

Les données ouvertes en sciences environnementales : concepts et méthodes de qualification et de partage des données à l'ère du big data

Campus CNRS de Gif-sur-Yvette

28-31 octobre 2019



Sommaire

Description générale	3
Objectifs de formation	3
Grands axes du programme 2019	4
Organisation et contexte	4
Coordination	4
Comité d'organisation	4
Comité scientifique	4
Programme complet	5
Documents associés et liens utiles	7
Description des enseignants	8
Liste de contact des enseignants	9
Liste de contact des participants	10

Description générale

Les Infrastructures de Recherche pour l'Environnement (ENVRI) fournissent aux communautés scientifiques, aux décideurs politiques et économiques et plus généralement aux citoyens des **informations sur les variables caractérisant les milieux atmosphériques, océaniques, la terre interne, les écosystèmes terrestres et la biodiversité**. Ces informations, dont certaines contribuent à des réseaux internationaux de mesure, permettent la caractérisation de l'état de l'environnement et sont partie intégrante des systèmes d'observation de la Terre. Les données d'observation in-situ et d'expérimentation sont par nature très hétérogènes tant dans leur dimension spatiale et temporelle.

Ainsi, le futur des sciences de l'environnement est marqué par la **nécessité de l'assemblage d'informations de sources et natures très variées** :

- des analyses et des intégrations sur de grandes échelles spatiales et temporelles (e.g., réseaux de satellites et de capteurs) ;
- des couplages entre des données et des modèles mathématiques ou statistiques (e.g., modèles d'écosystème ou modèles de circulation atmosphérique) ;
- des données issues de capteurs mais également issues d'observateurs aussi bien chercheurs que grand public (ex., suivi de populations ou sciences participatives).

Les enjeux liés à cette hétérogénéité des données sont de permettre **une plus grande interopérabilité entre les bases et les jeux de données** issues de communautés historiquement indépendantes et fragmentées, et de natures souvent fondamentalement différentes, tout en facilitant l'utilisation de référentiels communs et de standards de métadonnées suffisamment précis pour aider à leurs réutilisations.

Les conséquences pour la communauté scientifique sont fortes et l'on voit apparaître de **nouveaux métiers de type « Data Scientist » dont l'objectif est de produire des informations dites FAIR** (Findable, Accessible, Interoperable and Reusable) pour la science et la société. Ceci nécessite de définir et de mettre en œuvre des standards communs pour faciliter l'acquisition, le traitement et l'annotation des données, ainsi que de nouvelles pratiques de partage et d'exploitation afin de répondre aux enjeux scientifiques d'une analyse globale de la crise environnementale actuelle, de l'amélioration de la compréhension des processus et de la modélisation prédictive.

L'école thématique e-Envir dont le premier cycle est organisé en 2019 vise à répondre à cet enjeu par des actions annuelles de formation du personnel.

Objectifs de formation

Les objectifs scientifiques de l'école thématique consistent à l'acquisition des connaissances et des compétences du FAIR dans la gestion des grands jeux de données des sciences de l'environnement, notamment :

- Identifier et partager les enjeux que représente le partage des données pour consolider les résultats scientifiques et traiter les problématiques complexes des sciences de l'environnement ;

École thématique 2019 – e-Envir – Gif-sur-Yvette

- Développer la culture scientifique et technique sur la production/utilisation de données FAIR et sur les opportunités offertes par le partage des données ;
- Informer sur les stratégies et outils de valorisation de données FAIR afin de faire émerger des questions de recherche et projets basés sur ces ressources ;
- Connaître les référentiels notamment sémantiques de caractérisation des ‘données’ afin de permettre leur bonne utilisation dans les projets de recherche mais aussi favoriser la contribution des scientifiques français à ces référentiels ;
- Faire mieux connaître le cadre législatif national et international.

