



Les données ne sont pas données

Nils Ferrand, Mathilde Boissier, Peter Sturm, Mathieu Mangeot

► To cite this version:

Nils Ferrand, Mathilde Boissier, Peter Sturm, Mathieu Mangeot. Les données ne sont pas données : Quel protocole d'évaluation participative ex-ante des besoins réels des acteurs dans le système données, modèles, usage décisionnel et impact transformatif?. Journées INRAE systèmes d'information pour les données agro-environnementales, réseau SYSINFO INRAE, Nov 2022, Aubière, France. hal-04233874

HAL Id: hal-04233874

<https://hal.inrae.fr/hal-04233874>

Submitted on 9 Oct 2023

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Les données ne sont pas données

Quel protocole d'évaluation participative ex-ante des besoins réels des acteurs dans le système données, modèles, usage décisionnel et impact transformatif ?

Nils FERRAND*#

Mathilde BOISSIER#, Peter STURM#, Mathieu MANGEOT#

*UMR G-EAU Montpellier & #INRIA STEEP Grenoble

Nils.Ferrand@inrae.fr / Nils.Ferrand@inria.fr

Mathilde.Boissier@inria.fr , Peter.Sturm@inria.fr

“Les données ne sont pas données”

“Toute donnée se caractérise par un vaste faisceau de conventions (sémantique, nomenclatures, formats, etc.), et par l’infrastructure de connaissances dans laquelle elle s’inscrit, impliquant des choix qui n’ont rien de neutre. Une donnée se révèle aussi dépendante de l’environnement qui lui a donné naissance, et des processus productifs qui l’utilisent. On constate alors que les données ne sont pas pures et parfaites, ne vont pas de soi : paradoxalement, les données ne sont pas données. [...] elles sont intrinsèquement liées à un usage. Elles ne sont là que pour concourir à un objectif et non pour mettre à disposition du public une information de référence. [...] Elles requièrent en amont un travail de modélisation, d’abstraction, de spécification des concepts, puis des domaines, avant d’imaginer produire des valeurs.”

[P. Rivière, INSEE,

<https://www.insee.fr/fr/statistiques/fichier/5008707/courstat-5-8.pdf>]

“Les données ne sont pas données” (bis)

cf. Alain Desrosières → les données sont issues de conventions sociales pour les fabriquer et les utiliser

*“Toute quantification implique [...] une série de conventions en amont des prises de mesure afin de construire un espace d’équivalence permettant de prendre appui sur une rationalité partagée. [...] La] construction [des données] permet de « faire des choses qui tiennent », c’est-à-dire de s’entendre sur les termes mêmes dans lesquels se joue le débat. [...] La statistique a toujours participé d’une biopolitique, c’est-à-dire d’une politique qui entend encadrer la vie au nom de sa préservation future par la formulation d’une série d’impératifs (qu’ils soient médicaux, politiques, sexuels, sécuritaires ou personnels).” → “**gouvernementalité algorithmique**”* [De Becker, A. 2017. Défaire l’évidence de la gouvernementalité algorithmique. UCL.

https://dial.uclouvain.be/memoire/ucl/fr/object/thesis%3A12428/datastream/PDF_01/view]

“Les données ne sont pas données” (ter)

Cf. coûts de la production, traitement, stockage, diffusion, maintenance des données

