



UNIVERSITÉ DE
MONTPELLIER

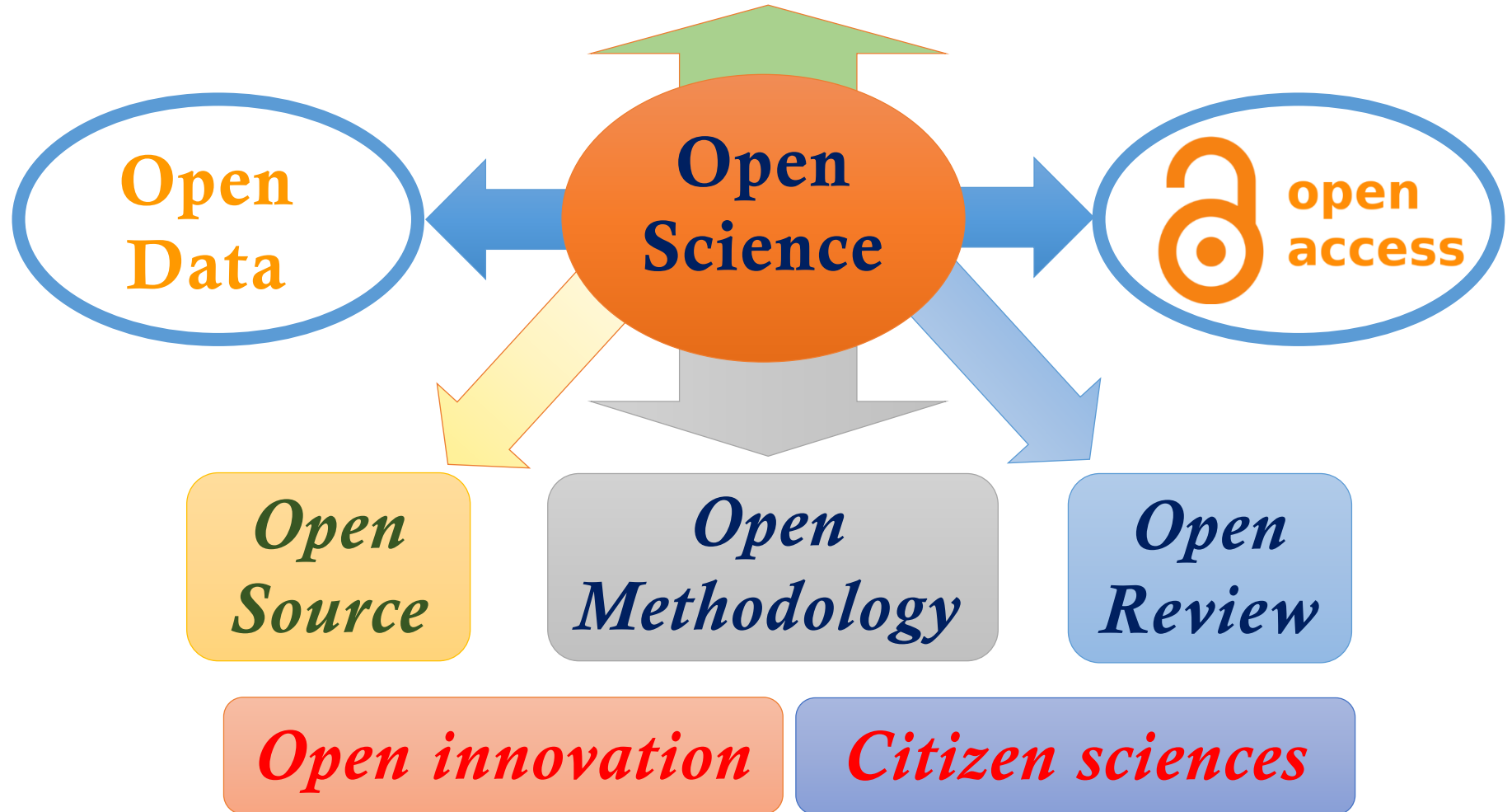


Data management capacity building at UM

