

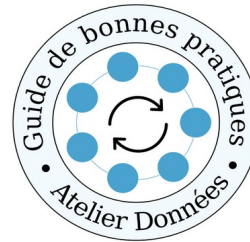
# *Guide de bonnes pratiques sur la gestion des données de la recherche*

---

---

**Webinaire SIST  
Juin 2021**

<https://mi-gt-donnees.miti.cnrs.fr/>



---

---

*Vous êtes invités à poser vos questions sur le chat, nous y répondrons à la fin de l'exposé*

# Genèse du GT Atelier Données

---

- Des initiatives du réseau Renatis
  - ✓ Identifié une thématique commune à plusieurs réseaux : la donnée
- Rencontres des réseaux professionnels du CNRS (13-14/01/2016)
  - ✓ Atelier : le jeu « inter-réseaux » :  
la donnée de l'acquisition à la valorisation»
  - ✓ 19 inscrits
  - ✓ Restitution : [https://webcast.in2p3.fr/video/restitution\\_des\\_ateliers](https://webcast.in2p3.fr/video/restitution_des_ateliers)
  - ✓ De nombreux besoins recensés
- Proposition de création du GT « Atelier Données »
  - ✓ Mettre en place une journée thématique sur l'interopérabilité
  - ✓ Rédiger un guide pratique sur la traçabilité des activités de recherche

# Lettre de mission du GT Atelier Données de la MITI

---

- Construire et diffuser une vision transversale de la gestion des données
  - ✓ enrichir la pratique de chaque réseau dans le domaine des données et permettre le développement de la complémentarité entre réseaux
- Valoriser l'apport des expériences et expertises « métier » dans une vision transversale de gestion de données
- Sensibiliser les communautés professionnelles de l'appui à la gestion des données (organisation de journées thématiques par exemple) ;
- Identifier les problématiques concernant les « data » dans chaque réseau (livrables à définir).
- Mettre en commun et partager de nouvelles pratiques en réseau et au sein de chaque réseau.

# GT Atelier Données aujourd'hui

---

- 18 membres
- Deux duo d'animateurs :
  - ✓ Emmanuelle Morlock & Caroline Martin (Janv. 2016 – Juin 2019)
  - ✓ Pierre Brochard & Marie-Claude Quido (Juillet 2019 – Juin 2021)
- Différentes « structures »
  - ✓ Réseaux MITI : Calcul, Devlog , Medici, QeR, rBDD, Renatis, Resinfo
  - ✓ Réseaux Institut : SIST
  - ✓ Autres réseaux : Relier
  - ✓ Structures : DDOR, INIST
- Des moyens de communication
  - ✓ <https://gt-atelier-donnees.miti.cnrs.fr/>
  - ✓ [donnees-inter-reseaux@services.cnrs.fr](mailto:donnees-inter-reseaux@services.cnrs.fr)
  - ✓ [https://twitter.com/miti\\_gt\\_donnees](https://twitter.com/miti_gt_donnees) / [https://mamot.fr/@miti\\_gt\\_donnees](https://mamot.fr/@miti_gt_donnees)
  - ✓ Lettre d'information mensuelle

# Actions réalisées (1/2)

---

- « Interopérabilité et pérennisation des données de la recherche. Comment FAIR en pratique ? Retours expériences »
  - ✓ Séminaire, Paris, Novembre 2018, 150 personnes
  - ✓ <https://gt-donnees2018.sciencesconf.org/>
- « Comment améliorer le dépôt de données de recherche »
  - ✓ Hackathon, Grenoble, Janvier 2020, 50 personnes
- « Data paper, une incitation à la qualification et à la réutilisation des jeux de données »
  - ✓ Webinaire, Novembre 2020, 150 personnes connectées
  - ✓ [https://www.canal-u.tv/video/medici/introduction\\_du\\_webinaire.58157](https://www.canal-u.tv/video/medici/introduction_du_webinaire.58157)
- Foire Aux Questions
  - ✓ 24 entrées sur 2 grands thèmes : entrepôts de données et les data papers
  - ✓ <https://mi-gt-donnees.pages.math.unistra.fr/site/FAQ.html>

