



How to publish research data? The example of Recherche Data Gouv

Améline Vallet – AgroParisTech

26/03/2024

Why publishing data?

- Enhances transparency and credibility in research
- Supports the replication and validation of results, strengthening scientific integrity
- Facilitates collaboration and information sharing across disciplines
- Accelerates the pace of research and findings

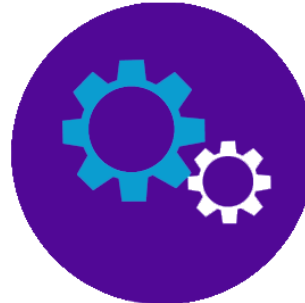
Findable



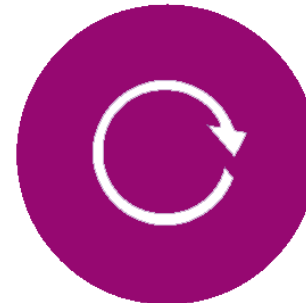
Accessible



Interoperable



Reusable



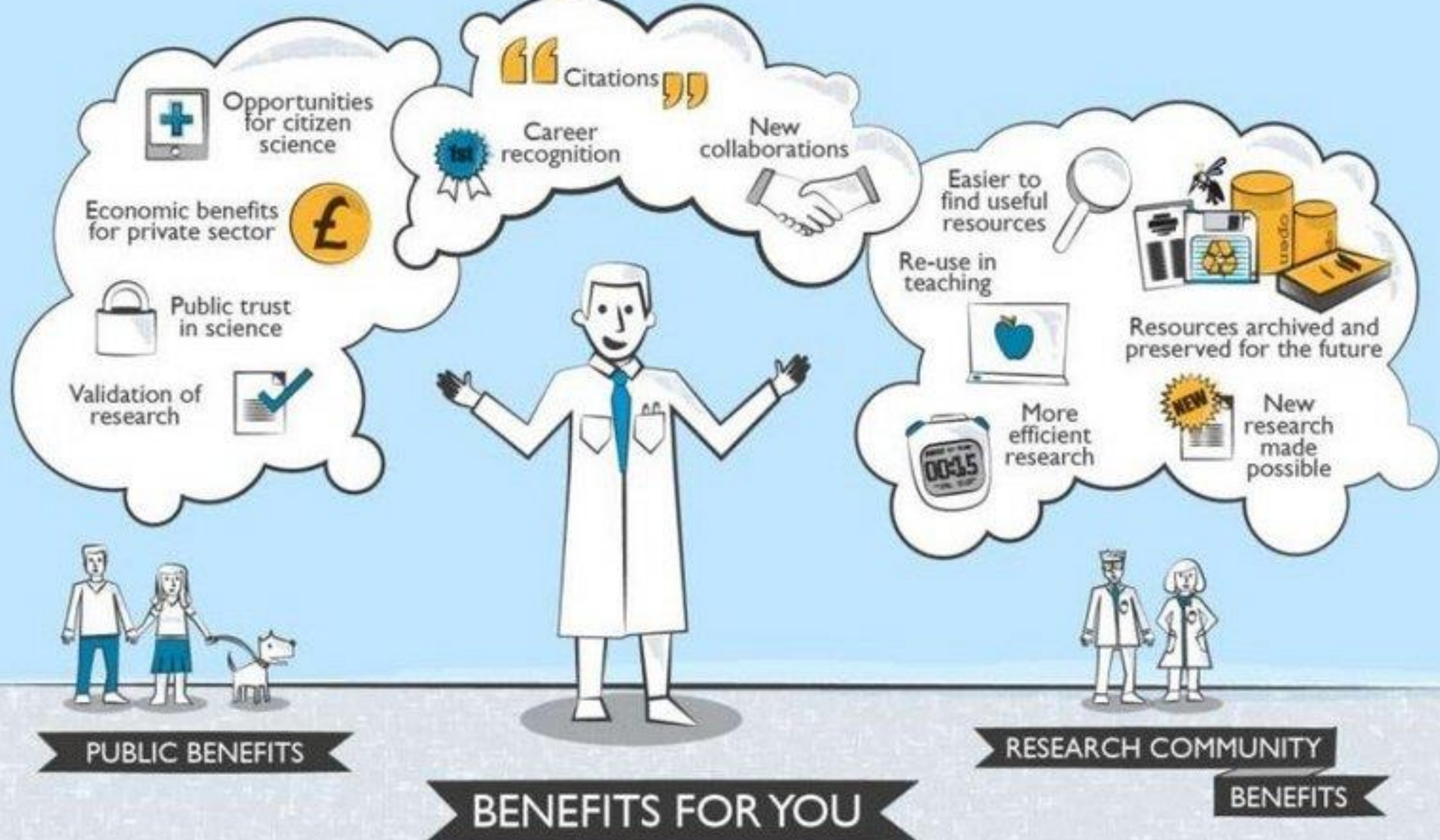
« *Open data citation advantage* »

