

Projet FooSIN : tâche 3.1 Formations en relation avec les principes FAIR. Rapport des résultats de l'enquête

Sylvie Cocaud, Anne-Sophie Grenier, Dominique l'Hostis

► **To cite this version:**

Sylvie Cocaud, Anne-Sophie Grenier, Dominique l'Hostis. Projet FooSIN : tâche 3.1 Formations en relation avec les principes FAIR. Rapport des résultats de l'enquête. [Autre] INRAE. 2021, 16 p. hal-03260886

HAL Id: hal-03260886

<https://hal.inrae.fr/hal-03260886>

Submitted on 15 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Projet FooSIN : tâche 3.1

Formations en relation avec les principes FAIR. Rapport des résultats de l'enquête

Sylvie Cocard, Dominique L'Hostis, Anne-Sophie Grenier. INRAE, Direction pour la Science Ouverte (DipSO)- 2021

Rappel de la commande

La tâche 3.1 du WP Go Train du projet [FooSIN](#)¹ doit permettre d'identifier les **matériels et actions de formation existant dans les organisations partenaires** en rapport avec les principes FAIR, et de proposer des pistes d'amélioration apportées par **l'examen des actions de formation menées par des acteurs internationaux**.

1- Enquête

Un questionnaire d'enquête a été rédigé et diffusé à l'ensemble des partenaires du projet FooSIN pour identifier les formations entièrement dédiées aux principes FAIR, ou intégrées à des formations plus larges, en français ou anglais (voir questionnaire en annexe).

L'invitation à renseigner ce questionnaire a été diffusée une première fois en juin 2020 via une liste d'adresses de messagerie établie à partir des partenaires du projet, puis diffusée une seconde fois, via la liste de diffusion RDA France-FAIR à sa création en novembre 2020.

Dix-neuf enregistrements ont été saisis directement par des organisateurs de formations, dont 13 par des partenaires FooSIN.

Vingt-huit enregistrements ont été saisis par l'équipe projet. Ils décrivent des formations identifiées à partir de requêtes sur internet, ou de sites de référence sur la science ouverte.

Les formations axées sur l'analyse et la visualisation des données, ciblant un niveau expert (data scientists), n'ont pas été retenues.

Les analyses décrites dans ce rapport sont réalisées à partir de ces 47 enregistrements. Ce corpus représente un instantané non exhaustif des formations abordant les principes FAIR. Les données sont accessibles sur le portail data INRAE([10.15454/BP9PBL](#)).

¹ [FooSIN](#) (ANR-19-DATA-0019-01) est la participation française au GO FAIR Food Systems Implementation Network – FooSIN. Il vise à 1) accélérer la mise en œuvre des principes FAIR dans les communautés agriculture et alimentation, et 2) positionner la France comme un leader dans cette démarche en donnant de la visibilité aux actions et productions françaises au niveau international.

2- Résultats

2.1- Organismes des formations

Treize formations sont organisées par les partenaires du projet FooSIN :

- 2 par le CIRAD
- 3 par les différentes composantes de l'Université de Montpellier
- 8 par INRAE seul ou à travers des partenariats (IFB)

34 formations proviennent d'organisations externes :

Typologie - Organisateur	Pays	Organisation
Collectif recherche international	INT	CODATA
		Open Science MOOC
		DataONE
		FORCE11 Scholarly Communication Institute (FSCI)
		GODAN
		ESIP Federation
Collectif recherche européen	EUR	CESSDA
Projet européen	EUR	PARTHENOS
		FOSTER
Collectif recherche national	FRA	MSH Val de Loire
	NLD	Research Data Netherlands
Infrastructure de recherche	FRA	ANAEE-France
Organisme de recherche	FRA	INIST-CNRS
		INRIA
Université	DNK	Denmark Universities
	FRA	Université de Lorraine
		Université de Paris
	INT	The University of North Carolina at Chapel Hill ; The University of Edinburgh
	NLD	Delft University of Technology
		Maastricht University

	USA	Vanderbilt University
		Institute of Museum and Library Science (IMLS)
		Johns Hopkins University
		Network of the National Library of Medicine (NNLM)
		North Carolina State University
		University of Washington

2.2- Formations génériques, formations spécialisées

Les formations génériques représentent 36 des 47 enregistrements (soit 77%). Seuls 23% des formations sont orientées vers un domaine spécialisé. On retrouve ces pourcentages dans les deux groupes de formations : FooSIN et externe.