Grands axes du programme 2019

L'école thématique 2019 sera organisée autour de trois axes prioritaires pour cette première session du cycle de formation :

- une journée de présentation des principes généraux de l'open data en sciences de l'environnement abordant les enjeux conceptuels, juridiques et techniques. Ces présentations seront suivies de deux ateliers dédiés sur les Plans de Gestion de Données FAIR pour les projets et les infrastructures de recherche et à la Qualité des données;
- une journée et demi de présentation des enjeux et des principes de (i) l'annotation des données en sciences de l'environnement, (ii) la production de métadonnées standardisées, et (iii) les ontologies ;
- une demi-journée de formation la mise à disposition et la publication des données et métadonnées suivie d'un atelier avec toutes les infrastructures de recherche autour du programme Européen ENVRI-FAIR.

Organisation et contexte

L'école thématique 2019 est coordonnée par l'infrastructure nationale AnaEE France et l'UMS BBEES avec la communauté des infrastructures du domaine de l'environnement dont RZA et OZCAR (eLTER France), ACTRIS, et le Pôle de données Système Terre. Elle vient compléter les objectifs et actions des pôles de données et du GDR Semandiv pour la contribution au Web sémantique.

Coordination

Jean-François Le Galliard (CNRS, AnaEE France) et Chloé Martin (CNRS, BBEES - RBDD)

Comité d'organisation

Sylvie Galle (IRD, OSUG), Guillaume Brissebrat (CNRS, Sedoo/OMP, OZCAR), Valérie Thouret (CNRS, Envri), Christian Pichot (INRA, AnaEE), Marie Claude Quidoz (CNRS, CEFÉ - RBDD)

Comité scientifique

Christine Plumejeaud (CNRS, RZA), Olivier Lobry (CNRS, OSU OREME - IST), Isabelle Braud (IRSTEA, OZCAR), André Chanzy (INRA, AnaEE), Paolo Laj (CNRS, Envri)

École thématique 2019 – e-Envir – Gif-sur-Yvette

Les sponsors de l'Ecole thématique 2019

Centre National de la Recherche Scientifique		Finance le programme de l'Ecole thématique 2019
AnaEE France		Coordonne et finance le programme de l'Ecole thématique 2019
OZCAR		Aide au programme de formation 2019
Data Terra		Aide au programme de formation 2019
Zones Ateliers		Aide au programme de formation 2019
ACTRIS		Aide au programme de formation 2019
GDR Semandiv		Aide au programme de formation 2019

Programme complet

Lundi 28 octobre

9h30 à 10h15- Accueil et installation. Château de Gif.

10h15 – Mot de bienvenu et présentation du programme de l'Ecole thématique par Jean-François LE GALLIARD (CR CNRS, CEREEP-Ecotron IleDeFrance) et Chloé MARTIN (CNRS, BBEES). Grande salle.

10h30 à 12h - Introduction à l'Open data et la gestion des données

10h30. Cours magistral. « Produire, utiliser et diffuser les données en science ouverte : présentation du cadre juridique ». (Stéphanie RENNES, INRA). Grande salle.

Pause déjeuner (12h-14h).

14h à 17h30 - Principes FAIR et Plan de Gestion des Données

14h. Cours magistral. Introduction au cycle de vie des données, aux principes FAIR et au Plan de Gestion des Données ou DMP (Yvette LAFOSSE et Coralie WYSOCZYNSKI, CNRS, INIST). Grande salle.

Pause

15h30. Atelier. GopenDoRE – jeu coopératif permettant d'échanger sur les bonnes pratiques de gestion et de partage des données de recherche (Yvette LAFOSSE et Coralie WYSOCZYNSKI, CNRS, INIST). Grande salle et Salle des commissions.

19h30-20h30 Cocktail dinatoire dans le Château.

Mardi 29 octobre

9h-12h00 – Métadonnées et sémantique pour les données de la recherche en environnement

9h. Evaluer la qualité de ses données - exemple d'utilisation d'OpenRefine (Dr. Chloé MARTIN, CNRS, UMS BBEES). Grande salle.

9h30. Cours magistral. Les principaux standards de métadonnées à l'échelon international : comment s'y retrouver et comment les exploiter pour la recherche en sciences de l'environnement ? (Dr. Julien BARDE, IRD, UMR MARBEC). Grande salle.

10h30-11h Pause.

11h00. Cours magistral. Intérêt et apports scientifiques de la synthèse de données : importance des ressources terminologiques (Dr. Eric GARNIER, CNRS, UMR CEFÉ). Grande salle.