- Studies that provide access to underlying data are cited 10-25% more often than those that don't

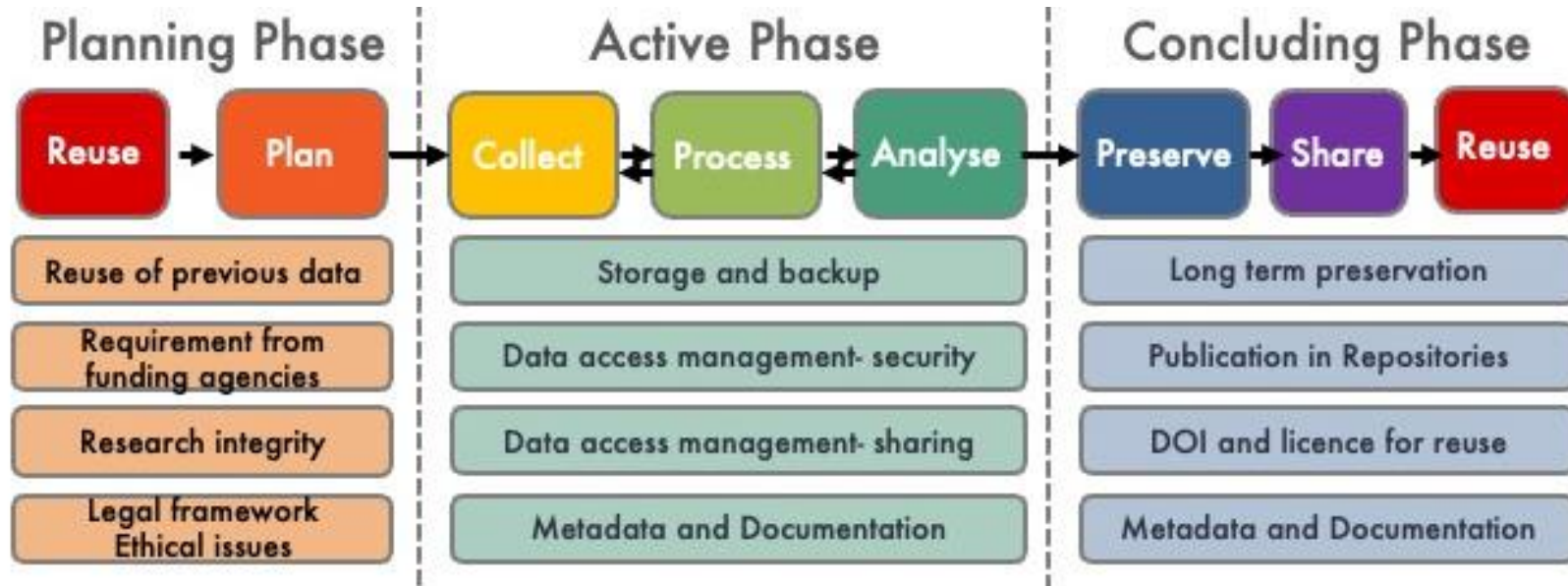
<https://doi.org/10.1371/journal.pone.0230416>