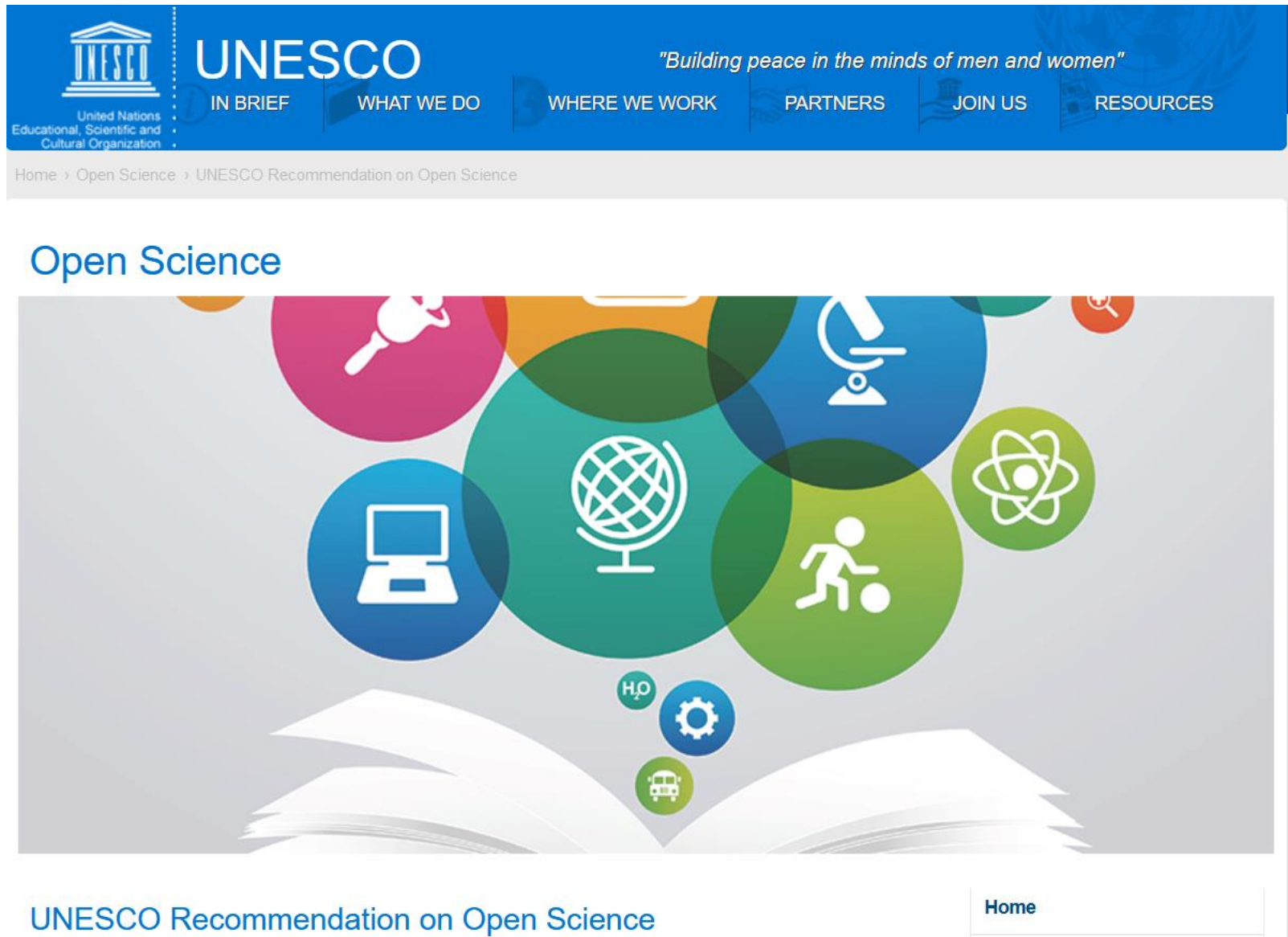
Minerva study visit - 16/06/2021

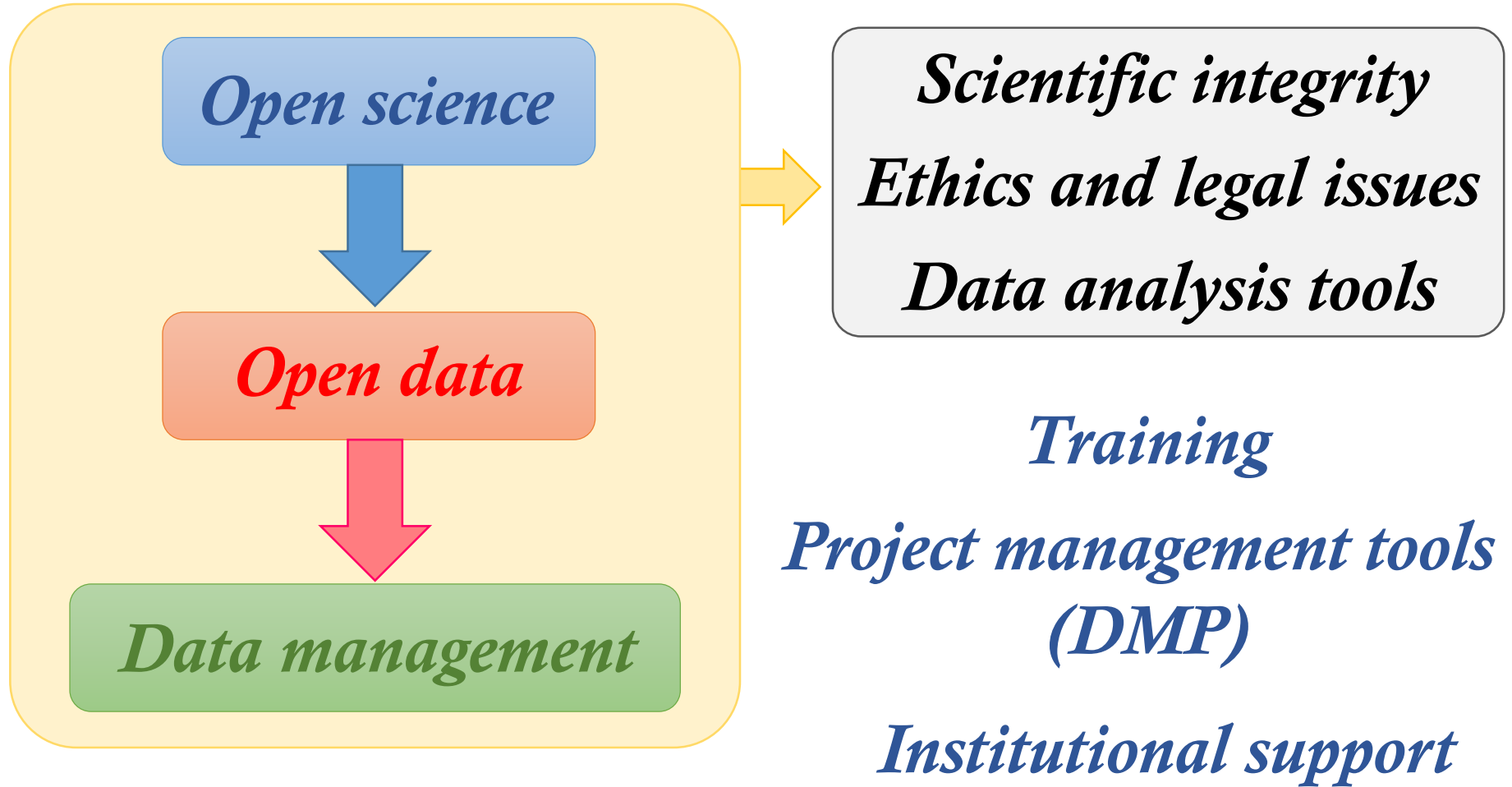
What is open science ?