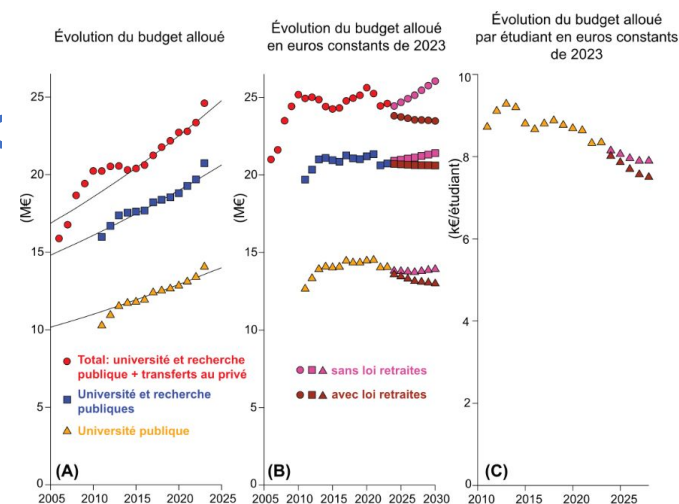
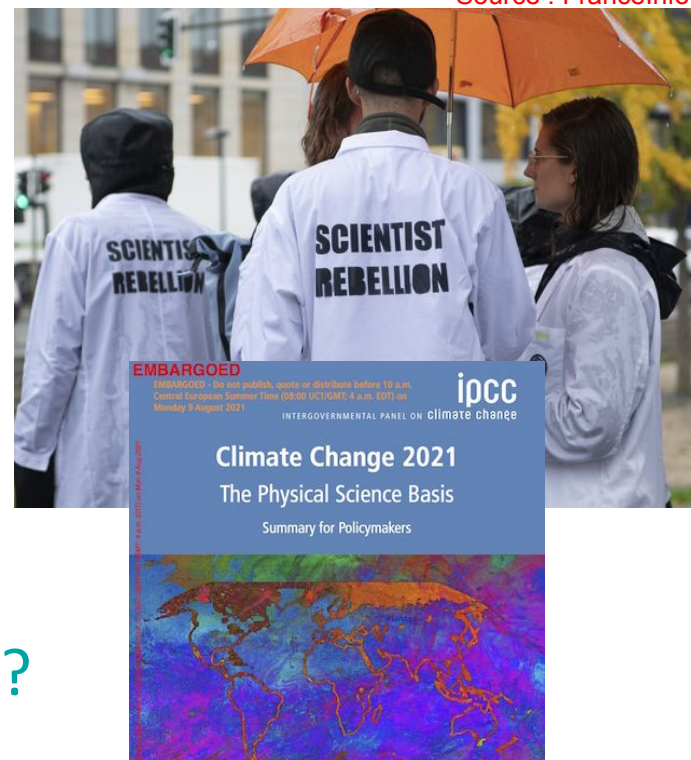
- économique : 3€/Go/an (2014) → 0,2€/Go/an (2022)
 - écologique : 7 à 40 gCO₂e/Go/an [Charret & al, 2020], empreintes matériaux, espace
 - sociaux → perte de contrôle des données personnelles (vol, chantage), impacts humains du New Public Management (“quantophrénie” administrative, cf. <http://www.penombre.org/> & De Gaulejac, 2010)
- + Questions sur les pratiques scientifiques fondées sur du “deep learning” et/ou sans modèle préalable [Varenne, 2013], vs. “artisanat et rhétorique” dans l’exploration et l’exploitation [Schmidt, 2018]

“What’s worth knowing ?”

- Urgences climatiques et sociales
Mais impact “discutable” / limité
des apports scientifiques
- Questionnements sur les relations
Science / Société / Politique
- Budgets recherche - enseignement :
affectation balistique / politique ?

⇒ Y-a t’il une méthode scientifique
pour le choix des données, connaissances
modèles, travaux pour des enjeux et
situations donnés ?