<https://peerj.com/articles/175/>

<https://doi.org/10.1371/journal.pone.0253129>



How to publish data? Data cycle






How to choose a data repository: exclusion criteria

- Absence of moderation
- Absence of a persistent identifier
- Absence of guarantee for infrastructures
- Repositories imposing transfers in property rights / copyrights
- Excessive pricing policy
- Data localization (outside the EU)
- Restrictions for access (e.g., institutional affiliation)

How to choose a data repository: listings

Nom	URL	Institution porteuse	Modération	Identifiant pérenne fourni	Pérennité de l'infrastructure/Engagement sur la durée de préservation des données	Discipline	Champs disciplinaires détaillés	Données acceptées= mots qui parlent à la communauté
data.InDoRES (inventaire des données de recherche en environnement et sociétés)	https://data.indores.fr/	CNRS-INEE et MNHN	La modération est opérée soit par l'équipe data.InDoRES soit par le référent local du réseau OpenDoRES de l'unité de recherche quand il existe. La complétude des métadonnées ainsi que la compréhension des contenus sont vérifiés par le modérateur. Le contenu scientifique est quant à lui laissé à la responsabilité du déposant.	DOI	Le portail InDoRES a été créé en 2016, l'entrepôt data.InDoRES a été lancé fin 2021. La pérennité de l'entrepôt est assurée par le soutien des deux tutelles (MNHN et CNRS-INEE). Le Centre de Calcul de l'IN2P3 assure le stockage et la sauvegarde des données de l'entrepôt.	Sciences de l'environnement	Ecologie, Environnement, Bio-archéologie	Données environnementales, écologiques et géographiques
EaSy Data	https://www.easydata.earth/	Data Terra et BRGM	Modération sur les métadonnées et le périmètre. Organisée par thématiques (liste thématique ici : https://teravocabulary.org/ncl/DataTerraRepositoryFairIncubator/_Voic_thematique). Prise en charge de la modération par des experts de domaine dans chaque pôle de données de Data Terra. Interface de modération fournie. Le GBIF fournit une modération à 2 niveaux : 1) au niveau de l'entité publiant les données (publisher), par exemple le laboratoire de recherche. Le GBIF s'assure que l'entité fournit des données dans le périmètre du GBIF. Ce contrôle est opéré une seule fois, lors de la création de l'entité déposante.	DOI	La pérennité de l'entrepôt est assurée par l'IR Data Terra et le BRGM.	Sciences de l'environnement	Sciences de la terre et de l'Environnement	Données de long terme du système et de l'environnement (exemple : issue d'un projet), hors données d'observation prise en charge par les producteurs, données d'occurrences, données d'échantillonnage, toutes standardisées selon les standards de la communauté.
GBIF Global Biodiversity Information Facility	https://www.gbif.org/	GBIF	2 niveaux : Basic curation : porte sur les métadonnées. Enhanced curation : Conservation des données	DOI	En cas d'interruption d'accès aux données ou de rupture de service, le GBIF transfèrera les données auprès de ses partenaires afin d'assurer la préservation de long terme des données.	Sciences de l'environnement	Sciences du vivant, Biodiversité, Biologie animale, Biologie végétale, Ecologie, Environnement	Données des géosciences (biologie, écologie, chimie)
PANGAEA	https://www.pangaea.de/	Alfred Wegener Institute, Helmholtz Center for Polar and		DOI	10 ans	Sciences de l'environnement	Sciences de la Terre, Sciences du vivant, Ecologie	

