertes 2020 : Prendre soin de ses données et les valoriser

<https://videos.univ-lorraine.fr/index.php?act=view&id=10967>

Découvrez OSMDData, le portail de démonstration des données d'OpenStreetMap sur la France. <https://www.openstreetmap.fr/osmdata-le-portail-de-la-donnee-osm/>

Announcing AWS Glue DataBrew – A Visual Data Preparation Tool That Helps You Clean and Normalize Data Faster <https://aws.amazon.com/fr/blogs/aws/announcing-aws-glue-databrew-a-visual-data-preparation-tool-that-helps-you-clean-and-normalize-data-faster/>

Ten simple rules for writing Dockerfiles for reproducible data science <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008316>

video Improving FAIRness with containers

The FAIR guiding principles state that published research objects should be made Findable, Accessible, Interoperable, and Reusable for other researchers. Data repositories provide research dissemination following FAIR principles while also developing standards and tools to facilitate them. However, increased use of advanced research methods, such as virtual containers, supercomputers and GPUs, is introducing new challenges for research sharing. There is no standardized way of describing and disseminating such research outputs in data repositories. Furthermore, dissemination of data within virtual containers like Docker may hinder some of the commonly supported principles, such as findability and accessibility. <https://sorse.github.io/programme/talks/event-020/>

Data Organization in Spreadsheets <https://www.tandfonline.com/doi/full/10.1080/00031305.2017.1375989>

Cadre commun COAR de bonnes pratiques en matière d'entrepôts
<https://zenodo.org/record/4118380> Ceci est la traduction française, réalisée par
le Noeud National RDA France, du document original "COAR Community
Framework for Best Practices in Repositories" (DOI: 10.5281/zenodo.4110829).

Nouvelles fonctionnalités relatives aux schémas de données sur data.gouv.fr
Les schémas de données permettent de décrire de manière précise et univoque
les différents champs qui composent un jeu de données et les valeurs possibles.
Ils permettent de valider qu'un jeu de données se conforme à un schéma,
de générer de la documentation automatiquement, de créer des jeux de
données d'exemple ou de proposer des formulaires de saisie standardisés.
[https://www.data.gouv.fr/fr/posts/nouvelles-fonctionnalites-relatives-aux-
schemas-de-donnees-sur-data-gouv-fr/](https://www.data.gouv.fr/fr/posts/nouvelles-fonctionnalites-relatives-aux-schemas-de-donnees-sur-data-gouv-fr/)

Logigramme de l'@institutpasteur <https://pasteur.fr/fr/file/20707/download>
Logigramme de l'@INRAE_France <https://www6.inrae.fr/datapartage/content/download/3749/39736/version/1>
AOQP+AFQN-11+juin2020.pdf et la fiche juridique associée <https://www6.inrae.fr/datapartage/content/download/3749/39736/version/1>

SIST20 Jean-Christophe Desconnets : Entrepôts de données, tour d'horizon et dis-
cussion PDF : https://sist20.sciencesconf.org/data/pages/SIST20_entrepot_de_donnees_Desconnets.pdf
Video : <https://nuage.osupytheas.fr/s/SaX9qEwR9NbDLag>

Aude Chambodut : la certification CoreTrustSeal PDF : https://sist20.sciencesconf.org/data/pages/SIST20_Cert
Video: <https://nuage.osupytheas.fr/s/SaX9qEwR9NbDLag>

Analyse I/IST n°33 – Novembre 2020 Ce devait être le web 2.0, puis
le web 3.0, voire 4.0... Le web sémantique ou web des données n'en finit
pas d'être la prochaine révolution sans cesse repoussée. [https://www.eprist.fr/wp-
content/uploads/2020/10/EPRIST_I-IST_Le-web-semantique_Novembre2020.docx](https://www.eprist.fr/wp-content/uploads/2020/10/EPRIST_I-IST_Le-web-semantique_Novembre2020.docx)