Open Educational ressources



What is open science ?





- 1) *Data management plan*
- 2) *Scientific Data Management training program*
- 3) *Support provided to researchers*
- 4) *Integrity and ethics at UM*

- 1) *Data management plan*
- 2) *Scientific Data Management training program*
- 3) *Support provided to researchers*
- 4) *Integrity and ethics at UM*

What is a DMP ?



An evolving document
(at least 3 versions)



Helps to organize
the data



Description of the data
according to the data
life cycle



Defines the
responsibilities



Helps to assess the
resources needed



Helps to obtain
reliable data

Why a DMP ?

20 years after publication, 80% of scientific data are lost.

Causes: destruction of media, file compatibility problems (software version), hardware or software obsolescence, indefinite storage location (usb key, doctoral student's laptop,...)

- Loss of time (data recovery)
- Loss of money
- Difficulty to compare results
- Lack of reuse (possibly different)
- Orphaned data, fossilized data

The objective of the DMP is to ensure that the data are reliable, well managed throughout the project, understandable, available and preserved for future use.



<http://depts.washington.edu/relpov/data-are-mine/>

To whom does the data belong?

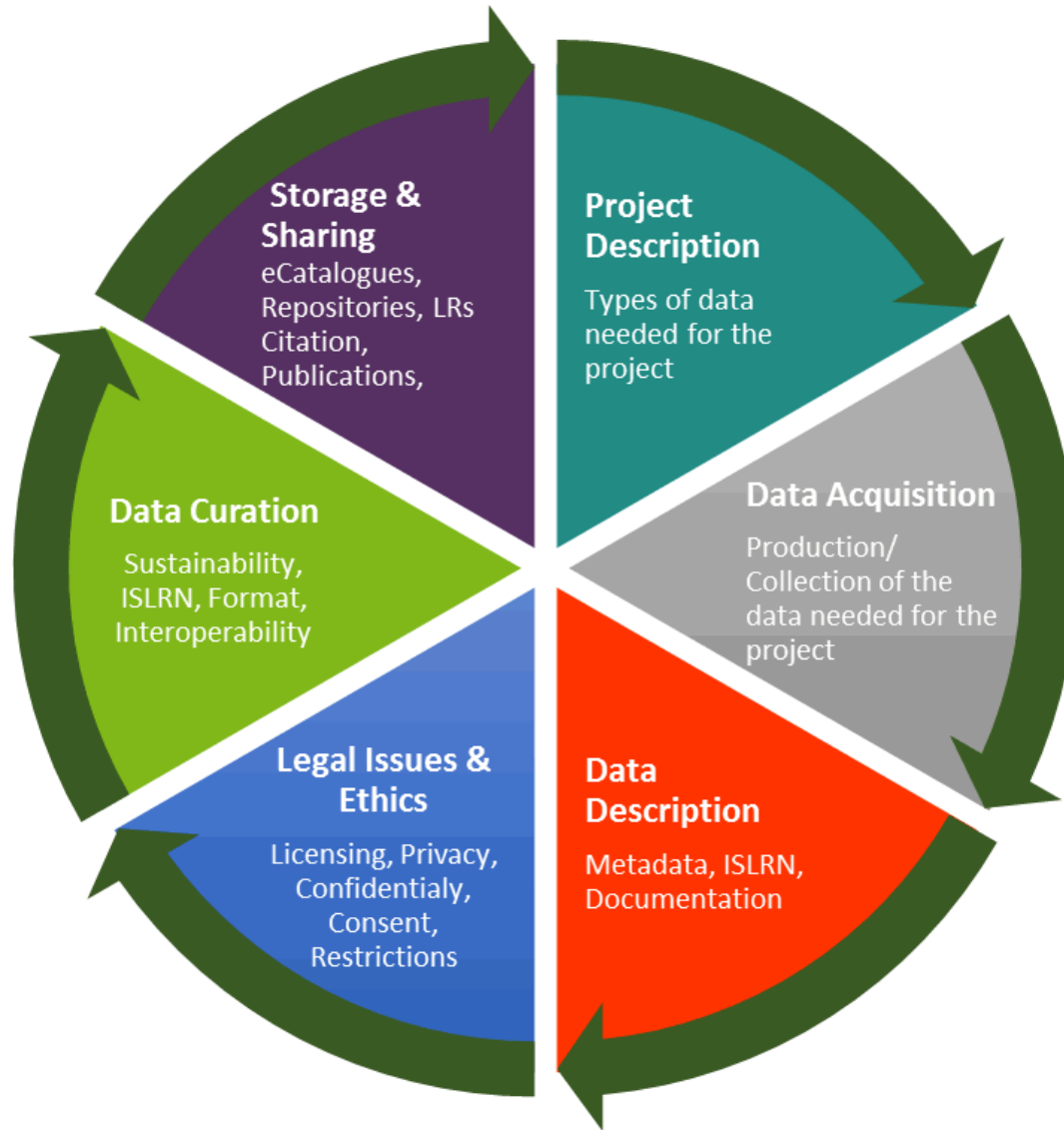
**Who owns
the data?**

**Who has the
right to do what
with this data?**

**Rights and
obligations of
dissemination ?**

**Conflicts IP/industrial
consortium and
openness ?**

DMP is an anticipation tool



<http://www.elra.info/en/services-around-lrs/dmp/>

DMP as a project management tool

- ✓ Anticipation and cost rationalization (volume of data)
- ✓ Quality of data
- ✓ Anticipation of data valorization and/or public sharing
- ✓ Storage and archive strategy for further use and value preservation

Development of novel powerful tools for data analysis - AI

- ✓ Powerful tools that accessible to most of researchers without specific programming skills
- ✓ Requires support for appropriate use and acculturation to its potential

DMP is a tool for researchers

What are the advantages for researchers of creating a DMP?

- **Efficiency:** Increase of work efficiency over the entire project duration (ensure retrievability, avoid data loss), as it is defined in time how data should be stored securely and possibly shared. Gain of time when data is easy to find and understand.
- **Security:** Avoiding risks in copyright matters (no transfer of data to commercial platforms, adequate licensing) by defining who may have access to which data and how sensitive data are protected.
- **Citations:** Enabling re-use and increasing the citation frequency of research results through a) the use of persistent identifiers (e.g. doi – digital object identifier), b) the provision of data in repositories such as Zenodo, and c) the re-use of own data by other researchers.
- **Credibility:** Improve your own credibility and confidence in other publications when data are freely accessible or when at least the published DMP shows how the data were generated and managed.

<https://www.uzh.ch/blog/hbz/2018/11/15/data-management-plan-in-a-nutshell/>

