

Lettre d'information GT Atelier Données n° 2 - Juillet 2020

vous trouverez le programme de la journée ci-dessous ou sur le site
<http://devlog.cnrs.fr/jdev2020>

Vous trouverez des canaux de chat pour chaque thématique: <https://jdev-chat.insa-rennes.fr/> Les informations de connexion. Le direct sur le web :
<https://prismes.univ-toulouse.fr/player.php?code=E7l71762&width=100%&height=100%>
<https://jdev-chat.insa-rennes.fr/> <https://mediapod.u-bordeaux.fr/live/6/> Vous êtes geek, connectez-vous directement au flux réseau avec votre lecteur vidéo préféré:

Vous trouverez des serveurs de flux réseau près de chez vous (plus stables, moins de latence):

Vérifiez ou installez un lecteur vidéo : ffplay, vlc, Gnome/totem, Mplayer (microsoft et mac)

Pour toute info et assistance : <https://jdev-chat.insa-rennes.fr/channel/live>

Pour les GT d'aujourd'hui. Merci de vous inscrire (mode interactif ou mode écoute):

<https://evento.renater.fr/survey/inscription-aux-gt-des-jdev-2020-halwehr5>

Finalement, ils se feront en salle1 (choisissez votre mode à la connexion):

<https://greenlight.lal.cloud.math.cnrs.fr/b/day-6kg-ky3>

13h30-15h00 - T4T6.GT07 - Kubernetes et la gestion de grands volumes de données: des pétaoctets de données cosmiques dans Kubernetes, c'est possible! - Fabrice Jammes (IN2P3/Clermont-Ferrand)

15h30-17h00 - T4.GT03 - Les usines logicielles Manuel Grizonnet (CNES)

Bonne journée,

Le comité d'organisation

PS: la visio étiquette La visio étiquette

je prends connaissance de BBB si besoin avec cette salle.

je teste mon micro (niveau sonore, saturation, ...) et mon dispositif préalablement avec cet

je ferme mon micro quand je ne parle pas.

Lors des tables rondes, j'allume ma caméra.

Si il y a des problèmes de qualité et de bande passante, je ferme ma caméra pour laisser la

Je viens en session interactive avec un micro opérationnel pour pouvoir intervenir. A défaut

J'utilise la discussion publique pour échanger avec les personnes du mode interactif (deman

Je consulte les questions en ligne (onlineQuestion) et le fl de discussion du rocketchat po

Si je veux diffuser un contenu(transparents, films, ...), je le teste préalablement avec cet

PS: Mercredi 08/07 - Plénière de la T8 - Programmer et déployer votre IA

09h00-09h30 - Accueil dans le salon en ligne JDEV (tester votre connexion) Canal du live

09h30-17h30 - Plénière de la T8 - Programmer et déployer votre IA

13h30-15h00 - T4T6.GT07 - Kubernetes et la gestion de grands volumes de données: des pétaoctets

Nous vous invitons à contribuer à ce GT en proposant un REX ou une participation à la table ronde qui suivra

Le “Large Synoptic Survey Telescope” (LSST) est un télescope à grande ouverture et à champ large qui couvre le ciel austral. Il est actuellement en construction au Chili. Son dispositif d’imagerie prendra environ 2000 expositions par nuit, pour un volume total de données brutes de 15 TB par 24 heures. Des processus automatisés détecteront les objets célestes présents dans ces images, mesureront leurs propriétés physiques et construiront un catalogue d’objets célestes qui contiendra à terme des milliers de milliards d’entrées. Ce catalogue sera au cœur des découvertes scientifiques permises par LSST. Durant cette présentation, nous présentons les travaux en cours pour la réalisation et la mise en œuvre de la base de données distribuée en charge de ce catalogue: Qserv. Nous détaillerons comment nous avons utilisé les techniques Devops/DevSecOps les plus abouties, telles que Kubernetes, les conteneurs, les Politiques, les opérateurs, les pipelines CI/CD, pour industrialiser le déploiement de Qserv. Grâce à ces techniques, Qserv peut aujourd’hui être mis en œuvre en quelques secondes sur des centaines de serveurs en simultané, avec des volumes de données de l’ordre du Pétaoctets, que ce soit dans les grands Datacenter internationaux tels que le CC-IN2P3, ou dans des “Public Clouds” tels que Google. Ces travaux novateurs permettent notamment de consolider auprès de la communauté internationale la pertinence de Kubernetes en terme de gestion de grands volumes de données.

15h30-17h00 - T4.GT03 - Les usines logicielles Manuel Grizonnet (CNES)

Plénière de la T8 du 8 Juillet 2020

09h30-09h45 : Introduction : E. Gondet (OBS-MIP/CNRS) et L. Risser (ANITI/3IA & IMT). Le sujet

09h45-10h15 : Apprentissage Automatique :

Un autre paradigme pour les sciences numériques. J.L Parouty (CNRS).

10h15-10h45 : L'IA dans les grands organismes :

G. Eynard-Bontemps pour le CNES.

G. Larvor, L. Perez pour METEO-France : LabIA et Infrastructures.

10h45-11h15 : Infrastructures nationales pour le déploiement de l'IA en France et en Europe.

11h15-12h00 : Séances Questions. Modérateurs : E. Gondet (CNRS) et L. Risser (CNRS).

Réponses des intervenants aux "Tchat questions".

Pause

14h00-14h50 : Machine Learning et Deep Learning.

