

# T5 - Infrastructures logicielles et science ouverte

- <http://devlog.cnrs.fr/jdev2017>
- <http://devlog.cnrs.fr/jdev2017/t5>

## Objectifs

Comprendre l'enjeu des infrastructures numériques scientifiques en émergence. Apprendre à les architecturer et à contribuer à leur développement. Comprendre et développer son projet en utilisant les infrastructures numériques dans le contexte du big data et de l'open-access.

Comprendre pourquoi et comment développer une infrastructure logicielle. Comment organiser une communauté de développement pour co-construire une infrastructure. Comment m'appuyer sur une infrastructure de données et/ou de calcul pour développer des applications web pour mes utilisateurs.

## Description

L'explosion du volume des données, leur répartition dans le nuage nécessitent de réfléchir aux infrastructures facilitant leur interopérabilité, leurs usages et leurs valorisations.

Dans un contexte de convergence des ressources de calcul et de données, du web des données, les infrastructures numériques permettent l'automatisation de la collecte et des traitements des données indispensable pour assimiler de grands volumes, certifier, fiabiliser les traitements en vue de rendre les expérimentations, les simulations numériques et la science reproductibles.

Nous nous intéresserons à comprendre l'enjeu des infrastructures numériques scientifiques en émergence, à apprendre à les utiliser pour mettre en place ses solutions et aussi à contribuer à leur essor en proposant des ressources de données et des ressources de calcul additionnelles.

En complément à leur développement, nous apprendrons à les architecturer (architectures orientées ressources, pilotage par processus, référentiels de contenu, référentiel de type, méta-données, indexation...). Nous parlerons de leur urbanisation avec en ligne de mire la convergence du HPC et du cloud (OpenStack, SaaS, PaaS...) et la bonne utilisation des infrastructures matérielles sous-jacentes.

Nous ferons un focus particulier sur les données liées et l'open-access, les humanités numériques et l'utilisation des ressources dans les systèmes d'information scientifique.

## Mots clés

- Infrastructures numériques
- Big data

- Référencement, traitement et analyse des données
- Architecture micro-service, SOA, ROA
- Fondamentaux du Big Data (algorithmes pour le Big Data)
- Ouverture des données de la recherche (open-access)
- Ressources de données (BD , persistance)
- Ressources de calcul
- Open-access, open-data, open-édition
- Humanités numériques
- Sécurité,
- Systèmes d'Information, patterns et urbanisation
- SOA
- ROA
- Intergiciel, bus logiciel
- Micro-service
- Orchestration, chorégraphie
- Infrastructures et IOT
- RDA
- LIMS Laboratory Information Management System
- Web API
- Openstack, IAAS, PAAS, SAAS, virtualisation

## Public

- Réseaux: Devlog, Calcul, RBDD
- GDR GPL, MADICS

## Planning

- Mardi 4/07 Matin: A01, A04, GT14(fusionné avec GT13)
- Mercredi 5/07 Matin: A03, A02, GT01/GT10, GT04/GT05
- Mercredi 5/07 après-midi: A06, A08, GT06/GT07
- Mercredi fin de journée:
- Jeudi 6/07 Matin: A05, A07, A09, GT03, GT09/GT12
- Jeudi 6/07 après-midi: T5.P plénière (présentation)

## Presentations

\* **14h00-14h25:** - SOA, late-binding et agilité technique - **Linagora, Christophe DENEUX, Bertrand ESCUDIE**

\* **14h25-15h10:** - Logisland, Event Mining libre basée sur Spark & Kafka - **Thomas Baillet**

\* **15h10-15h40**: Les concepts d'une infrastructures logicielles- apis, scalabilité, résilience, couche métier, couche données,intéropérabilité - **Olivier Sallou**

\* **15h40-16h00**: - **Pause**

\* **16h00-16h35**: - SOA/WOA/ROA tenants et aboutissants - **Linagora, Christophe DENEUX, Bertrand ESCUDIE**

\* **16h35-17h00**: - [T5.PR06](#) Portage d'une architecture SOA sous Docker, exemple du système d'information du réseau d'observation ReefTEMPS - **Andry Andriatiana**([Observatoire Midi-Pyrénées](#))