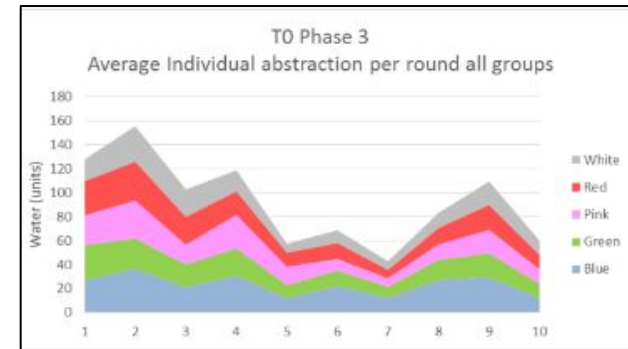
Source : FranceInfo



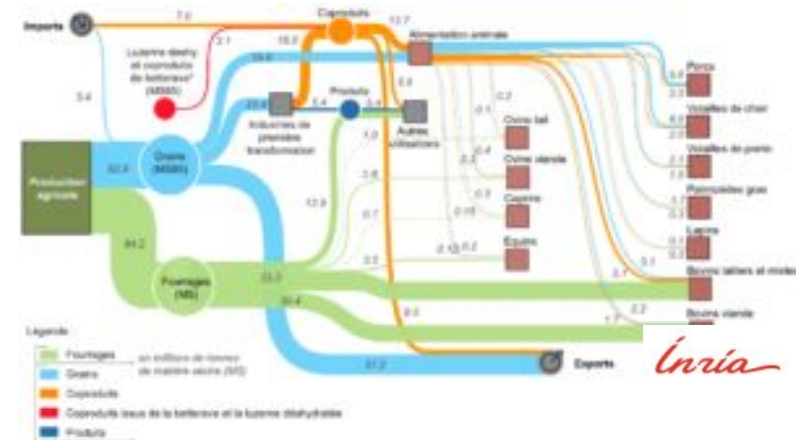
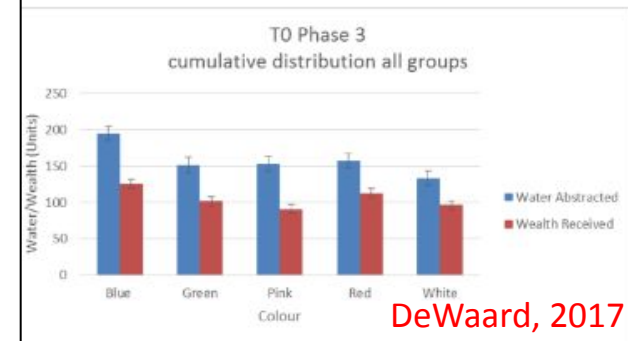
Source : <https://roqueesr.fr/category/billets/>

Des questions opérationnelles, en contexte

- Recherche - intervention (cf. SYSINFO 2021 - cadre CoOPLAGE [<http://set.Watagame.info>]) sur l'ingénierie de l'action publique participative ⇒ quelles méthodes et outils pour accompagner la participation de tous les acteurs (citoyen.ne.s, gouvernants, opérateurs) dans des trajectoires de changement durable ?
 - Projets “en opération” sur des terrains internationaux
 - Sans ou avec numérique (cf. CoOPILOT + Résurgences R&D)
- + expérimentation comportementale / sociale (PRECOS - Bonté & al, 2019 - De Waard, 2017)
- Avec INRIA STEEP (projet SYSPART 2021-...) ⇒ qualification par mise à l'épreuve du terrain et des acteurs des modèles de métabolisme territorial (Analyse de flux de matières et énergie - Courtonne & al, 2022)