# Actions réalisées (2/2)

---

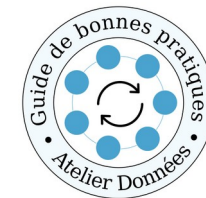
- « Cartographie des actions des réseaux métiers autour de la gestion des données »
  - ✓ Mai 2017
  - ✓ <https://mi-gt-donnees.pages.math.unistra.fr/site/download/GTInterreseaux-CartoSyntheseV6-optimize.pdf>
- « Guide de bonnes pratiques sur la gestion des données de la recherche »
  - ✓ Juillet 2019 -> Février 2021
  - ✓ <https://mi-gt-donnees.pages.math.unistra.fr/guide/00-introduction.html>
- *Save the date : 5 Juillet 2021*
  - ✓ Webinaire sur la "qualité des données"

---

**« Guide de bonnes pratiques sur la  
gestion des données de la recherche »**

<https://mi-gt-donnees.pages.math.unistra.fr/guide/00-introduction.html>

# Originalité



L'originalité de ce guide réside dans son application aux données de la recherche **sous l'angle des pratiques de différents métiers de la recherche :**

- production de quelques réseaux de l'Atelier Données de la MITI:
  - ✓ *DDOR, Calcul, RBDD, QeR et SIST*
  - ✓



## Auteurs

Ce guide de bonnes pratiques sur la gestion des données dans les réseaux métiers, a été réalisé par:

- Christine Hadrossek : [DDOR, réseau Renatis](#)
- Joanna Janik : [DDOR, réseau Renatis](#)
- Maurice Libes : [réseau SIST](#)
- Violaine Louvet : [réseau Calcul](#)
- Marie-Claude Quidoz : [réseau rBDD](#)
- Alain Rivet : [réseau QeR](#)
- Geneviève Romier : [réseau rBDD](#)

## Relecteurs

- Pierre Brochard : [réseau DevLog](#)
- Dominique Desbois : [réseau DevLog](#)
- Emilie Lerigoleur : [réseau SIST](#)
- Caroline Martin
- Pierre Navaro : [réseau Calcul](#)



# Originalité



- Il fournit une vision *transversale* à travers une compilation de diverses pratiques métiers dans la gestion des données.
- Il présente :
  - ✓ les nombreuses **actions de formation** ou de sensibilisation des réseaux ;
  - ✓ les compétences et expertises développées issues de **pratiques standardisées** qui font leurs preuves sur le terrain ;
  - ✓ des recommandations et des solutions techniques et organisationnelles grâce à la **veille technologique et juridique** réalisée très régulièrement.
- Il traduit le soutien mis en place par les membres des réseaux métiers, dans la gestion et la valorisation des données scientifiques.

# Une initiative des réseaux MITI

---

Ce Guide valorise nos réseaux et nos activités car il traduit les efforts et le soutien des membres des réseaux dans la gestion et la valorisation des données et le soutien à la communauté scientifique.

- réflexion et mise à disposition des *outils, méthodes et infrastructures* en matière de gestion et de partage des données scientifiques,
- préconisation et mise en place des bonnes pratiques de gestion des données,
- organisation de formations (ANF), séminaires, et journées d'études.

**A travers ces actions nous témoignons des activités de soutien des réseaux, en fournissant les meilleures pratiques du moment en matière de gestion des données.**

# Le guide est dans le "move"

---

Ces initiatives et actions des réseaux concordent avec les initiatives nationales :

- le “Plan National pour la Science Ouverte” (2018),
- le Noeud National RDA-France (2018),
- La Feuille de route du CNRS (2019),
- le “Plan Données de la recherche du CNRS » (2020) qui fournissent les orientations en matière de gestion “FAIR” et d’ouverture des données.



# Contenu : cycle de vie des données

Pour adopter un **point de vue commun** aux différents métiers et activités de nos réseaux :

- Nous nous basons sur le **cycle de vie des données** :
- Les étapes du cycle de vie des données donnent un **cadre structurant** et un **vocabulaire commun**

Le guide fournit une **lecture nouvelle** des actions des réseaux, enrichie des approches complémentaires des pratiques des différents réseaux



# Le contenu du guide



Rechercher dans ce livre ...