Comparison: Zenodo-Dryad-figshare

Zenodo 	Dryad 	Figshare 
FAIR Nonprofit organization (CERN & H2020)	FAIR Nonprofit organization	FAIR Commercial company-Digital Sciences (NPG)
Muldi-disciplinary (physics)	Muldi-disciplinary (life science, medical)	Muldi-disciplinary (life science)
<ul style="list-style-type: none"> ▪ <u>FBM community</u> ▪ Multiformat (All file types) ▪ Manuscript submission workflow ▪ Private link sharing (not friendly procedure) ▪ Quick & simple upload ▪ Satisfy funder (H2020, SNSF) & publisher (check the journal) open data mandates 	<ul style="list-style-type: none"> ▪ Multiformat <u>"All data types ,specific formats "</u> ▪ Manuscript submission workflow ▪ Private link sharing (friendly procedure) ▪ Quick & simple upload ▪ Satisfy funder (H2020, SNSF) & publisher (check the journal) open data mandates 	<ul style="list-style-type: none"> ▪ Multiformat <u>All file types</u> ▪ Manuscript submission workflow ▪ Private link sharing (friendly procedure) ▪ Quick & simple upload ▪ Don't satisfy SNSF funder ▪ Satisfy H2020 & publisher (check with journal) open data mandates
<ul style="list-style-type: none"> ▪ 50 GB limit per package. ▪ Flexible licensing (CC0 or CC BY) ▪ Easily access and reuse ▪ DOI to make data citeable. ▪ Exposing robust metadata in landing page markup with html meta tags ▪ Dataset directly link to publication. 	<ul style="list-style-type: none"> ▪ 20 GB limit per package: \$50 extra per additional 10 GB ▪ Licence CC0 only. ▪ Easily access and reuse ▪ DOI to make data citeable. ▪ Exposing robust metadata in landing page markup with html meta tags ▪ Dataset directly link to publication 	<ul style="list-style-type: none"> ▪ Unlimited public space but 5 GB limit per file. ▪ 20 GB of free restricted private space ▪ Licence CC0 or CC BY ▪ Easily access and reuse ▪ DOI to make data citeable. ▪ Exposing robust metadata in landing page markup with html meta tags ▪ Dataset directly link to publication
<ul style="list-style-type: none"> ▪ Datafile versioning ▪ Contents secure and curated ▪ Long-term preservation 	<ul style="list-style-type: none"> ▪ Possibility to update datafiles ▪ Contents secure and curated ▪ Long-term preservation 	<ul style="list-style-type: none"> ▪ Possibility to update datafiles ▪ Contents secure and curated ▪ Long-term preservation
<ul style="list-style-type: none"> ▪ Free for individuals 	<ul style="list-style-type: none"> ▪ Fee \$120 for individuals if not link to a sponsored journal paper ▪ Big data storage (>20 GB): \$50 extra per additional 10 GB 	<ul style="list-style-type: none"> ▪ Free for individuals
<ul style="list-style-type: none"> ▪ Only simple search tool ▪ Relevant ranking ▪ Altmetrics 	<ul style="list-style-type: none"> ▪ Simple and advanced search tool ▪ Relevant ranking 	<ul style="list-style-type: none"> ▪ Simple and advanced search tool ▪ Relevant ranking ▪ Views/share

How to choose a data repository: guidelines

Subject

UK Data Service



EMBL-EBI



The Cambridge Crystallographic Data Centre



General-purpose



Institutional

University of Reading Research Data Archive



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*

recherche.data.gouv.fr

- Disciplinary repository
- Institutional repository
- Research Data Gov

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES



Liste des entrepôts
de confiance

Research Data Gouv

- Une offre souveraine de dépôt, publication et de signalement des données
- Plateforme nationale pluridisciplinaire et fédérée des données de la recherche
 - Complémentaire des entrepôts thématiques
 - Une solution opérationnelle à destination des équipes de recherche
 - Une alternative à la publication sur des plateformes commerciales ou ne répondant pas à des critères de confiance
 - Devenir un des services de *l'European open science cloud*
- Un maillage d'offres d'accompagnement généraliste et thématique
 - Sensibiliser, former ou accompagner les équipes de recherche en proximité thématique et/ou géographique