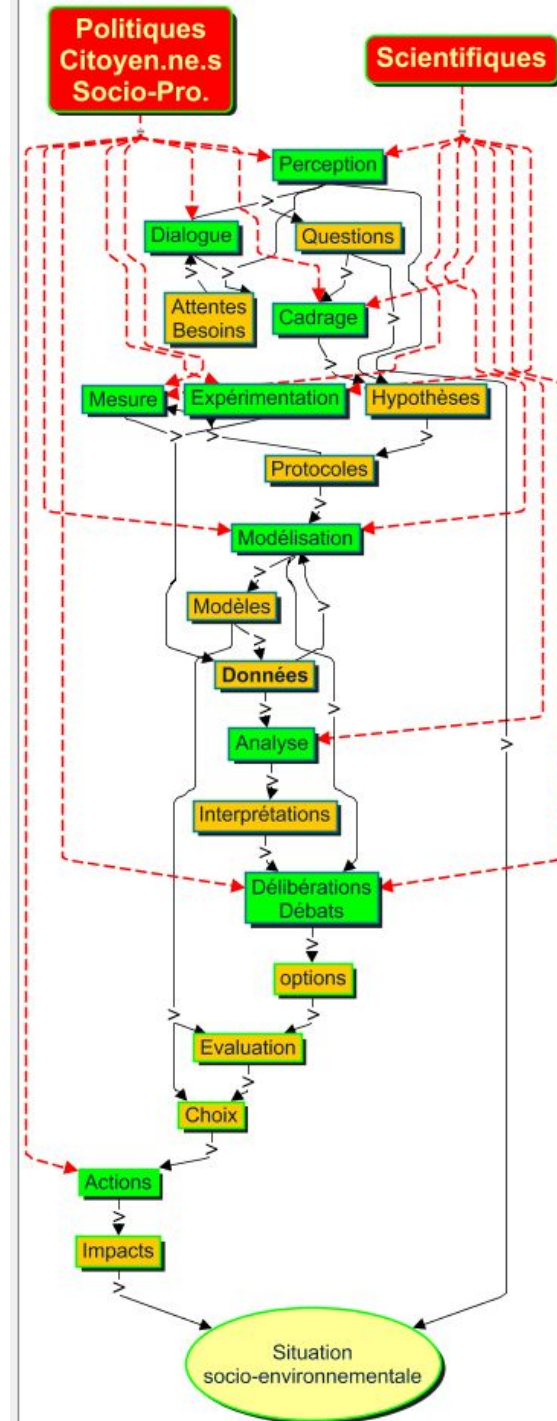
Phase 1 vs Phase 3



Les chemins de valeur de la donnée à l'impact transformatif

- Hyp1 (GIECC) : Situation problématique → analyses scientifiques incrémentales (mesures, modèles, données, information) → alerte (/ média) → saisissement ? → actions politiques, industrielles et individuelles → impacts ?
- Hyp2 : Situation problématique → attentes & débats publics → enquête sociale (Zask, 2019), cadrage et options socio-techniques → requête aux scientifiques → dialogue de recadrage → réponses scientifiques ? → dialogue d'appropriation → actions → impacts

Push vs pull ?



Pré-évaluer pour “choisir” des données & leurs usages ?

(Vs. post-évaluer = usage effectif par les acteurs des données *rendues* disponibles (p.e. Pornon, 2017) → imputabilité décisionnelle “*quand j’ai su que ..., j’ai...*”)

NB1. données indissociables des contextes et modes d’usage
⇒ on évalue un composite acteurs - contexte - données - modèles - processus décisionnel
⇒ le résultat sera contingent, faiblement généralisable !

Critère d’évaluation : **induction effective d’un changement décisionnel et opérationnel conduisant à un impact cohérent avec les préférences préalablement exprimées par les acteurs**

Pré-évaluer (protocole de principe)

1. Cadrage acteur - système : avec qui, sur quoi
2. Les acteurs engagés précisent des objectifs transformatifs (même incohérents ou divergents) sur le système socio-environnemental
3. 1ère modélisation participative (p.e. cartographie conceptuelle, fonctionnelle, problématique) → précadrage
4. Identification par les scientifiques & experts des :
 - a. connaissances disponibles vs. manquantes
 - b. modèles et outils disponibles vs. possibles
 - c. données & observations disponibles vs. collectables
 - d. processus d'usage, protocoles
5. Forum d'évocation (Ferrand, 2008) avec exposition de l'existant
6. Sélection et bascule en mode expérimental face à une situation type (simulation participative, i.e. jeu de rôle)
 - a. Sur données / outils réels
 - b. Sur données / outils simulés (Magicien d'Oz - Baccino, 2009)
7. 2nde modélisation participative → recadrage
8. Evaluation délibérative des processus / modèles / outils

Protocoles expérimentaux initiés



Projet + CNRS MITI 2022 → évaluation participative de l'usage décisionnel et de l'impact transformatif des modèles de flux de matière et énergie

- En laboratoire (collab. INNOVACS UGA) 2023
- Sur le terrain (PETR Briançonnais → concertation énergie / biodiversité / eau / économie locale)



⇒ cadrage préalable
⇒ 35 enquêtes de positionnement avec acteurs
⇒ pré-tests protocoles d'usage modèles AFME
⇒ identification des données & modèles manquants
⇒ tests "réels"
⇒ impacts ?