Bienvenue !

DMP OPIDoR vous accompagne à travers l'élaboration et la mise en pratique de plans de gestion de données et de logiciels.



Accessible à la communauté scientifique de l'ESR et à ses partenaires français ou étrangers



Personnalisable par tout organisme de recherche pour la mise en place de sa politique de données



Enrichi par des exemples et des recommandations adaptés à l'environnement de recherche



Collaboratif : il facilite les échanges entre les partenaires d'un même projet et les services d'accompagnement

DMP OPIDoR évolue grâce à vos retours. Les développements s'inscrivent dans le cadre d'une collaboration internationale autour du logiciel open source DMPRoadmap

[Rejoignez la communauté des utilisateurs de DMP OPIDoR](#)

<https://dmp.opidor.fr/>

[Connexion](#)[S'inscrire](#)

Accès institutionnel

Connexion

Accès individuel

* Courriel

* Mot de passe






[Mot de passe oublié ?](#)

☐ Se souvenir de moi

Connexion

Comment | Published: 31 May 2021

Best practices in machine learning for chemistry

Nongnuch Artrith , Keith T. Butler , François-Xavier Coudert , Seungwu Han , Olexandr Isayev ,
Anubhav Jain  & Aron Walsh 

Nature Chemistry **13**, 505–508 (2021) | [Cite this article](#)

8075 Accesses | **179** Altmetric | [Metrics](#)

Statistical tools based on machine learning are becoming integrated into chemistry research workflows. We discuss the elements necessary to train reliable, repeatable and reproducible models, and recommend a set of guidelines for machine learning reports.

Data's value - new opportunities

Beyond the limits where simple analytical expressions are applicable or sophisticated numerical models can be computed, statistical modelling and analysis are becoming valuable research tools in chemistry. These present an opportunity to discover new or more generalized relationships that have previously escaped human intuition. Yet, practitioners of these techniques must follow careful protocols to achieve levels of validity, reproducibility, and longevity similar to those of established methods.

The purpose of this Comment is to suggest a standard of 'best practices' to ensure that the models developed through statistical learning are robust and observed effects are reproducible. We hope that the associated checklist (Fig. 1 and Supplementary Data 1) will be useful to authors, referees, and readers to guide the critical evaluation of, and provide a degree of standardization to, the training and reporting of machine learning models. We propose that publishers can create submission guidelines and reproducibility policy for machine-learning manuscripts assisted by the provided checklist. We hope that many scientists will spearhead this campaign and voluntarily provide a machine learning checklist to support their papers.

- 1) *Data management plan*
- 2) *Scientific Data Management training program*
- 3) *Support provided to researchers*
- 4) *Integrity and ethics at UM*



"Scientific Data Management" university degree:

aims at training a wide audience in the issues, practices and tools of scientific research data management.