\* **17h00-17h30**: Illustration de la mise en oeuvre d'une infrastructure logicielle dans le monde des SHS - perspective

## Ateliers préparatoires

### Ateliers

- [T5.A01](#) : Usage des données et des API ISTEEX - <http://www.istex.fr/> (INIST - **Nicolas Thouvenin & Stéphane Gully**)
- [T5.A02](#) : Usage des API de HAL, des méta-données et des données. Indexation et RI (CCSD - **Yannick Barborini**).
- [T5.A03](#) : Recherche d'Information (RI) à partir de plusieurs référentiels - (Alexandre Delanoë - travaille sur le projet Gargantext à l'ISC / analyse du réseau et traitement automatique des langues)
- [T5.A04](#) : Déploiement avec Docker ( **Sébastien Musso & Nicolas Muller, Treptik**)
- [T5.A05](#) : Découverte d'elastic search : beats + elasticsearch + kibana (**David Pilato**)
- [T5.A06](#) : TP BD orientée graph - Fouille de données et jointure - (**Neo4j Benoit Simard**)
- [T5.A07](#) : Openstack : configurer un tenant/configurer un IAAS. Utilisation de [Cloudwatt](#) ou de la POC **Philippe Saby OMP et Objectif Libre**
- [T5.A08](#) : Roboconf, Orchestration de conteneurs pour l'intégration continue. Encapsulation de composants logiciels ou de machines virtuelles pour leur liaison avec le Cloud. Déploiement hybride -privé et public-(en lien avec la T8 atelier T8.A01) - **Vincent Zurczak/Linagora**
- [T5.A09](#) : Atelier Flux d'événement massif: mettre en place une chaîne de traitement, collecte, acquisition, configurer, création et exécution d'un algo de matching learning, plugin - (**LogIsland & Thomas Bailet**)

### Groupes de travail

\* Groupe 1

- [T5.GT01](#) : Openstack pour modéliser et tester son architecture SOA ou ROA - (**Olivier Sallou et ??**)

\* Groupe 2

- T5.GT03 : Infrastructure numérique et communauté de développement - (**Stéphane Ribas**)
- [T5.GT11](#) : Intégrer ses développements au sein d'une infrastructure - (**Yvan Le Bras**), voir [T5.GT04 Stratégie de publication pour l'OpenScience](#)

\* Groupe 3

- [T5.GT04](#) : Stratégie de publication pour l'OpenScience (publication, code et données) - (**Marie Farge**)
- [T5.GT08](#) : Comment protéger son jeu de données dans une infrastructure ouverte? (comité éthique, représentation des données interface, embargo, ...) - (**Yvan Le Bras**) voir [T5.GT04 Stratégie de publication pour l'OpenScience](#)

\* Groupe 4

- [T5.GT06](#) : Recherche textuelle - (**Patrice Bellot**)
- [T5.GT07](#) : Les API pour la recherche et la fouille de documents scientifiques et techniques - (**Alexandre Delanoë- Gargentext**)

\* Groupe 5

- [T5.GT05](#): La question de la pérennité des données de la recherche : les plateformes et les infrastructures - (**Nicolas Larrousse**)
- [T5.GT12](#) : Standardisation des métadonnées pour l'interopérabilité des infrastructures : illustration par l'infrastructure européenne EPOS et des observatoires virtuels - (**Emmanuel Delage - OPGC**)

\* Groupe 6

- T5.GT09 Quels usages pour la recherche du traitement temps réel de flux massifs d'événements (**LogIsland - Thomas Bailet**)
- [T5.GT10](#) : Mise en place d'une architecture robuste / Sécurité & scalabilité : Retour d'expérience sur la plate-forme Hypothèse - (**Florentin Clouet Cléo / Open Edition / CNRS**)

\* Groupe 7

- [T5.GT13](#) - les identifiants permanent, pourquoi, comment? La problématique de la citation des données (et des requêtes!) - RDA : Françoise Génova et/ou François-André (OMPs) / Raphael Ritz Fusionné avec [T5.GT14](#)
- [T5.GT14](#) - Intéropérabilité des référentiels de données pour lier les ensembles de données variés - (**Françoise Génova et/ou François-André (OMPs) /Raphael Ritz**)