INRAE *Inria*

Questions ouvertes → discussion

- Impact de l'ergonomie des substrats AFME vs. données
- Demande primaire anté-expé de formation → inclusion dans le protocole ?
- Protocole “méta-évaluatif” (cf thèse Sarah Loudin, 2019)
- Quel appui croisé entre expérimentations en laboratoire et expérimentations de terrain sur cas réel ?
- Modèles / jeux supports minimaux vs. réalistes → validité interne vs. validité externe
- Impact de l'imprécision modèles - données sur la perception d'ensemble du processus et donc sur ses résultats
- Croisement (nexus) et qualité différentielle des données
- Impact de la réconciliation dans les AFME sur l'appropriation

Références

- Bonte Bruno, Farolfi Stefano, Ferrand Nils, Abrami Géraldine, Diallo Mamadou Ciss, Dubois Dimitri, Johannet Anne, Gaudi Wanda Aquae. 2019. Building new kinds of meta-models to analyse experimentally (companion) modelling processes in the field of natural resource management. Environmental Modelling and Software, 120:104486, 24 p. <https://doi.org/10.1016/j.envsoft.2019.07.011>
- Loudin, S. 2019. Can we use a social experiment to assess the impact of participatory processes for water management? Studying a generic method tackling the evaluation of capabilities. thèse ABIES. <https://www.theses.fr/2019IAVF0020>
- De Waard, S. 2017. DECISION MAKING UNDER UNCERTAINTY IN WATER MANAGEMENT. Exploring experimental methods in multi-actor systems. - MSc thesis - Uni. TWENTE & IRSTEA - <https://drive.google.com/file/d/1xClOou0U7RLBwfkxo6ySpkxu16Tltcvh/view?usp=sharing>
- Grey Anderson, Louise Beaumais, Louise Carnapete, Iris Lambert, Sami Makki, et al.. Exploring Practices of Conflict Data Production, Analysis, Dissemination, and Practitioner Reception: Methodological Framework and Preliminary Findings. 2021. ffhaf-03218342
- Guillaume Charret, Alexis Arnaud, Françoise Berthoud, Bruno Bzezniak, Anthony Defize, et al.. Estimation de l'empreinte carbone du stockage de données. [Rapport de recherche] CNRS - GRICAD. 2020. ffhaf-03573790
- Gaulejac, Vincent. (2010). La NGP: Nouvelle gestion paradoxante. Nouvelles pratiques sociales. 22. 83. 10.7202/044221ar.
- Eglantine Schmitt. Explorer, visualiser, décider : un paradigme méthodologique pour la production de connaissances à partir des big data. Histoire, Philosophie et Sociologie des sciences. Université de technologie de Compiègne, 2018. Français. ffnNT : ff. fftel-01960545
- Franck Varenne, Marc Silberstein. Modéliser & simuler. Epistémologies et pratiques de la modélisation et de la simulation, tome 1. Franck Varenne et Marc Silberstein. Editions Matériologiques, 1, pp.974, 2013, Sciences et Philosophies, 978-2-919694-19-8. <hal-00826655>
- Jean-Yves Courtonne, Mathilde Boissier, Alexandre Borthomieu, Léon Fauste, Vincent Jost, et al.. Des outils numériques d'aide à la décision pour la conception et l'évaluation participative d'organisations alternatives de l'économie. ROADEF 2022 - 23ème congrès annuel de la Société Française de Recherche Opérationnelle et d'Aide à la Décision, INSA Lyon, Feb 2022, Villeurbanne - Lyon, France. pp.1-2. ffhaf-03595273
- Zask, J. (2019). L'enquête et ses obstacles. Recherche & formation, 92, 83-94. <https://doi.org/10.4000/rechercheformation.5721>
- Pornon, H. 2017. Quel ROI pour la ville intelligente ? <https://henripornon.wordpress.com/2017/10/18/quel-roi-pour-la-ville-intelligente/>
- Ferrand, N. 2008. Forum des besoins. Projet ANR INTERMED. https://docs.google.com/presentation/d/1d_rJUwnCxyxU5tZFmJuFYRi7pU6iHkxzQSTGqT6vVek/edit?usp=sharing
- Baccino, T. (2009). Prototypage. Document numérique, 12, 133-144. <https://www.cairn.info/revue--2009-2-page-133.htm>.