The proposed degree is part of the implementation of the National Plan for Open Science and is attached to the Data Science Institute of Montpellier.



DU Scientific Data Management



Head of program

Dr. Agnès Robin

Associated professor (HDR)
Law school
University of Montpellier
(ERCIM/UMR 5815)

agnes.robin@umontpellier.fr

Training program

- ✓ Initial education / professional training
- ✓ 186 hours / one week per month
- ✓ Full program/ by-module program
- ✓ Diplomation / Certification

More informations:

<https://sdm.edu.umontpellier.fr/>





Who is the program intended for?

- any person interested and concerned by the handling of data sets from research activities;
 - any discipline.
-
- Researchers;
 - Professors;
 - Post-doctoral fellows;
 - PhD degree candidates;
 - Engineers;
 - Innovative project leaders (incubated or not);
 - Master students

DU Scientific Data Management

understanding
the
environment
of scientific
data

MODULE 1.1

territories, innovation and funding

MODULE 1.2

Research, governance and data

MODULE 1.3

Data, innovation and law



INSCRIPTION

mastering
the tools of
data analysis
in science

MODULE 2.1

Data extraction

MODULE 2.2

Management of data and
metadata

MODULE 2.3

Data exploration and data mining

MODULE 2.4

Data security

MODULE 2.5

Protection of personal data

managing
the opening
of scientific
data

MODULE 3.1

Redaction of a Data Management
Plan (DMP)

MODULE 3.2

Data storage, data archive

MODULE 3.3

Data dissemination
and data sharing

MODULE 3.4

Data valorization and reuse

<https://sdm.edu.umontpellier.fr/programme-de-la-formation/>

Next session: septembre 2021

Possibility to register on a rolling basis for each module



- ✓ Europe's first **complete** training course on research data
- ✓ Training available in modules
- ✓ Initial education: Master/Doctorate level
- ✓ DU open to professional training
- ✓ Open to people either from private or public sector
- ✓ Pre-requisites: minimum Master level

- 1) *Data management plan*
- 2) *Scientific Data Management training program*
- 3) *Support provided to researchers*
- 4) *Integrity and ethics at UM*

Research data support at UM

← Catalogue de services


🏠 Accueil > Catalogue de services > Données de la recherche

Parcourir les catégories <

- Tout(e)(s)
- Données de la recherche
- Informatique et Numérique
- Logiciels
- Vidéo


Rechercher... Rechercher

Disponible(s)




Accès à une Infra. de données

Demande




Aide au choix du modèle de PGD

Demande




Aide au dépôt d'un PGD

Demande




Aide au remplissage / relecture d'un PGD

Demande




Choix des formats, standards et métadonnées

Demande




Contrats et Propriété des données

Demande




Expertise ou coopération en IA Intelligence Artificielle

Demande




Obtenir un DOI ou publication data paper

Demande



Questions éthiques et autre demande sur les données de la re
...


Demande



Sensibilité de données / données personnelles

Demande

A single access for different services of UM



Infra stockage de données

Accès à une Infra. de données

 Demande



Déposer son modèle

Aide au dépôt d'un PGD

 Demande



Aide à l'écriture de plan

Aide au remplissage / relecture d'un PGD

 Demande



Classement des données

Choix des formats, standards et métadonnées

 Demande



Contrats et propriété

Contrats et Propriété des données

 Demande



Autres questions

Questions éthiques et autre demande sur les données de la re

...