- 1. Imaginer et préparer
- 2. Concevoir et planifier
- 3. Collecter**
- 4. Traiter
- 5. Analyser
- 6. Préserver et archiver
- 7. Publier et diffuser
- Conclusion
- Glossaire
- Infrastructures
- Reproductibilité
- Autres guides de bonnes pratiques
- Crédits
- Document pdf

[contact](#)



Contenu

## 3.1. Utiliser des normes et des standards d'interopérabilité

- 3.1.1. Les standards de métadonnées
- 3.1.2. Les référentiels de métadonnées

## 3.1.3. Les protocoles standards en information Géographique

- 3.2. Les systèmes d'acquisition : maîtriser l'acquisition et la collecte des données
- 3.3. Environnements de stockage - Sauvegarder les données

### ↳ [Les Normes OGC \(Open Geospatial Consortium\)](#)

François André, Aeris  
Séminaire SIST15 - OSU Pytheas Marseille 2015

Parmi les standards de l'OGC les plus utilisés dans nos réseaux métiers chez les gestionnaires de données environnementales, on peut citer :

- **CS-W** - Catalog Service for the Web : ce protocole est destiné à diffuser des métadonnées ISO 19139 et permettre l'interrogation de catalogues de métadonnées. Une très bonne implémentation de ce protocole est réalisée dans le logiciel "Geonetwork" utilisé pour constituer des catalogues et des inventaires de jeux de données et les présenter sur le Web de manière interopérable. Ce logiciel est détaillé dans le chapitre 7 "Publier" du présent guide, dédié à la publication des jeux de données. Grâce à ce protocole, on peut constituer des réseaux de catalogues tels que demandés par la [Directive Européenne Inspire](#).
- **WMS** - [Web Map Service](#) est un protocole de communication standard qui permet de constituer des cartes de données géoréférencées à partir de différents serveurs de données cartographiques. .

Le réseau **SIST** a organisé deux actions de formation nationale (ANF) sur ces logiciels mettant en oeuvre les standards d'interopérabilité WMS, CSW et SOS qui permettent aux personnels d'améliorer la gestion et la diffusion de leurs données scientifiques d'observation en apprenant à installer, configurer et utiliser différents outils logiciels, choisis pour leur aptitude à répondre de manière standardisée à ces problématiques.

### ↳ ["Gestion des données d'observation : les outils informatiques pour la valorisation"](#)

ANF SIST17, Fréjus - ANF SIST18, Autrans

De nombreux instituts et auteurs, gestionnaires de données suivent ces standards OGC :

Sylvain Grelet communique par exemple le retour d'expérience sur l'utilisation et le déploiement des standards d'interopérabilité au BRGM :

# Et SIST là dedans ?

---

- La gestion des données environnementales est le coeur de métier de SIST
- Il m'a semblé important de contribuer à ce Guide au titre de SIST
  - ✓ Objectif : *rassembler la "majeure partie" des pratiques que nous utilisons et avons traitées dans les séminaires et formations SIST*
- La juxtaposition avec les connaissances et pratiques d'autres réseaux (DDOR, RbDD, QeR, Calcul)
  - ✓ Représente un Intérêt original et une Incitation à le lire et le diffuser auprès des chercheurs
  - ✓ De nouvelles sources de savoir pour SIST ? :
    - ▶ *la reproductibilité de la science,*
    - ▶ *les infrastructures*
    - ▶ *Les phases de préparation de projet*

# Recherche

Searching for multiple words only shows matches that contain all words.



## Résultats de la recherche

La recherche est finie, 9 page(s) trouvée(s) qui corresponde(nt) à la recherche.



🔍 SIST

- 1. Imaginer et préparer
- 2. Concevoir et planifier
- 3. Collecter
- 4. Traiter
- 5. Analyser
- 6. Préserver et archiver
- 7. Publier et diffuser
- Conclusion
- Glossaire
- Infrastructures
- Reproductibilité
- Autres guides de bonnes pratiques
- Crédits
- Document pdf [📄](#)

[contact](#)

