



ReproVirtuFlow



ReproVirtuFlow : Reproductibilité des expériences d'analyse de données scientifiques

Journée du Mercredi 16 mars 9h45 - 10h-16h Orsay LRI

Bâtiment 650 <https://www.lri.fr/info.pratiques.php>

Thématique : Vers un état de l'art sur les Workflows - Provenance - Machines virtuelles - approches émergentes - use cases

Objectifs: couvrir aux mieux les standards, les approches fonctionnelles ("qui marchent"), les approches les plus prometteuses...

Participants

- Sarah Cohen-Boulakia, LRI, Université Paris Sud, Orsay (CNRS INS2I UMR 8623)
- Christophe Blanchet, Institut Français de BioInformatique, Gif-sur-Yvette (CNRS INSB UMS3601)
- Khalid Belhajjame, LAMSADE, Université Paris Dauphine (CNRS INS2I UMR 7243)
- Konrad Hinsén, Centre de Biophysique Moléculaire (CNRS UPR4301, Orléans)
- Pierre Larmande, Institut de Biologie Computationnelle, Montpellier (IRD, UMR DIADE) (à distance)
- Christophe Pradal, Institut de Biologie Computationnelle, Montpellier (CIRAD, Inria, UMR AGAP)
- Jérôme Chopard, Institut de Biologie Computationnelle, Montpellier (INRA)
- Hervé Ménager, CIB, C3BI, Institut Pasteur
- Fabien Mareuil, CIB, C3BI, Institut Pasteur
- Alban Gaignard, Université de Nantes
- Yvan Le Bras, IRISA, Rennes (UMR INS2I CNRS 6074)
- Olivier Collin, IRISA, Rennes (UMR INS2I CNRS 6074)
- Angela Bonifati, Université Lyon 1, LIRIS (UMR INS2I CNRS 5205) (à distance)
- Christine Froidevaux, Laboratoire de Recherche en Informatique, Université Paris Sud, Orsay (CNRS INS2I UMR 8623)

Programme

9h30 Accueil

10h-11h30

- Sarah Cohen-Boulakia: *Breve introduction à la problématique de reproductibilité dans le groupe ReproVirtuFlow* (10 min)
[Slides](#)
- Fabien Mareuil: *reproductibilité et Galaxy* (10 min)
[Slides](#)
 - Galaxy + bio-tools : <https://github.com/C3BI-pasteur-fr/ReGaTE>, <http://regate.readthedocs.org/en/latest/index.html>, <https://github.com/bio-tools/galaxy/tree/biotools>
- Khalid Belhajjame: *Une visite rapide de PROV et certaines de ses extensions* (10-15 min)
[Slides](#)
- Alban Gaignard: *Provenance et Linked Data pour l'interprétation et le partage de résultats d'expérience in silico (domaine biomédical)* (10 min) [Slides](#)
- Hervé Ménager : *Les registres de ressources en bioinformatique* (10 min)
[Slides](#)
 - Registre bio.tools <http://bio.tools>
 - Autres <http://biocatalogue.org> ,

Discussions

11h30-12h

- Christophe Blanchet : *le cloud IFB et ses images, machines virtuelles et conteneurs* (10 min) [Slides](#)
- Yvan Le Bras : *Docker et Galaxy, un mariage heureux pour plus de reproductibilité* (10 min) [Slides](#)
 - Poster Galaxy Community Conference 2014 sur la POC [ISA-DataFlow](#)
 - [Présentation](#) Cloud + Galaxy + Docker pour une application en RADseq
 - Démo [IPython et RStudio sous Galaxy](#)
 - ISA recent info:
 - [Nature scientific data](#)
 - FAIR related recent documents :
 - <http://www.dtls.nl/the-fair-principles-herald-more-open-transparent-and-reusable-scientific-data/>
 - <http://www.nature.com/articles/sdata201618>



ReproVirtuFlow



Discussions

12h-12h45

- Christophe Pradal - Jerome Chopard: Notebooks et OpenAlea (10 min)
[Slides](#)
- Konrad Hinsen: Papiers reproductibles ([ActivePapers](#), [everpub](#)), reproductibilité des calculs HPC (10 min)
[Slides](#)

Discussions

12h45-13h45 Repas sur place (plateaux repas)

13h45-16h Discussion

Etat de l'art ? Autres approches à creuser / Autres communautés ?

Les défis restants : pratiques, techniques, théoriques... possibilité d'illustrer sur des use cases (données plantes, données médecine...)