 Demande



Analyse de sensibilité

Sensibilité de données / données personnelles

 Demande

Access to local data storage infrastructure



Aide au choix du
modèle de PGD

 Demande



Aide au dépôt d'un
PGD

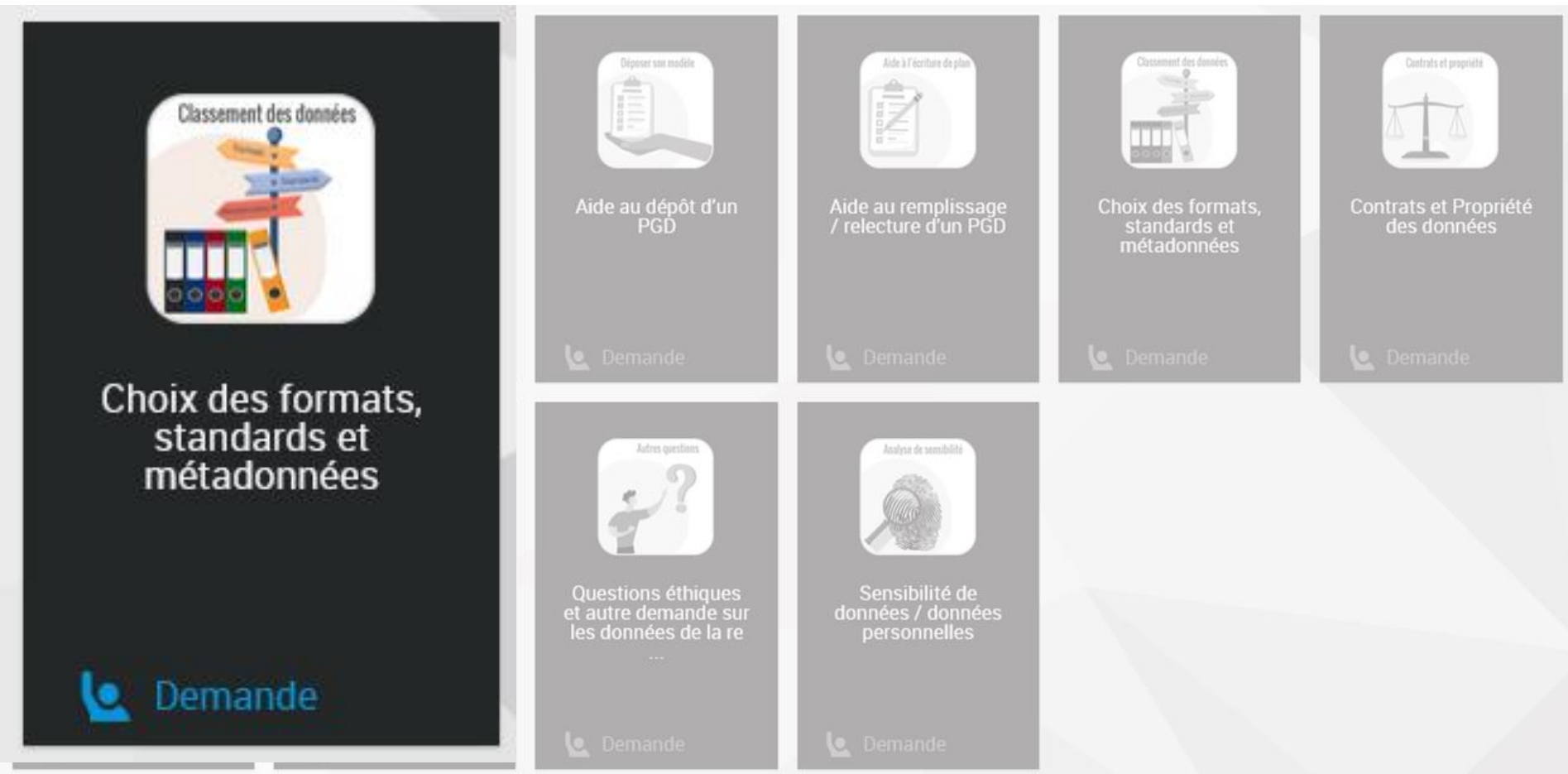
 Demande



Aide au remplissage
/ relecture d'un PGD

 Demande

Support for DMP redaction



Support for data format and metadata associated



Contrats et Propriété
des données

 Demande



Questions éthiques
et autre demande sur
les données de la re
...

 Demande



Sensibilité de
données / données
personnelles

 Demande

Support for data legal issues : property, ethics, GDPR,...



Obtenir un DOI ou
publication data
paper

 Demande



Aide au dépôt d'un
PGD

 Demande



Aide au remplissage
/ relecture d'un PGD

 Demande



Choix des formats,
standards et
métadonnées

 Demande



Contrats et Propriété
des données

 Demande



Questions éthiques
et autre demande sur
les données de la re
...