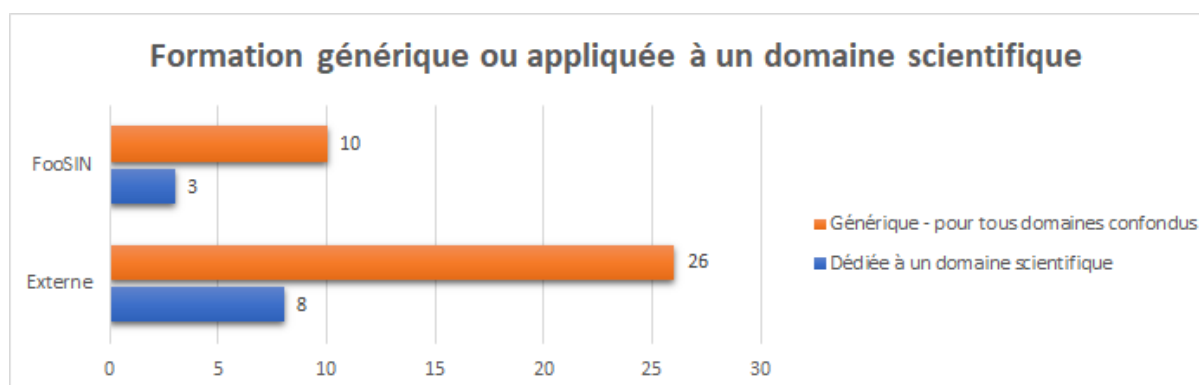


Figure 1 : Type de formation organisée

Formations appliquées à un domaine scientifique, tel qu'indiqué par les répondants :

- FooSIN : Sciences de la vie et de la santé, Bioinformatique, Sciences Animales (bovins)
- Externe : Sciences Biomédicales, Sciences environnementales, Agriculture et nutrition, Sciences de l'Ingénieur, Recherche clinique, Sciences humaines et sociales

2.3- Modalités de formation

- 16 formations en présentiel
- 21 formations en ligne et Mooc
- 4 classes virtuelles
- Autres : 2 guides en ligne, 2 multimodales, 2 au choix en présentiel ou en ligne

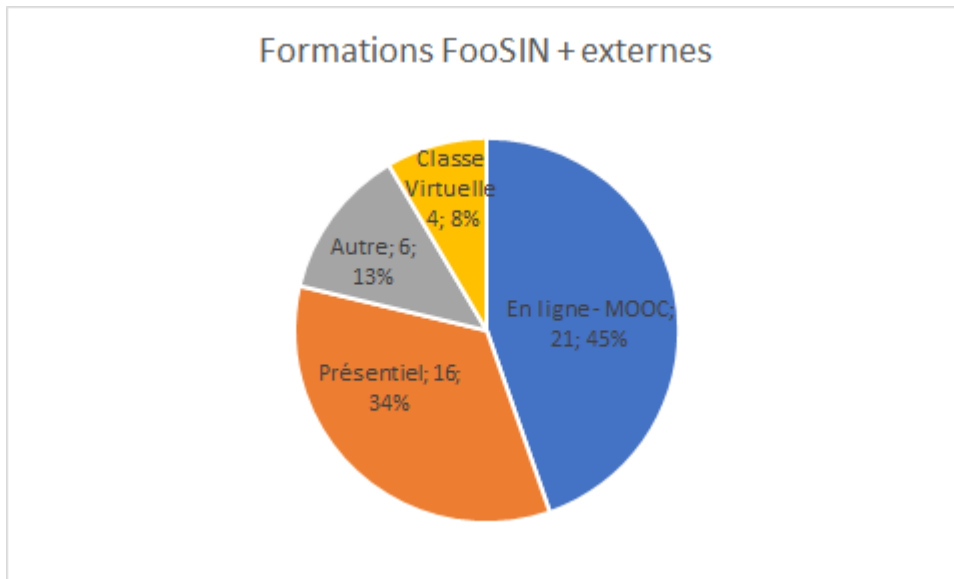


Figure 2 : Modalités des formations organisées

2.4- Durée des formations

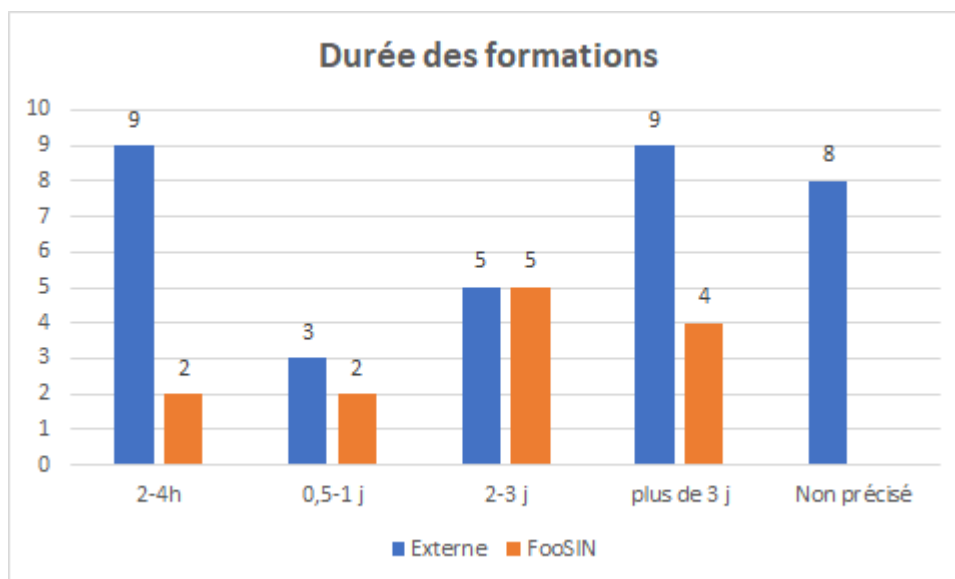


Figure 3 : Durée des formations organisées

Les formations organisées par les partenaires FooSIN sont plutôt des formations longues (2 jours ou plus).

