

## **Séminaire «Méthodes et outils pour l'Open Data»**

Montpellier 17-19 Décembre 2014

Hameau de l'étoile

### **La problématique**

Ce séminaire s'inscrit dans la démarche « Open science et Open data » engagée par l'Inra en 2013. Cette démarche est construite sur un ensemble de principes institutionnels visant à fournir un cadre et à développer notre capacité à gérer, partager et permettre la réutilisation de données scientifiques.

Le partage et la réutilisation de données demandent d'adopter des standards, de maîtriser des méthodes et des outils bien spécifiques ainsi que de mettre en place des infrastructures adaptées. Elles demandent aussi d'être capable de prendre en compte des aspects légaux, éthiques et de s'inscrire dans une politique de diffusion contrôlée et raisonnée.



### **Contexte et enjeux**

Ce séminaire fait suite à la mise en place de l'initiative INRA Open science et partage de la donnée lancée par la direction générale en mars 2013. Elle se positionne clairement dans des priorités du schéma directeur des systèmes d'information. Le succès de cette démarche institutionnelle passe par la connaissance approfondie et la mise en œuvre de méthodes nouvelles dans l'institut. Toutes les familles de données scientifiques (expérimentales, enquêtes, ORE, génomiques, bibliographiques, cohortes,...) produites et utilisées à l'INRA sont directement concernées. Cependant, leur partage suscite beaucoup de questions comme les formats d'échange de données, les modèles de métadonnées, les cycles de vie ou encore la qualité des données. Aujourd'hui, des méthodes et des outils originaux comme ceux développés dans le cadre du Web Sémantique et du « Linked data » permettent d'apporter des réponses à ces questions et d'avancer dans la réflexion autour de notre patrimoine numérique et sur les moyens associés à mettre en œuvre.



### **Objectifs et attendus**

- Rencontre pluri-organismes entre des équipes math-info
- Favoriser les collaborations entre équipes et consolider une communauté composite naissante
- Partager des méthodes pour la capitalisation et la pérennisation des données
- Identifier les futurs outils autour du partage de données
- Identifier des fronts de sciences
- Identifier les compétences à développer et leurs moyens de diffusion (tutoriels, formations...)

Ce séminaire permettra également de poser les premières briques pour l'organisation d'une école chercheurs / ingénieurs en 2015 :

- Faire évoluer les compétences d'un collectif ingénieurs/chercheurs « math-info » par l'acquisition de connaissances et de savoir-faire dans le domaine du partage de données
- Acquérir des compétences dans des méthodologies spécifiques comme la mise en œuvre et la construction de modèles de métadonnées



### Action envisagée

**Durée** : 3 jours autour de 4 sessions

- Session Introduction au Web de données
- Session Fouille de textes et d'images ; datamining
- Session Sémantisation, réutilisation et visualisation des données
- Session Linked open data

**Période** : 17-19 Décembre 2014

**Localisation** : Montpellier, Hameau de l'étoile, Saint Martin de Londres



### Responsables du projet :

- Pascal Neveu (INRA MISTEA) et Patrice Buche (INRA IATE)



### Commanditaire (s) :

- CDSI (comité directeur des Systèmes d'information) – Isabelle Blanc – Odile Hologne – Christine Gaspin (animateurs chantier gestion et partage des données de recherche)



### Public envisagé :

- Personnes sensibilisées à la problématique du partage de données : scientifiques, informaticiens, ingénieurs IST,...
- Personnes concernées par des actions de partage de données au sein de projets ou pour une diffusion très large.
- Public chercheurs et ingénieurs Inra et autres organismes : environ 30 participants

## Les suites envisagées

La finalité est de présenter à un public sensibilisé à la problématique du partage de données l'état de l'art et les fronts de science actuels dans le domaine du Web sémantique et de l'Open Data.

## Programme prévisionnel

### **17 Décembre**

#### Session Introduction au Web de données

- 14h00-14h30      Enjeux et stratégie pour l'Inra et objectifs du séminaire : Isabelle Blanc/  
Fred Garcia
- 14h30-15h15      Introduction au Web Sémantique : Olivier Corby/Catherine Faron-Zucker  
(INRIA WIMMICS)
- 15h15-15h45      Pause-café
- 16h00-19h00      Tutoriel RDF/RDFS/OWL/SPARQL : Olivier Corby/Catherine Faron- Zucker  
(INRIA WIMMICS)

### **18 Décembre**

#### Session Fouille de textes, d'images et datamining

- 09h00-10h00      Texte et datamining
- Fouille de données spatio-temporelles: Maguelonne Teisseire (IRSTEA TETIS)
  - Fouille de texte : Claire Nédellec (INRA MIG)
- 10h00-10h30      Pause café
- 10h30-12h00      Image/video mining
- Construction de résumés de vidéo : Yingbo Li (ENS)
  - Quelques outils du domaine public : Alain Trubuil (MIA Jouy)
  - Gestion du cycle de vie des images en microscopie pour la biologie: Perrine Paul-Gilloteaux (Institut Curie)
- 12h00-12h15      Bilan session – Discussions- Table ronde

#### Session Sémantisation, réutilisation et visualisation des données

- 13h45-15h15      Capitalisation et sémantisation des données
- Capitalisation et réutilisation de données scientifiques - le projet @Web : Patrice Buche (INRA IATE) et Juliette Dibie (AgroParisTech, MIA Paris)
  - Visualisation de réseaux : Ludovic Cottret (LIPM)
  - Intérêt et utilisation d'une ontologie médicale en lien avec des images, enrichissement avec des règles simples : Federico Ulliana (INRIA GraphIK)
  - Table ronde Sémantisation et visualisation
- 15h15-15h45      Pause-café

15h45-16h15	Alignement d'ontologies : Konstantin Todorov (LIRMM)
16h15-16h45	Recommandation : Esther Paccitti (INRIA Zenith)
16h45-17h15	Pause-café
17h15-18h15	Témoignages, retours d'expériences, besoins exprimés par des utilisateurs <ul style="list-style-type: none"> <li>• Emballages alimentaires : Valérie Guillard (UM2/IATE)</li> <li>• La plateforme ANAEE F : Christian Pichot (URFM)</li> </ul>
18h15-18h45	Bilan session - Discussions

## **19 Décembre**

### Session Linked open data

09h00-09h30	Linked open data : Marie-Christine (Rousset LIG)
09h30-10h00	Les chenilles processionnaires : Sophie Aubin (DV-IST)
10h00-10h30	Pause-café
10h30-11h30	Ateliers en deux sous-groupes : compétences, quelle démarche mettre en place pour mieux réutiliser les données ?
11h30-12h00	L'approche OBDA : Marie-Laure Mugnier (INRIA GraphIK)
12h00-13h00	Restitution des ateliers et discussion générale