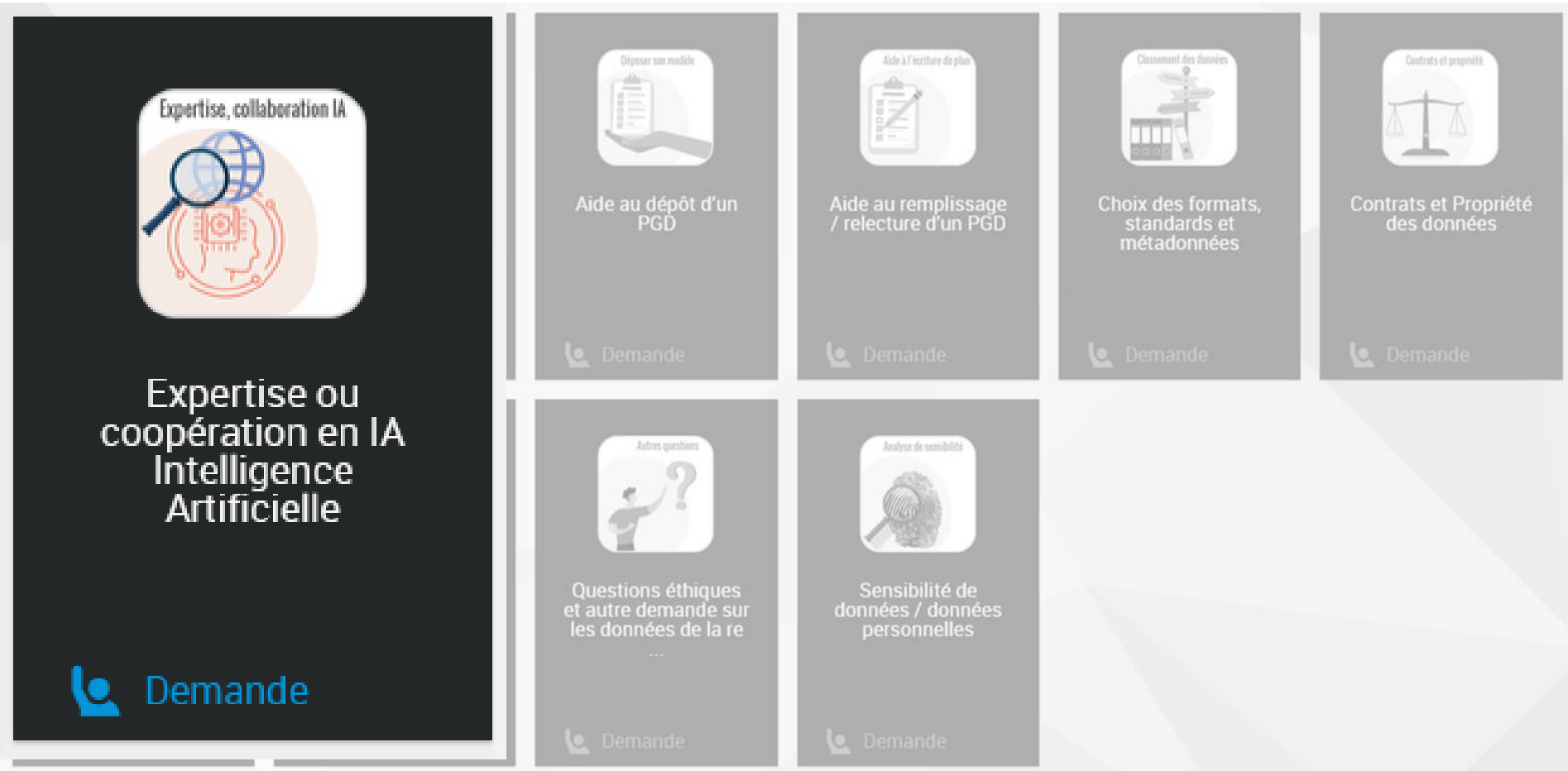
 Demande



Sensibilité de
données / données
personnelles

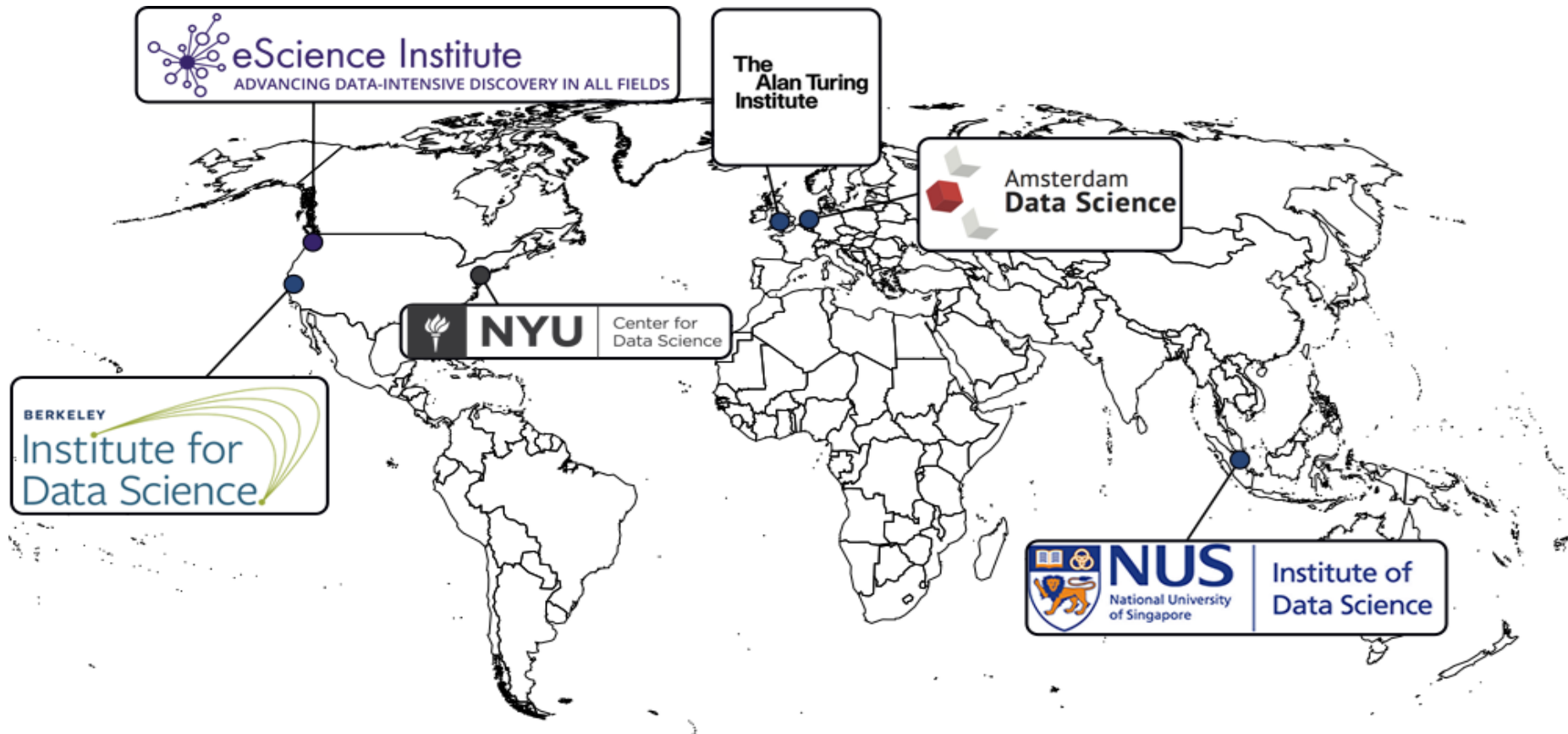
 Demande

Support for data paper or / and DOI attribution

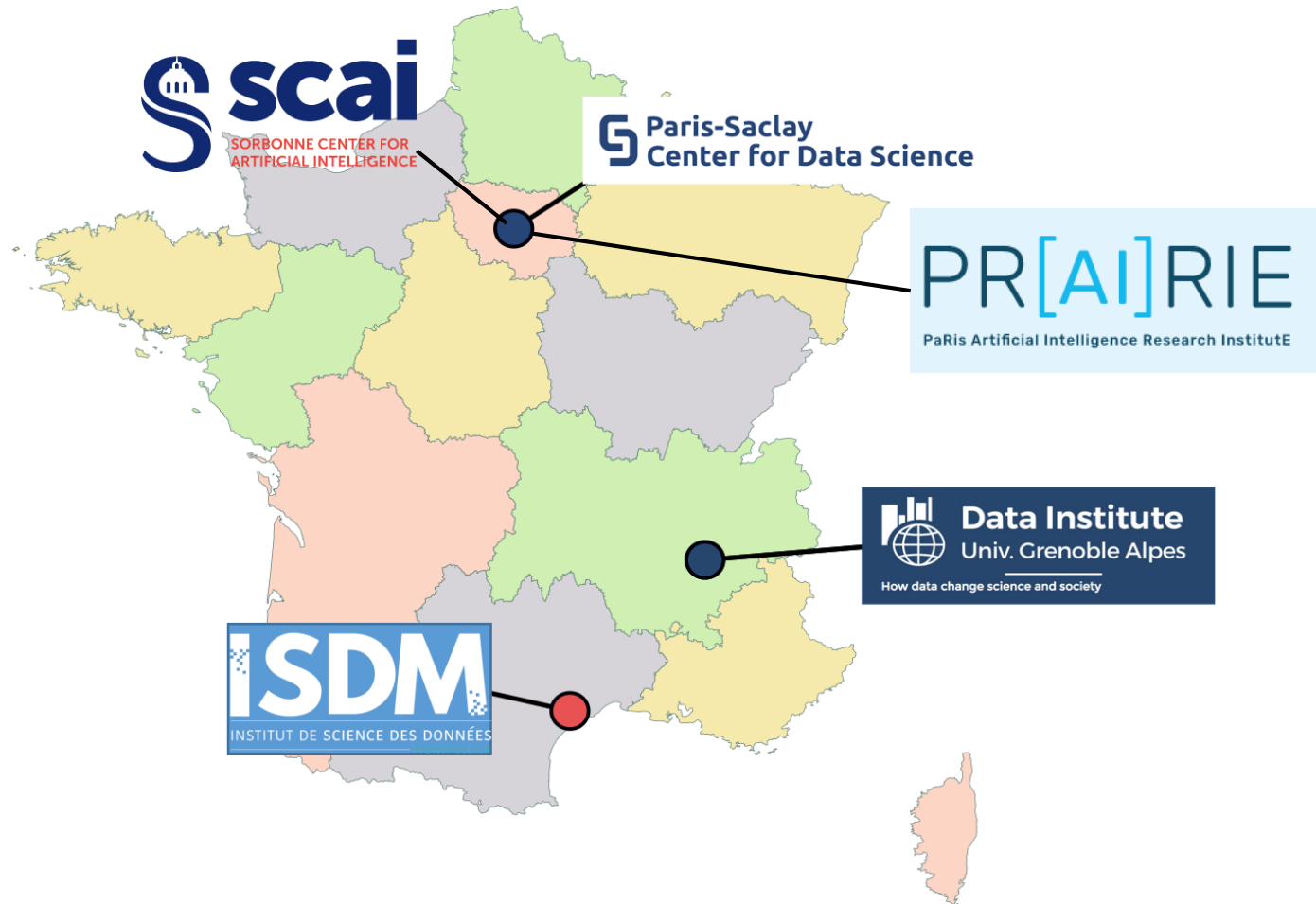


Support for AI expertise

Data science for researchers



Data science for researchers





MONTPELLIER UNIVERSITY OF EXCELLENCE

Institut de Science des Données de Montpellier

isdm@umontpellier.fr

isdm.umontpellier.fr

- 1) *Data management plan*
- 2) *Scientific Data Management training program*
- 3) *Support provided to researchers*
- 4) *Integrity and ethics at UM*

- ✓ **2016 : appointment of a Scientific Integrity Referent.**
The referent can be contacted directly by email.
- ✓ **2016 : creation of an internal procedure at the UM, in order to deal with cases of alleged breaches of scientific integrity.**
Scientific Integrity Commission, responsible for investigating procedures relating to the handling of allegations of breaches of scientific integrity in cases where the Referent alone cannot provide an answer
- ✓ **2017 : adoption of a charter of good practices relating to scientific integrity.**
- ✓ **2017 : appointment of a Vice President for Ethics and Deontology**

✓ *doctoral training:*

- *MOOC Research Ethics;*
- *MOOC Scientific Integrity in the Research Profession;*
- *MOOC Reproducible Research: Methodological Principles for Transparent Science;*
- *Conference "Crossed views on plagiarism", January 7, 2021 (available on Youtube).*



kenneth.maussang@umontpellier.fr