Comprendre les enjeux. - F. Camps (LAAS/CNRS).

14h50-15h00 : Jeu de données METEONET pour l'IA. G. Larvor (METEO-FRANCE).

15h00-15h30 : Séances Questions Réponses. Modérateurs : E. Gondet (CNRS) et L. Risser (CNRS)
15h30-16h00 : Pause
16h00-17h30 : Groupe de travail :
 T8.GT05 : Discussions sur les thèmes de l'apprentissage machine, automatique & co."
 A. Boucaud (CNRS), P. Navaro (CNRS), J.L Parouty (CNRS) & L. Risser (CNRS).
 Partie 1 : Modèles d'apprentissage et architectures de réseaux de Neurones
 Partie 2 : Langages, packages et usages
 Partie 3 : Hardware et grandes infrastructures
 Partie 4 : Julia et outils émergents

Projets en cours à Météo-France. G. Larvor (METEO-FRANCE).

Le data paper, un nouvel outil de communication scientifique? <https://ternumeric.hypotheses.org/374>

Un guide d'autoformation aux données de la recherche à destination des professionnels de l'information et de la documentation (production du gTIGRE (groupe de travail inter-établissements Grand Est))
<https://zenodo.org/record/3920869#.XvmdzufgrIU>

Vous êtes chercheur ou chercheuse, enseignant-chercheur ou enseignante-chercheuse, ingénieur ou ingénieure, doctorant ou doctorante et vous exercez vos missions en France ? Le Comité pour la Science ouverte a initié une enquête nationale sur les pratiques numériques des chercheurs et chercheuses en France qui aura lieu du 17 juin 2020 au 15 septembre 2020. Cette enquête a pour objectif de mieux connaître les pratiques des acteurs de la communauté scientifique française en matière d'outils numériques et des données de recherche. Elle est menée par une équipe de recherche pilotée par l'Université de Lyon (Mme Mariannig Le Béhec, ELICO, Ur fist de Lyon). Votre participation consistera à répondre en ligne à une quarantaine de questions réparties en 9 parties et nécessite une durée estimée à 20 minutes. Si vous êtes intéressé.e.s, vous accéderez au questionnaire ici : <https://questions.huma-num.fr/SurveyServer/s/SOSP-FR/outilsnumeriqueschezleschercheursenFrance/questionnaire.htm>

Les résultats donneront lieu notamment à une étude rédigée, un poster et des publications scientifiques. Une restitution publique en France et une restitution publique dans un événement international à l'étranger sont prévues. Vous pouvez également suivre le projet sur : <https://ecosysteme.hypotheses.org>

Participez à l'élaboration de la nouvelle feuille de route open data d'Etalab ! <https://www.etalab.gouv.fr/participez-a-lelaboration-de-la-nouvelle-feuille-de-route-open-data-detalab>

Parcours interactif sur la gestion des données de la recherche <https://doranum.fr/enjeux-benefices/parcours-interactif-sur-la-gestion-des-donnees-de-la-recherche/>

visio-séminaire : Les Nouveaux Paradigmes de l'Archive : L'agrégation des données numériques et leur conservation pérenne Le séminaire Les Nouveaux Paradigmes de l'Archive reprend son activité et propose un visio-séminaire via Teams le 8 juillet 2020, 10h00-12h30 Participation sur inscription auprès de claire.scopsi@lecnam.net. Nombre de places limitées à 20. Les captations des interventions seront mises en ligne quelques jours après la séance. Le tout et la partie : L'agrégation des données numériques et leur conservation pérenne.

Nous proposons de commencer cette année par une interrogation commune aux acteurs de l'information : l'agrégation de données numériques, leur conservation et leur interprétation. L'archiviste, l'historien sont confrontés à des collections, des lots de données pré-organisés par les producteurs et dont ils ignorent le sens et dont ils doivent percevoir les logiques internes. Comment conserver ce sens et le transmettre ? Quelle organisation et quelle culture de la conservation pérenne mettre en place chez les acteurs des projets liés aux data afin que les données agrégées conservent leur sens sur les temps longs ?

Nous proposons de faire dialoguer histoire, data analyse, et Knowledge Management pour mieux connaître les enjeux des données agrégées pour les historiens, et réfléchir aux actions, outils et acteurs à mettre en place pour que les données croisées des organismes et entreprises deviennent des sources exploitables pour l'Histoire. Intervenants : Gérald Kembellec, Maître de conférences en Sciences de l'information. Laboratoire Dicen-IDF Cnam Paris – Institut Historique Allemand (DHIP-IHA) Enjeux d'un conservation pérenne et accessibles des données agrégées pour les acteurs des Humanités Numériques Anna Nesvijejskaia - Directrice associée de la société QUINTEN experte en Data sciences - Docteur en Sciences de l'information et chercheuse au Laboratoire Dicen-IDF Cnam Paris. Enjeux de la transparence des algorithmes (et de leur conservation pérenne) à l'ère de l'Intelligence Artificielle Gonzague Chastenet Expert en Knowledge Management, société Ourouk. Professeur Associé à l'INTD-Cnam. L'apport du KM pour une conservation des data agrégées dotée de sens ?

Les nouveaux Paradigmes de l'Archive est un séminaire réalisé en partenariat entre le Labex Hastec, le laboratoire Dicen-IDF Cnam Paris, les Archives Nationales et le Centre Jean Mabillon (ENC) Ecoutez les interventions 2019 : <https://nparchive.hypotheses.org/captations-des-seances-2019>