RÉPUBLIQUE
FRANÇAISE

Liberté
Égalité
Fraternité

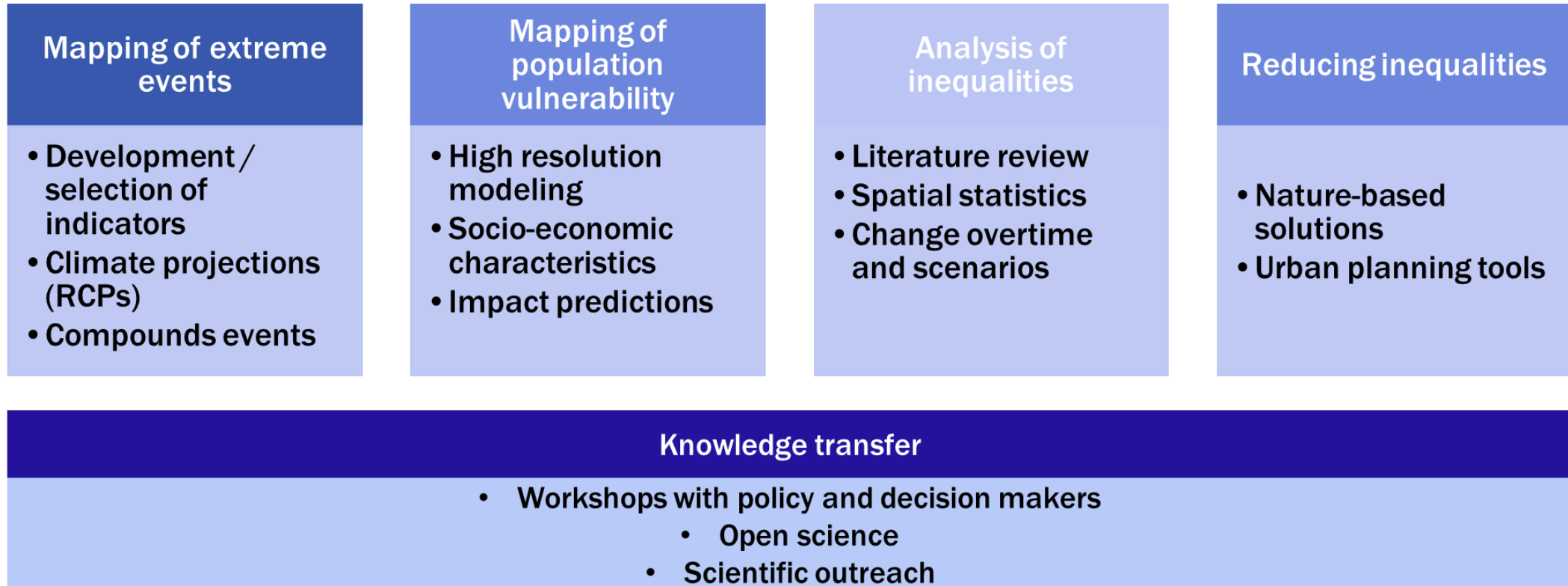
recherche.data.gouv.fr

Example: JUICCE Project