Pause déjeuner (12h00-14h).

14h-17h30 - Comment découvrir et exploiter les thésaurus existants

14h. Cours magistral. Technologies pour la gestion des thésaurus et de leur interopérabilité : standards, outils de gestion et d'alignement (Dominique VACHEZ, CNRS, INIST). Grande salle.

15h. Cours magistral. Les portails d'accès aux ressources sémantiques (Dr. Clément JONQUET, Université de Montpellier, UMR LIRMM). Grande salle.

16h-16h30 Pause.

16h30. Cours magistral. Panorama des thésaurus existants dans le domaine des sciences de l'environnement (Dominique VACHEZ, CNRS, INIST). Grande salle.

Atelier du soir (optionnel) 18h30-19h30 : Apprendre à utiliser OpenRefine

19h30-20h30 : Repas Bâtiment 20 salle Automne

Mercredi 30 octobre

9h-12h00 - Comment créer son thésaurus et l'aligner avec ceux qui existent ?

9h. Retour d'expérience. Mise en place d'un thésaurus de traits de plantes (Dr. Eric GARNIER, CNRS, UMR CEFÉ). Grande salle.

10h. Atelier. Créer son thésaurus avec OpenThésos (Dr. Eric GARNIER, CNRS, UMR CEFÉ ; Marie-Claude QUIDOZ, CNRS, UMR CEFÉ). Grande salle.

11h. Retour d'expérience. Le thésaurus AnAEE : processus d'élaboration et exploitation (Philippe CLASTRE, INRA, URFM). Grande salle.

Pause déjeuner (12h-14h).

14h-17h30 – Exploitation web sémantique et ontologies

École thématique 2019 – e-Envir – Gif-sur-Yvette

14h-14h10 Introduction au web sémantique (Pr. Danielle ZIEBELIN, Université Grenoble Alpes, UMR LIG)

14h10-14h35 Projet de portail de données Theia/OZCAR (Véronique CHAFFARD, IRD, IGE et Isabelle BRAUD)

14h35-15h Gestion et valorisation sémantiques de données de biodiversité et d'études d'écosystèmes dans l'infrastructure ANAEE-France (Christian PICHOT ou Damien MAURICE, INRA, UMR EEF)

15h-15h25 Projet Trajectories 20 min + 5 min questions - (Pr. Danielle ZIEBELIN, Université Grenoble Alpes, UMR LIG)

Pause.

16h. Table ronde animée par Pr. Danielle ZIEBELIN, Véronique CHAFFARD, Dr. Eric GARNIER et Damien MAURICE/Christian PICHOT. Discussion libre sur la thématique.

Atelier du soir (optionnel) 18h30-19h30 : PI4*4 et centipède : des initiatives pour le terrain

19h30-20h30 Cocktail dinatoire dans le Château.

Jeudi 31 octobre

9h-12h00 – Publier et accéder aux données

Cours magistral. Les entrepôts de données ou comment rendre les données trouvables et accessibles des données publiées (Dr. Jean-Christophe DESCONNETS, IRD). Grande salle.

Pause.

Cartographie des solutions institutionnelles et des infrastructures : présentation du pôle système Terre et des actions en cours (Dr. Jean-Christophe DESCONNETS, IRD & Damien BOULANGER, CNRS, OMP).

Pause déjeuner (12h-14h)

14h-16h - Table ronde de clôture animation par Damien BOULANGER & Dr. Jean-Christophe DESCONNETS, IRD. Grande salle.

Présentation du projet Européen ENVRI-FAIR (Damien BOULANGER, CNRS, OMP)

Table ronde sur les pôles de données et les infrastructures de l'environnement : enjeux et conséquences du programme Européen ENVRI-FAIR et des pôles de données sur la communauté nationale.