Formations courtes

L'échantillon intègre 11 formations courtes (2-4h) qui sont soit des séances introductives génériques, soit des séances spécialisées sur un point ou un outil.

Formations longues

Vingt-trois formations dépassent une durée de 2 jours : pour des publics avertis (7), débutants (11), tous niveaux (5). Huit délivrent une attestation de type badge ou certificat.

Ces formations longues sont de typologie diverse et assez difficiles à caractériser. Nous distinguons cependant les formations de type :

- **MOOC** avec interaction entre apprenants et enseignants :
 - [Recherche reproductible : principes méthodologiques pour une science transparente](#)
 - [Open Science: Sharing Your Research with the World](#)
 - [Gestion des données pour la recherche clinique](#)
- **Auto-formations** comme le site [Doranum](#) ou [Open Science MOOC](#)
- **Formations universitaires diplômantes** :
 - [DU Scientific Data Management](#) (SDM)
 - [Junior Research Lab](#)

2.5- Niveau pré-requis pour les formations

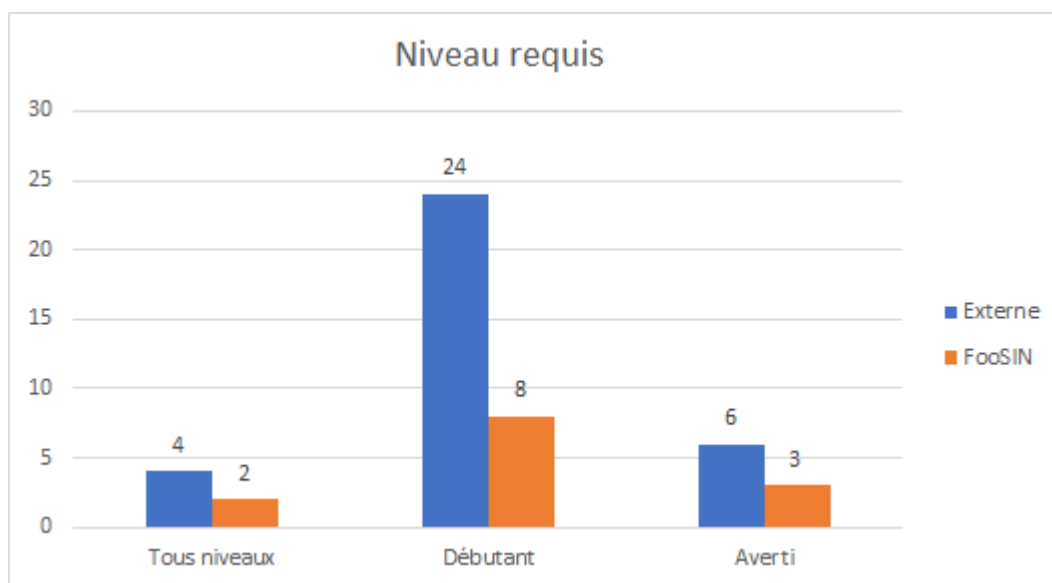


Figure 4 : Niveau pré-requis par les formations

La majorité des formations cible un public débutant et n'exige aucun pré-requis. Neuf formations (soit 19% seulement) s'adressent à un public averti :

- 3 sont dédiées à un domaine scientifique : Bioinformatique, Sciences Biomédicales, Sciences de la vie et de la santé, 6 sont génériques
- 1 est de courte durée, les 8 autres sont des formations longues de plus de 2 jours

2.6- Langue

Rappel : seules les formations en anglais et/ou en français ont été prises en compte.