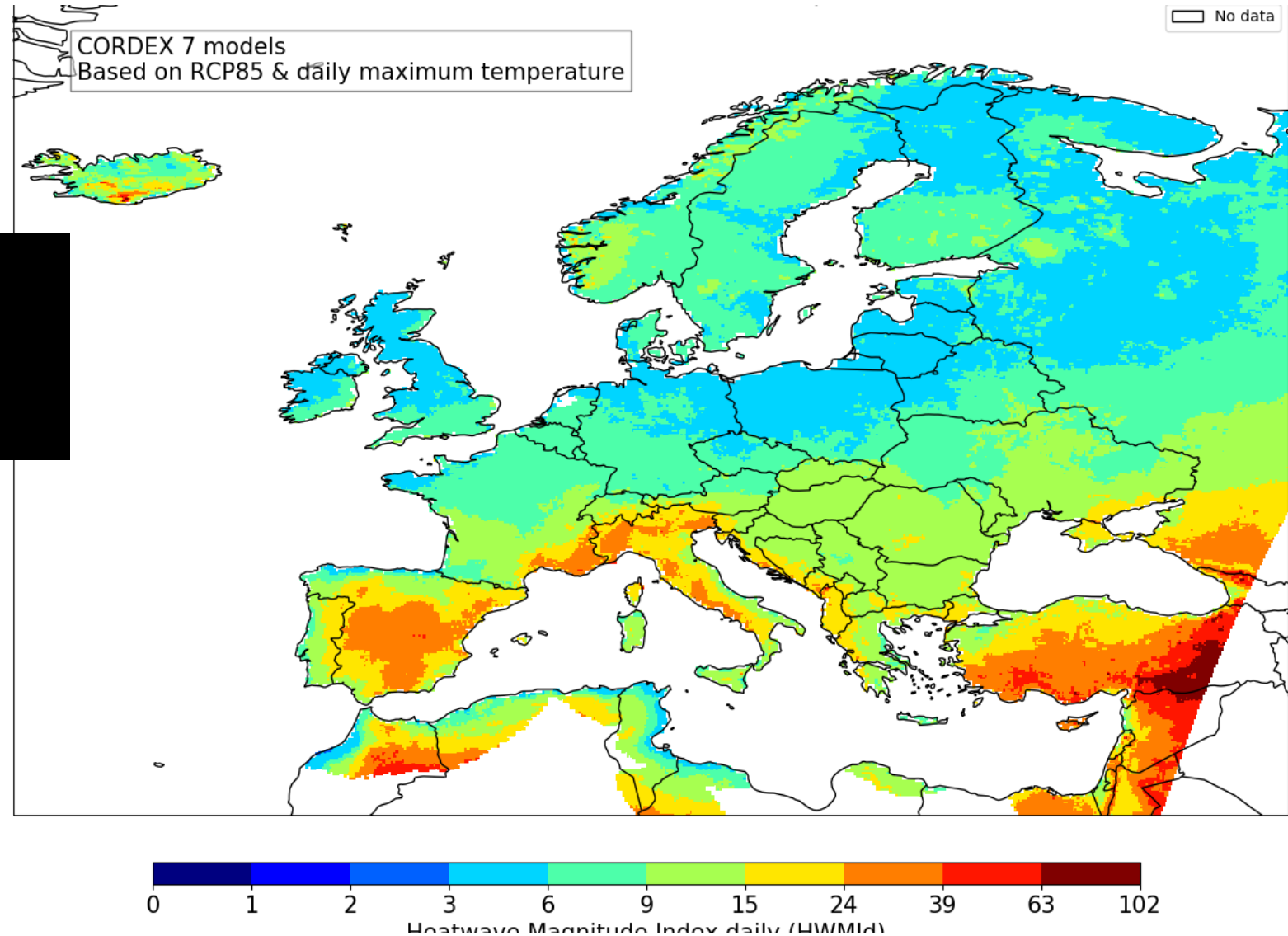
JUICCE

- Research project started in 2019
- Focus on climate justice
- Distribution of the impacts of climate change related extreme events

