

Réutilisation des données sismologiques

Journées du Réseau National de la Science Reproductible
Jonathan Schaeffer <jonathan.schaeffer@univ-grenoble-alpes.fr>

27 mars 2024



<https://www.epos-france.fr>

<https://seismology.resif.fr>

datacentre@resif.fr



Contexte

Le centre de données sismologique Epos France

- › Certification Core Trust Seal
- › Données in-situ de toute la planète
- › Consultées depuis le monde entier
- › Protocole de distribution standardisé pour la sismologie

Physionomie d'une donnée sismologique

- › Données in situ, parfois dans des conditions complexes.
- › Organisation des données hiérarchique par réseau de stations
- › Données en séries temporelles sans métadonnées
- › L'interprétation de la donnée n'est possible qu'avec des métadonnées métier riches.
 - ▶ position géographique
 - ▶ orientation des capteurs
 - ▶ réponse instrumentale
 - ▶ conditions au sol
- › Métadonnées de citation (datacite) par réseau sismologique (identifiant persistant DOI)

FAIR data

La sismologie prend un soin tout particulier pour répondre aux critères FAIR

- > Findable
 - ▶ DOI sur les réseaux sismologiques
 - ▶ Métadonnées métier et datacite complètes
 - ▶ Lien métadonnée donnée
- > Accessible
 - ▶ Métadonnées métier consultables selon un standard
 - ▶ Métadonnées survivent à la donnée
- > Interoperable
 - ▶ Représentation des métadonnées par vocabulaire contrôlé
- > Reusable : Détaillés dans cette présentation

Ré-utilisabilité des données

Définition

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

Ce qu'on vise

1. Faciliter l'accès à la donnée
2. Réaliser des corrections sur la métadonnées dans le long terme
3. Garantir que son interprétation sera toujours possible

R1.1 Licence sur les données qui favorise l'utilisation

- › Conformément à la législation sur la science ouverte, les données sont sous licences CC-by 4.0
- › Certaines données sont sous embargo sur une période définie à l'avance
- › Les licences et leur date de prise d'effet sont décrites dans les métadonnées datacite

R1.2 Identification claire de la provenance des données

En s'appuyant sur les métadonnées datacite, nous documentons:

- › le distributeur: le centre de données
- › le producteur: le chercheur responsable du réseau sismologique
- › le curateur: l'organisme ou la personne ayant travaillé sur la validation des données

Quand c'est possible, nous attachons des identifiants uniques à ces informations (OrcID pour les personnes, RE3DATA pour les centres de données, RoR pour les organismes impliqués)

R1.3 Données et métadonnées au standard de la communauté

La sismologie internationale est organisée en fédération depuis des décennies. Elle établit les standards, et on s'y conforme.

Des efforts à poursuivre

Généralisation des DOI au niveau international (2014) :

- › objectif initial : la citation des données dans les publications

Évaluation en cours, pour R1.1

- › 8% des réseaux sismologiques indiquent une licence
 - ▶ dont 81% sont des réseaux Européens

Reproductibilité ?

Ré-utilisabilité vs reproductibilité

Nos systèmes ne garantissent pas la reproductibilité

- › les données peuvent être modifiées par le producteur (corrections, réorientations, re-qualification, complétion, ...)
- › pas de versionning des données en place
- › pas de standard du domaine pour figer une version des données et l'identifier de façon pérenne