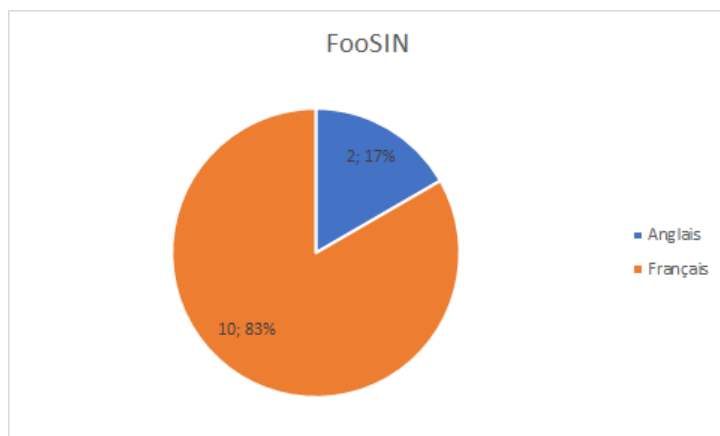


Figure 5 : Langue des formations FooSIN

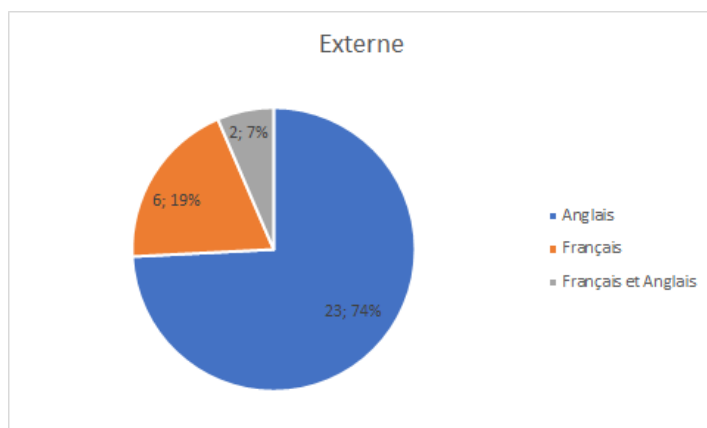


Figure 6 : Langue des formations externes

2.7- Sujets abordés dans les formations

Une liste de sujets a été créée pour identifier le contenu des formations et faciliter la saisie dans le formulaire à renseigner. Un champ libre permettait aux répondants d'indiquer d'éventuels sujets complémentaires, mais la plupart d'entre eux étaient déjà dans la liste, en particulier les outils liés à la reproductibilité des données.

Liste complète : voir le [questionnaire en annexe](#).