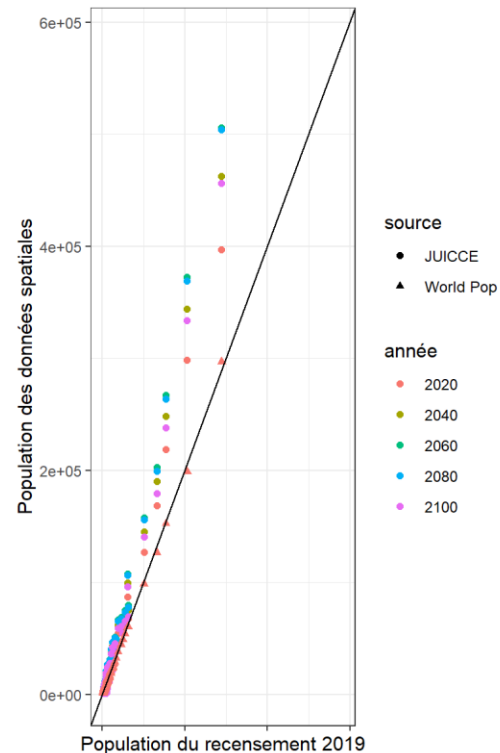
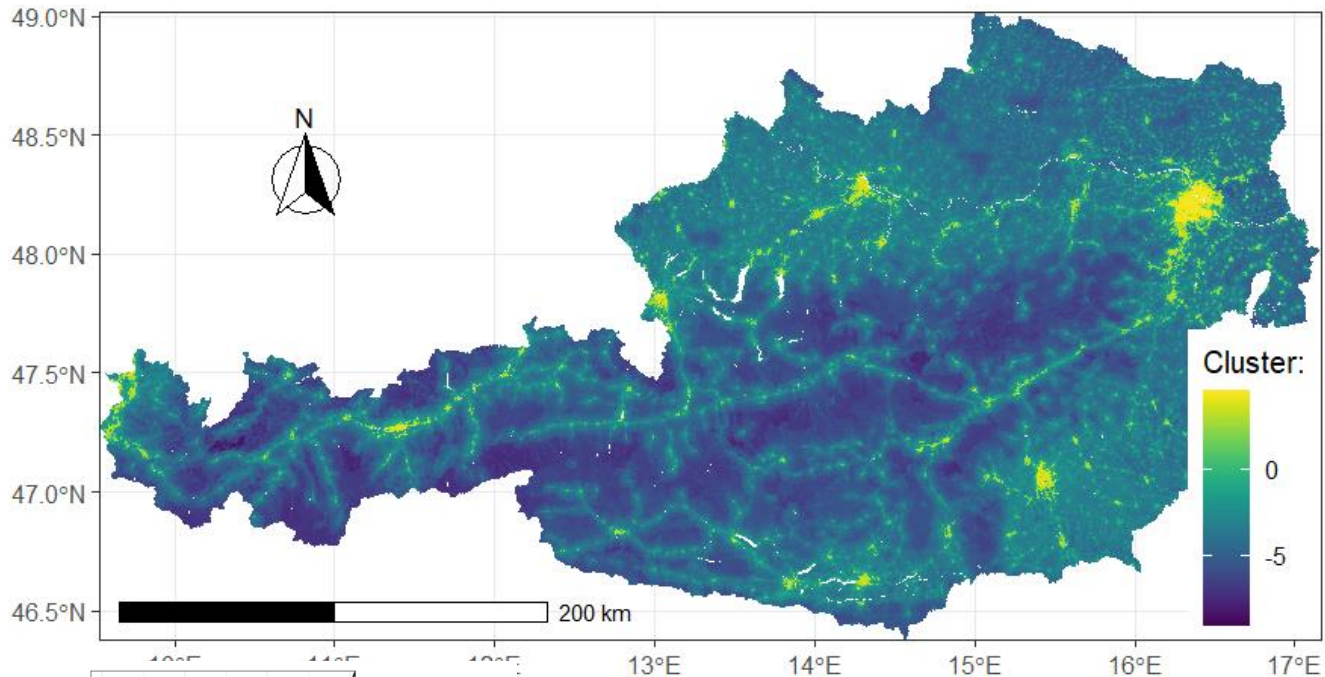