Damien BOULANGER, CNRS, OMP

Jean-Christophe DESCONNETS, IRD

Yvan LE BRAS, MNHN, UMS PatriNat

Documents associés et liens utiles

Pour les grands principes de l'Open Data et du FAIR Data Management

- la publication de référence du FAIR : <https://doi.org/10.1038/sdata.2016.18>
- le rapport de la Royal Society sur l'Open Data : <https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/>
- le [guide d'analyse du cadre juridique en France 2017](#)
- le site web de la [Research Data Alliance](#) dédié au partage des données sans barrières
- le site du Comité Science Ouverte (COSO) et leur base documentaire : <https://www.ouvrirelascience.fr/category/ressources/?type=0>

Pour la conception d'un plan de gestion de données FAIR

- DoRANum : <https://doranum.fr/>
-

Pour Thésaurus, ontologie et web sémantique

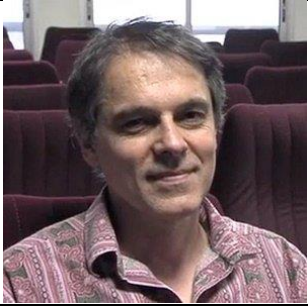
- Terminologie des langages documentaires et du web sémantique : http://rbdd.cnrs.fr/IMG/pdf/terminologie_des_langages_documentaires_et_du_web_s_emanique.pdf?429/5cd9b95a74a404b8adc7875d6452ba5c77361190

Espace de partage des ressources documentaires du cours : se connecter l'espace MyCore indiqué par les coordinateurs de l'Ecole thématique.

Description des enseignants

Dans l'ordre d'apparition lors de l'Ecole thématique

	Stéphanie RENNES Stéphanie Rennes est juriste, chargée des questions relatives à la science ouverte à la Direction des affaires juridiques de l'INRA. Doctorante au Bureau d'économie théorique et appliquée de l'Université de Strasbourg (UMR BETA 7522 CNRS, INRA UNISTRA, UL, AgroParisTech), ses recherches portent sur l'ouverture des données de recherche dans les partenariats science-industrie.
	Yvette LAFOSSE Titulaire d'une formation universitaire en sciences agronomiques et en IST, Yvette est chargée d'ingénierie et d'animation de formation au sein du service de Formation-DoRANum de l'Inist-CNRS. Elle est coordinatrice de l'évolution de la plateforme d'autoformation DoRANum.
	Coralie WYSOCZYNSKI Chargée d'ingénierie et d'animation de formation en IST à l'Inist-CNRS, Coralie travaille à l'élaboration de ressources et parcours pédagogiques, notamment pour la plateforme d'autoformation à la gestion et au partage des données de recherche DoRANum. Coralie développe une nouvelle génération de formations mixtes qui mêlent apprentissage numérique et d'autres modalités comme le jeu coopératif.
	Julien BARDE Docteur en sciences informatiques, Julien développe à l'IRD des systèmes d'information favorisant l'interopérabilité des jeux de données de biodiversité marine et de l'environnement afin d'améliorer la gestion des écosystèmes marins. Il participe à des projets Européens de l'IRD dans le domaine.
	Chloé MARTIN Ingénieure de recherche au CNRS, titulaire d'une thèse en archéozoologie, Chloé est affectée à l'UMS BBEES du CNRS et du MNHN où elle travaille comme experte base de données. Elle coordonne le réseau RBDD de la Mission pour les initiatives Transverses et Interdisciplinaires (MITI) du CNRS.

	Eric GARNIER Directeur de recherche au CNRS où il est un spécialiste internationalement reconnu de l'écologie fonctionnelle des plantes, Eric a largement participé au développement d'une éco-informatique relative aux traits des organismes et à des projets de synthèse exploitant ces ressources informatiques.
	Dominique VACHEZ Ingénieur de recherche à l'Inist-CNRS dans le service Ingénierie terminologique et spécialiste des systèmes d'information, Dominique est ingénieur agronome de formation. Il travaille au développement d'outils (Loterre) et référentiels sémantiques (Thésaurus de la biodiversité) et intervient dans le GDR SémanDiv en co-animant l'axe consacré à l'inventaire des ressources terminologiques pour les bases de données de biodiversité
	Clément JONQUET Maître de conférences à l'Université de Montpellier et titulaire d'une thèse de doctorat en informatique, Clément étudie les problèmes sémantiques et ontologiques posées par les bases de données d'agronomie et de biodiversité. Il a développé l'AgroPortal (https://cv.archives-ouvertes.fr/jonquet)
	Marie-Claude QUIDOZ Responsable technique de la plateforme « système d'information en écologie » du laboratoire CEFE du CNRS à Montpellier, Marie-Claude est architecte des systèmes d'informations. Elle développe des outils d'aide à la collecte et à la gestion des données pour l'écologie.
	Philippe Clastre Ingénieur d'étude en charge des systèmes d'informations pour l'unité URFM (écologie de la forêt méditerranéenne). Il gère la plateforme géomatique du centre PACA (GeOpen4S), et les outils sémantiques du projet AnaEE-France. Il intervient principalement sur le thésaurus anaeeThes.
	Danielle ZIEBELIN Professeur en informatique à l'Université Grenoble-Alpes, Danielle est spécialiste du web sémantique pour l'exploitation de données spatiales et temporelles. Elle est notamment experte de la modélisation des données environnementales et travaille sur des projets de recherche de caractérisation de l'Arc Alpin.