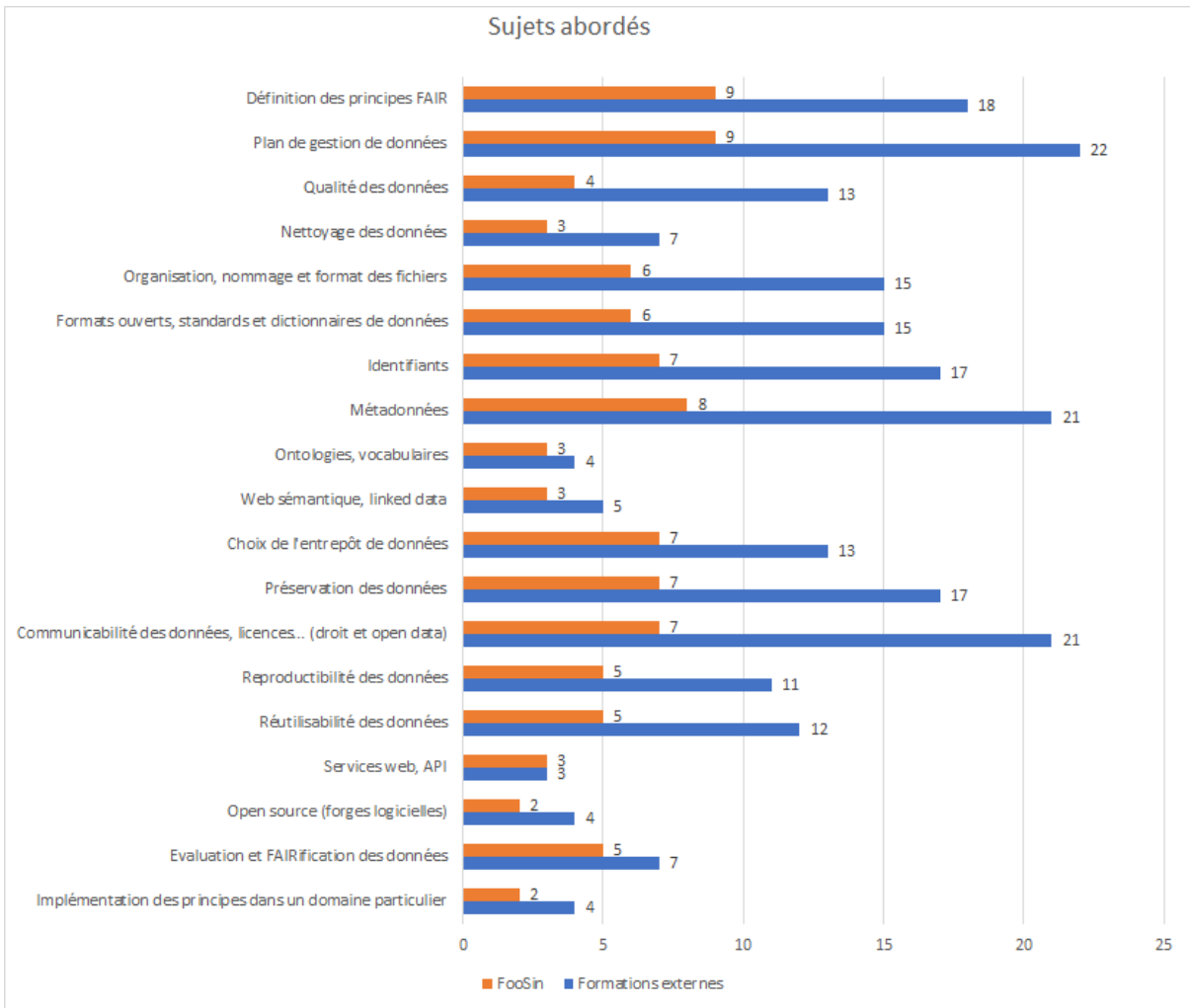


Figure 7 : Sujets abordés dans les formations

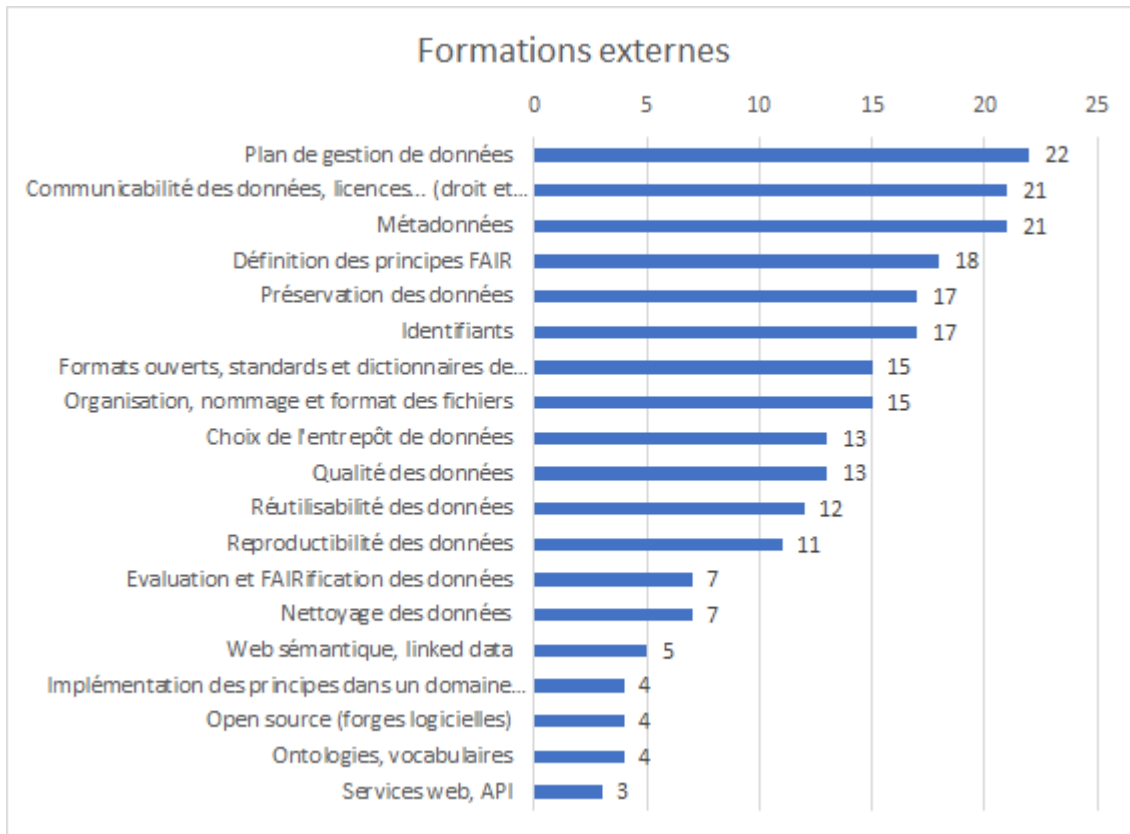


Figure 8 : Les sujets des formations externes par nombre d'occurrences décroissant