HEATWAVES MAPPING

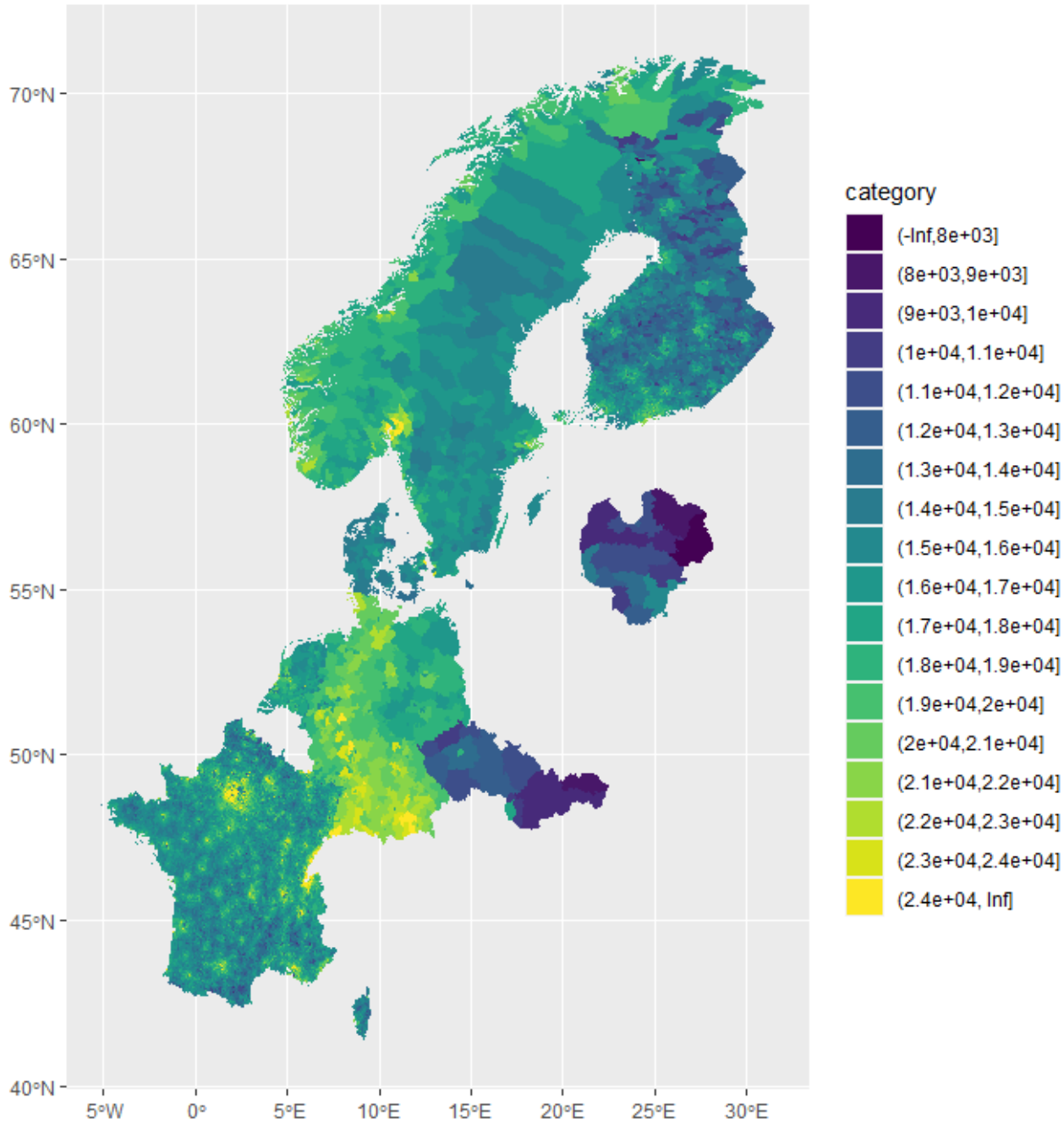
Philémon Poux
Clément Devenet



SPATIALLY EXPLICIT PROJECTIONS OF POPULATION (AGE, GENDER)



Mehdi Mikou
Juliette Piketty-Moine



INCOME MAPPING AND PROJECTIONS

Thérèse Blanchet
Mehdi Mikou



- Demographic, economic and climatic spatial datasets
- High resolution (large volume of data)



Thank you!

Améline Vallet
ameline.vallet@agroparistech.fr