	Véronique CHAFFARD Ingénieure d'étude à l'IRD dans le LTHE, Véronique administre la base de données et le site web de l'observatoire AMMA-CATH au sein du réseau OZCAR. Elle est spécialiste des données spatialisées et environnementales.
	Damien MAURICE Ingénieur d'étude INRA dans une unité d'écologie forestière, Damien est responsable d'une plateforme de gestion des données spécialisée en administration de systèmes d'information et en analyse, modélisation et développement web. Il a contribué au développement des outils de sémantique et de modélisation ontologique de l'infrastructure nationale AnaEE France.
	Jean-Christophe DESCONNETS Ingénieur de recherche à l'IRD avec une double formation en informatique et hydrologie, Jean-Christophe est spécialiste des systèmes d'information spatialisés pour les infrastructures de recherche en environnement. Il pilote la mission infrastructures et données numériques au sein de l'IRD (https://cv.archives-ouvertes.fr/jean-christophe-desconnets).
	Damien BOULANGER Damien Boulanger est ingénieur de recherche et travaille au sein du Service de Données de l'Observatoire Midi-Pyrénées (SEDOO) à Toulouse. Il fait partie du bureau exécutif du Pôle de Données et Services Atmosphère AERIS. Il est responsable du centre de données de l'Infrastructure de Recherche Européenne IAGOS. Dans ce cadre il coordonne un des Work Package du projet ENVRI-FAIR.
	Yvan LE BRAS Yvan LE BRAS est ingénieur de recherche en charge du projet du Pôle national des Données de Biodiversité au MNHN dans l'Unité Patrinat. Docteur en écologie marine, il s'est spécialisé en bioinformatique au cours de son post-doctorat et dans un second temps sur le développement d'outils informatiques pour les sciences participatives.

Liste de contact des enseignants

Liste nominative des enseignants dans l'ordre des interventions

Stéphanie RENNES	Direction des affaires juridiques INRA, Paris	stephanie.rennes@inra.fr 01 42 75 90 58
Yvette LAFOSSE	Inist-CNRS, Nancy Département Valoriser les Données de Recherche, Service Formation- DoRANum	yvette.lafosse@inist.fr 03 83 50 94 04
Coralie WYSOCZYNSKI	Inist-CNRS, Nancy Département Valoriser les Données de Recherche, Service Formation- DoRANum	coralie.wysoczynski@inist.fr 03 83 50 46 00 (poste 7995)
Julien BARDE	UMR MARBEC - IRD Site de la Réunion	julien.barde@ird.fr
Chloé MARTIN	UMS 3568 BBEES CNRS-INEE, Paris	chloe.martin@mnhn.fr 01 40 79 80 42
Eric GARNIER	UMR CEFE – CNRS Montpellier	eric.garnier@cefe.cnrs.fr 0 4 67 61 32 42
Dominique VACHEZ	Inist-CNRS, Nancy Département Analyser et fouiller l'information scientifique	dominique.vachez@inist.fr 0383504600 (poste 7964)
Clément JONQUET	UMR LIRMM Université de Montpellier	jonquet@lirmm.fr 04 67 41 97 43
Marie-Claude QUIDOZ	UMR CEFE – CNRS Montpellier	Marie-Claude.Quidoz@cefe.cnrs.fr 04 67 61 32 39
Philippe CLASTRE	URFM-INRA Avignon	philippe.clastre@inra.fr 04 32 72 23 74
Danielle ZIEBELIN	UMR LIG Université Grenoble Alpes Grenoble	Danielle.Ziebelin@imag.fr 04 57 42 15 74
Véronique CHAFFARD	UMR IGE, OSUG-B IRD Saint Martin d'Hères	veronique.chaffard@univ-grenoble.alpes.fr 04 56 52 03 98
Damien Maurice	UMR EEF, INRA Nancy	damien.maurice@inra.fr 03 83 39 41 31
Jean-Christophe DESCONNETS	UMR ESPACE-DEV IRD, Montpellier	Jean-Christophe.Desconnets@ird.fr 04 67 55 86 33
Damien BOULANGER	UMS831 - OMP (CNRS-INSU) Toulouse	damien.boulanger@obs-mip.fr 05 61 33 27 71
Yvan LE BRAS	UMS2006 – Patrinat (MNHN) Concarneau	yvan.le-bras@mnhn.fr 02 98 50 99 35