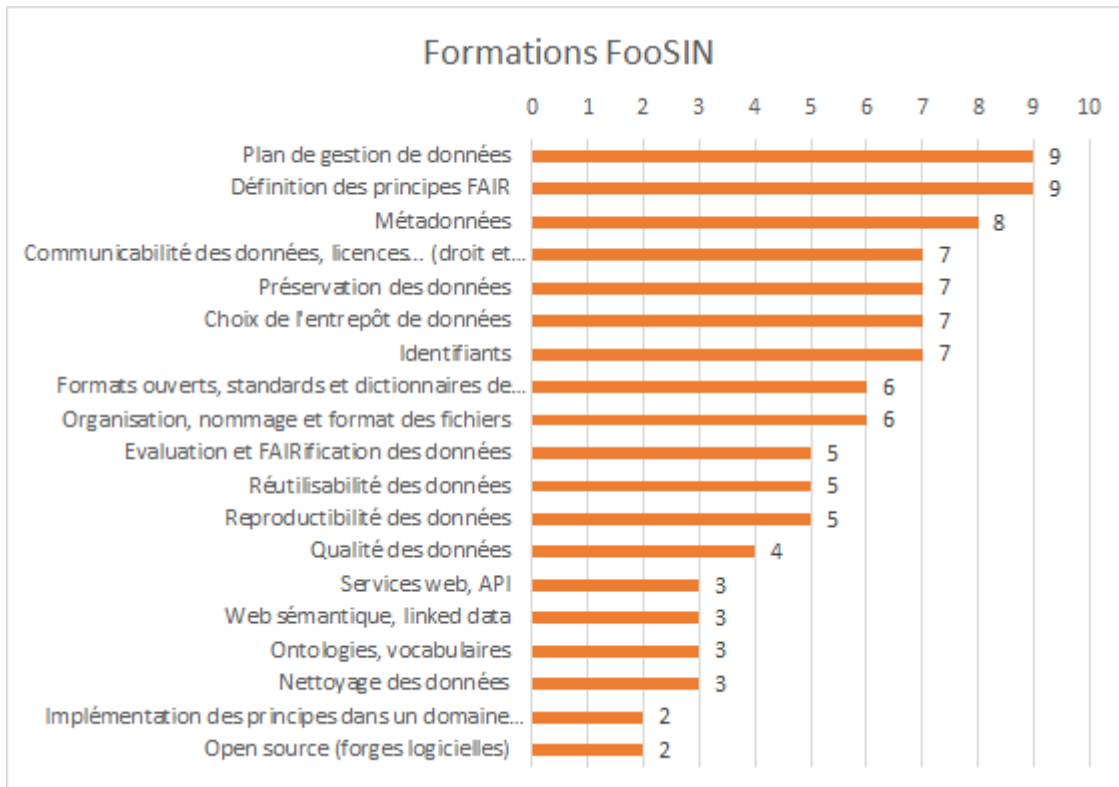


Figure 9 : Les sujets des formations FooSIN par nombre d'occurrences décroissant

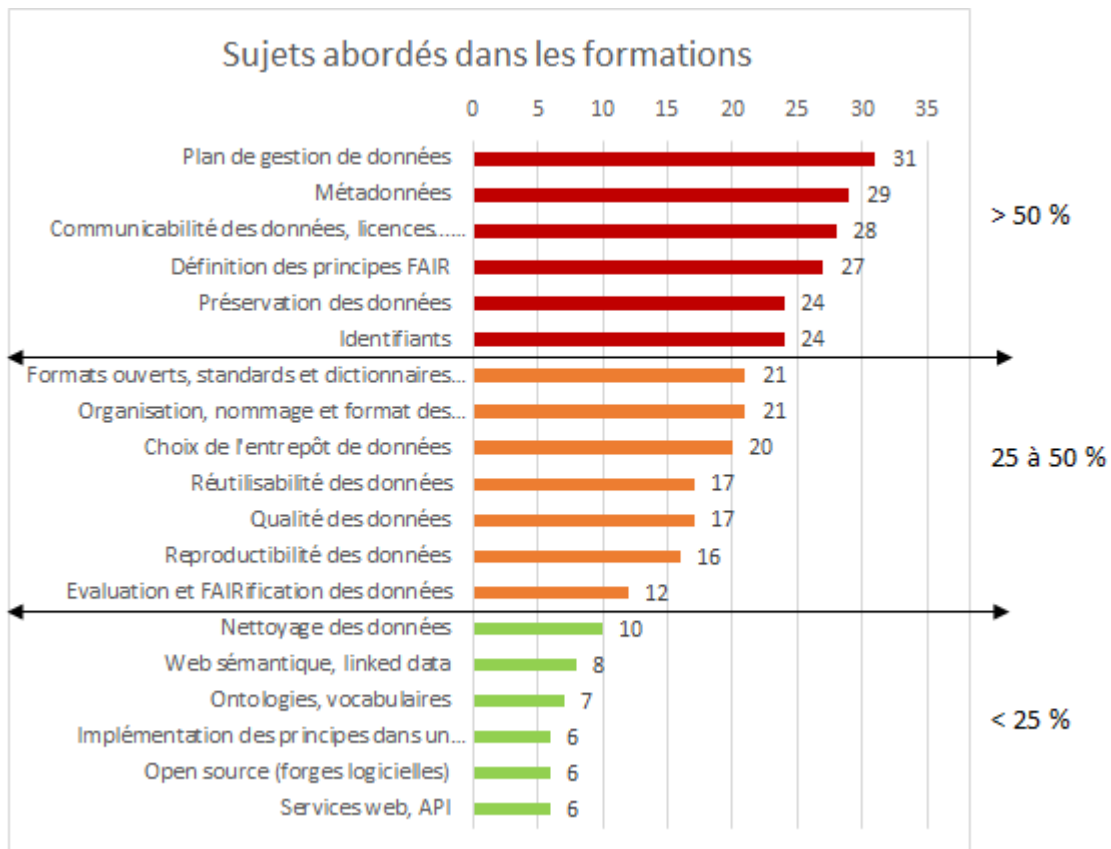


Figure 10 : Classement des sujets selon leur nombre d'occurrences sur l'ensemble des formations

Plus de la moitié des formations (FooSIN comme externes), abordent les sujets suivants, à considérer comme des contenus incontournables.

- Plan de gestion de données,
- Métadonnées,
- Communicabilité des données, licences... (droit et open data),
- Définition des principes FAIR,
- Préservation des données,
- Identifiants.

A l'inverse, les sujets les moins souvent abordés et toujours intégrés à des formations plus larges sont :

- Nettoyage des données,
- Web sémantique, linked data,
- Ontologies, vocabulaires,
- Implémentation des principes dans un domaine particulier,
- Open source (forges logicielles),
- Services web, API.

2.8- Type de contenu

Les formations FooSIN semblent donner une plus grande part aux “ateliers pratiques” et “retours d’expériences” que les formations externes.