- [7. Publier et diffuser](#)  
...oréférencé. L'utilisation des logiciels GeoNetwork et GeoServer a été traitée par une action de formation ANF du réseau SIST pour laquelle on trouvera les informations sur les pages formation du site SIST Documentations sur les logiciels é...
- [1. Imaginer et préparer](#)  
...ultats d'une enquête réalisée auprès des directeurs d'unité pour aborder la question du point de vue du chercheur et insiste sur la nécessité de réinventer nos métiers et de s'approprier la gestion des données. Gestion des données de la r...
- [2. Concevoir et planifier](#)  
...plan ? Plan de gestion de données ? DMP OPIDoR vous guide ! Laurent RASSINOX, Marie-Christine JACQUEMOT-PERBAL, INIST, SIST 2018 : Séries Interopérables et Systèmes de Traitement, Guyancourt, 2018 Des discussions, lors des Journées Calcul...
- [3. Collecter](#)  
...ou systèmes existants ou futurs, et ce sans restriction d'accès ou de mise en œuvre". Développer l'interopérabilité consiste donc à mettre en place et utiliser des normes et des standards qui fixent des règles permettant d'assurer le bon f...
- [4. Traiter](#)  
...interface de programmation (API) en Python pour créer des fichiers NetCDF par programme ont été présentés au séminaire SIST19 à l'OMP de Toulouse, par Joël Sudre, Maurice Libes et Didier Mallarino : Présentation du format NetCDF Joël Sudr...
- [5. Analyser](#)  
...ions le filtrage les agrégations le stockage en base de données Plusieurs présentations issues des journées du réseau SIST illustrent des mises en oeuvre de chaîne logicielle d'analyse de données: Filtrage interactif de données multidime...
- [6. Préserver et archiver](#)  
...sés, nous vous proposons les définitions suivantes. 6.1.1. Définitions générales StockerC'est l'étape première qui consiste à déposer les données sur un support numérique pour les rendre accessibles. Cela peut être un ordinateur personnel...
- [Crédits](#)  
...été réalisé par: Christine Hadrossek : DDOR, réseau Renatis Joanna Janik : DDOR, réseau Renatis Maurice Libes : réseau SIST Violaine Louvet : réseau Calcul Marie-Claude Quidoz : réseau rBDD Alain Rivet : réseau QeR Geneviève Romier : résea...
- [Guide de bonnes pratiques sur la gestion des données de la recherche](#)  
...groupe composé de plusieurs réseaux métiers de la MITI (Calcul, Devlog, QeR, rBDD, Renatis, Resinfo, Medici), du réseau SIST, (labellisé par l'INSU et regroupant les gestionnaires de données environnementales), de l'INIST, et de la Directio...

# Et SIST là dedans ?

---

- ✓ Les Thèmes traités par SIST apparaissent plus particulièrement
  - ▶ *Chapitre 3 : Collecter*
  - ▶ *Chapitre 7: Diffuser*
- ✓ Pratiques et outils d'interopérabilités OGC
- ✓ Diffusion de données par CSW, WMS, DAP, SOS
  - ▶ *Geonetwork, geoCMS, GeoServer, Thredds, Erddap,*
  - ▶ *IstSOS.... DMP, DOI... pôles de données*
- Pour autant le Guide *n'est pas exhaustif et un certain nombre de thèmes manquent, et devraient être ajoutés ou développés.*
  - ✓ *Méthodes et outils de Qualité et qualification des données*
  - ✓ *Utilisation des thesaurus et vocabulaires contrôlés etc.*
  - ✓ *Nouveaux Outils, développements*
  - ✓ *etc.*



# Et SIST là dedans ?

---

- *SIST19 Toulouse : Atelier / table ronde : Bonnes pratiques pour la gestion des données d'observation*
  - ✓ *Réflexions et élaboration d'un mini guide de recommandations publié sur le site web SIST*
    - ▶ <https://sist.cnrs.fr/ressources/guide-de-bonnes-pratiques/guide-de-bonnes-pratiques>
- **Traitement des données (chapitre 4)**
  - ▶ *Qualité, qualification, correction, versionnement, ...*
- **Pérennité des données (chapitre 6)**
  - ✓ *Comment rendre les données utilisables dans le futur*
- **Accessibilité des données (chapitre 7)**
  - ✓ *Comment rendre les données trouvables, identifiables, accessibles,*
- **Ré-utilisabilité, interopérabilité (chapitre 7)**
  - ✓ *Comment faire en sorte que les données soient interopérables entre elles et avec les services*



# Appel à contribution



- Ce guide n'est pas exhaustif puisqu'il est le reflet des thèmes abordés et actions des réseaux impliqués dans la rédaction du guide.
- Il demande à évoluer et à être enrichi de nouveaux thèmes et pratiques métiers dans la gestion des données
  - ✓ Une version 2 est prévu pour début 2022 avec de nouveaux apports
  - ✓ Appel à Identifier les présentations, axes, thèmes et problématiques qui pourraient manquer
  - ✓ Faire remonter des pratiques, outils utilisées dans SIST dans le domaine des données
- Nécessité d'être 2 par réseau ⇒ *recherche d'un.e 2eme représentant.e* SIST dans l'atelier de données

[contact-guide@groupes.renater.fr](mailto:contact-guide@groupes.renater.fr)

---

- Questions ?