Liste de contact des participants

Liste nominative des participants dans l'ordre alphabétique

Nom	Prénom	Employeur	Contact email
Abbassi	Sarra	CNRS	Sarra.Abbassi@univ-savoie.fr
Arnaud	Fanny	CNRS	fanny.arnaud@ens-lyon.fr
Boër	Michel	CNRS	michel.boer@anaee.eu
Braud	Isabelle	IRD	isabelle.braud@irstea.fr
Caesar	Maram	MNHN	maram.caesar@edu.mnhn.fr
Damy	Sylvie	Université Bourgogne Franche Comté	sylvie.damy@univ-fcomte.fr
Girard	Virginie	CNRS	virginie.girard@univ-grenoble-alpes.fr
Hajj-Hassan	Hicham	CNRS	hishamhh@cnrs.edu.lb
Heintz	Wilfried	INRA	wilfried.heintz@inra.fr
Henon	Amandine	MNHN	amandine.henon@mnhn.fr
Hoebeke	Mark	CNRS	mark.hoebeke@sb-roscoff.fr
Jacqueton	Céline	INRA	celine.jacqueton@inra.fr
Jean-Charles	Arnaud	CNRS	jeancharles@eccorev.fr
Joly	Dominique	CNRS	dominique.joly@cnrs-dir.fr
Kouadio	Sekedoua	IFSTTAR	jules.kouadio@ifsttar.fr
Lafont	Sébastien	INRA	sebastien.lafont@inra.fr
Le Galliard	Jean-François	CNRS	cereep@biologie.ens.fr
Le Moal	Francoise	CNRS	francoise.le-moal@univ-rennes1.fr
Lerigoleur	Emilie	CNRS	emilie.lerigoleur@univ-tlse2.fr
Liger	Lucie	CNRS	lucie.liger@irstea.fr
Martel	Julien	Inrap	julien.martel@irsn.fr
Massol	Florent	CNRS	florent.massol@ens.fr
Moulard	Nathalie	CNRS	nathalie.moulard@cnrs-orleans.fr
Ogereau	Francis	Université Clermont Auvergne	francis.ogereau@uca.fr
Ohanessian	Gilles	CNRS	gilles.ohanessian@polytechnique.edu
Pons	Marie-Noëlle	CNRS	marie-noelle.pons@univ-lorraine.fr
Ramet	Anaïs	CNRS	anaïs.ramet@cefe.cnrs.fr
Ravel	Olivier	CNRS	olivier.ravel@ecotron.cnrs.fr
Rouan	Mathias	CNRS	mathias.rouan@univ-brest.fr
Sarramia	David	Université Clermont Auvergne	david.sarramia@clermont.in2p3.fr
Souton	Corentin	CNRS	souton@biologie.ens.fr
Tarniewicz	Jérôme	Université Versailles Saint Quentin	jerome.tarniewicz@lsce.ipsl.fr
Vannière	Boris	CNRS	boris.vanniere@univ-fcomte.fr