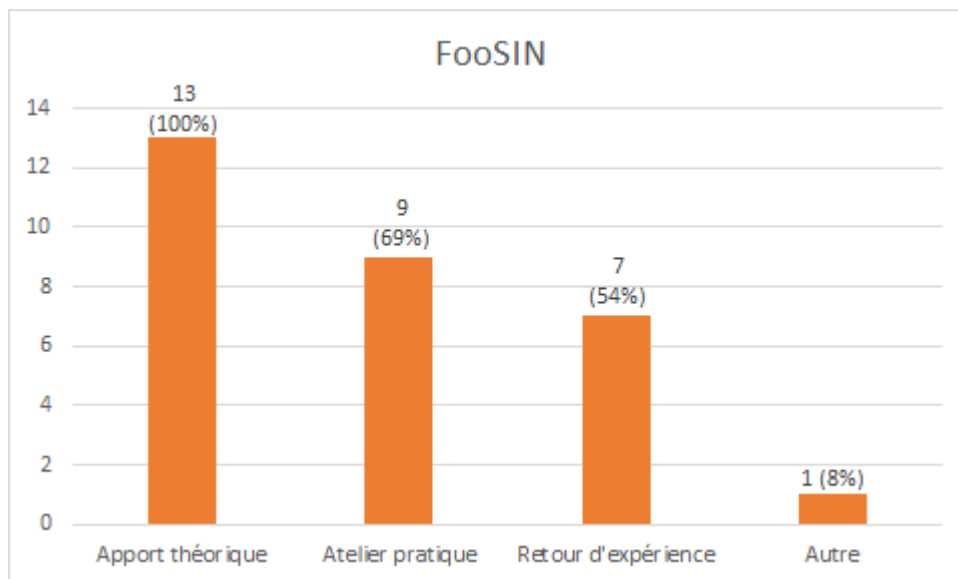


Figure 11 : Type de contenu des formations FooSIN

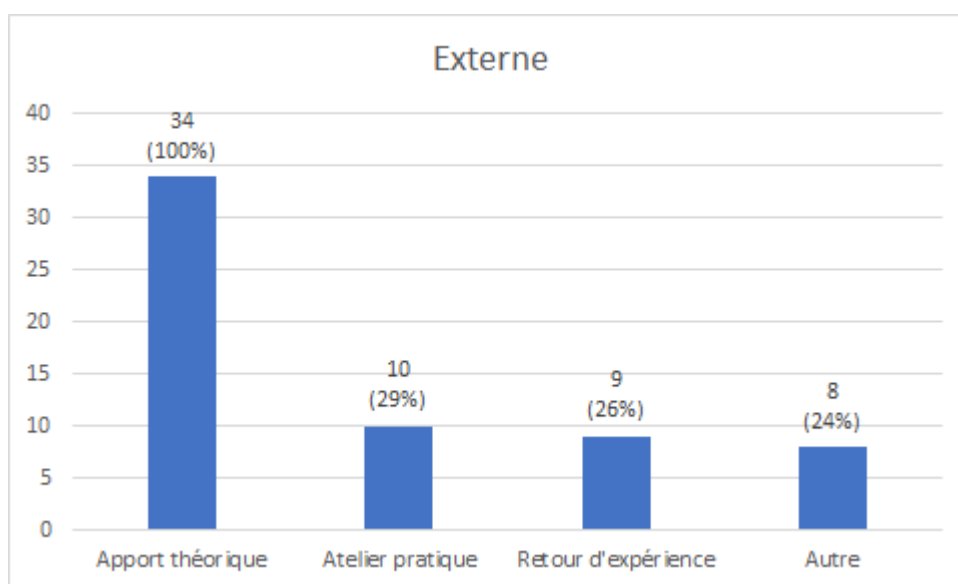


Figure 12 : Type de contenu des formations externes

La catégorie “Autre” a été utilisée pour 9 formations. Elle représente des exercices, des quiz, du tutorat, et des échanges.

Conclusion

L'enquête réalisée entre le 3/06/2020 et le 25/02/2021 porte sur un échantillon très modeste (47 réponses) et peut comporter des biais :

- Les enregistrements saisis par l'équipe projet provenant de sites web n'ont pas toujours permis de renseigner de façon précise les informations attendues dans le questionnaire.
- Différencier de façon claire les formations réelles de sites internet, de conférences ou guides d'information s'est révélé souvent difficile. Nous avons par exemple choisi de ne pas intégrer les Journées Nationales Science Ouverte (JNSO - COSO) et les "Swiss Research Data Day" dans l'enquête.
- Choix d'écartier a priori les formations orientées analyse et visualisation des données, ciblant un niveau expert (data scientists)

Les résultats ne révèlent pas de différences notables entre les formations FooSIN et les formations externes sur les différents critères examinés. L'enquête met en évidence les sujets moins abordés et qui mériteraient d'être intégrés à de futures formations, en particulier l'évaluation du niveau FAIR et la FAIRification des données (les principes FAIR étant présentés le plus souvent de façon théorique), la reproductibilité, la qualité et la réutilisation des données.

Annexe : questionnaire d'enquête

Formations FAIR

Les objectifs du [projet ANR FooSIN](#) sont de :

- Développer la culture FAIR et d'encourager sa mise en oeuvre,
- Prendre en compte les besoins et pratiques des acteurs français de l'agriculture et de l'agro-alimentaire,
- Encourager les approches sémantiques et linked open data.

Le projet prévoit la réalisation d'un état des lieux des formations en relation avec les principes FAIR, entièrement dédiées à ces principes ou intégrées dans des formations plus larges.

Ce questionnaire a pour objectifs de permettre à tout organisateur de signaler une formation répondant à ces critères.



Merci de renseigner le questionnaire pour chaque formation signalée.

Il y a 23 questions dans ce questionnaire

INFORMATIONS GENERALES

[[Titre de la formation : *

Veillez écrire votre réponse ici :

[[Description (mots-clés ou texte court) : *

Veillez écrire votre réponse ici :

[[Durée de la formation (en heures) : *

Veillez écrire votre réponse ici :

[[Date de création de la formation *

Veillez entrer une date :

[]Organisme formateur *

Veillez écrire votre réponse ici :

[]Lien vers le programme ou une présentation de la formation :

Veillez écrire votre réponse ici :

CONTENU DE LA FORMATION

[]

Sujets abordés au cours de la formation ?

*

Veillez choisir toutes les réponses qui conviennent :

- Définition des principes FAIR
- Plan de gestion de données
- Qualité des données
- Nettoyage des données
- Organisation, nommage et format des fichiers
- Formats ouverts, standards et dictionnaires de données
- Identifiants
- Métadonnées
- Ontologies, vocabulaires
- Web sémantique, linked data
- Choix de l'entrepôt de données
- Préservation des données
- Communicabilité des données, licences... (droit et open data)
- Reproductibilité des données
- Réutilisabilité des données
- Services web, API
- Open source (forges logicielles)
- Evaluation et FAIRification des données
- Implémentation des principes dans un domaine particulier
- Autre:

Plusieurs réponses possibles

PUBLIC DE LA FORMATION

[[Formation générique ou appliquée à un domaine scientifique *

Veillez sélectionner une seule des propositions suivantes :

- Générique - pour tous domaines confondus
- Dédiée à un domaine scientifique

[[Merci de préciser lequel :

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse était 'Dédiée à un domaine scientifique' à la question '8 [Q9]' (Formation générique ou appliquée à un domaine scientifique)

Veillez écrire votre réponse ici :

[[Profil professionnel ciblé *

Veillez choisir toutes les réponses qui conviennent :

- Technicien et ingénieur de recherche
- Chercheur, enseignant-chercheur, post-doctorant
- Etudiant, doctorant
- Documentaliste, data Librarian (spécialisé / données)
- Informaticien
- Bioinformaticien
- Grand Public

Plusieurs réponses possibles. SVP, à renseigner selon son métier, et non son statut

[[Niveau pré-requis *

Veillez sélectionner une seule des propositions suivantes :

- Débutant
- Averti
- Expert
- Tous niveaux

FORMATEURS

[[Animation

	Intervenant interne	Intervenant externe
Informaticien	<input type="checkbox"/>	<input type="checkbox"/>
Scientifique	<input type="checkbox"/>	<input type="checkbox"/>
Documentaliste	<input type="checkbox"/>	<input type="checkbox"/>
Juriste	<input type="checkbox"/>	<input type="checkbox"/>
Autre	<input type="checkbox"/>	<input type="checkbox"/>

Plusieurs réponses possibles

[[Vous avez sélectionné la réponse "Autre" à la question précédente (intervenant interne). Merci de préciser ci-dessous :

Répondre à cette question seulement si les conditions suivantes sont réunies :
La réponse était '1' à la question '12 [Q12]' (Animation)

Veillez écrire votre réponse ici :

**[[Vous avez répondu "Autre" à la question précédente (intervenant externe).
Merci de préciser :**

Répondre à cette question seulement si les conditions suivantes sont réunies :
La réponse était '1' à la question '12 [Q12]' (Animation)

Veillez écrire votre réponse ici :

AUTRES CARACTERISTIQUES

[[Nombre de participants *

Veillez écrire votre réponse ici :

Merci d'indiquer seulement un nombre ou une tranche A-B

[[Rythme de la formation *

Veillez écrire votre réponse ici :

Précisez par exemple : le nombre de séances par mois, année ... ou "sur demande" ...

[[Type de formation *

Veillez sélectionner une seule des propositions suivantes :

- Présentiel
- En ligne - MOOC
- En ligne - SPOC
- Classe virtuelle
- Serious Game
- Autre

Faites le commentaire de votre choix ici :

[[Type de contenus *

Veillez choisir toutes les réponses qui conviennent :

- Apports théoriques
- Atelier pratique
- Retour d'expérience
- Autre:

[[Langue *

Veillez sélectionner une seule des propositions suivantes :

- Français
- Anglais
- Français et Anglais

Sylvie COCAUD (sylvie.cocaud@inrae.fr)

Anne-Sophie GRENIER (anne-sophie.grenier@inrae.fr)

Dominique L'HOSTIS (dominique.lhostis@inrae.fr)

15/02/2021 – 10:18