

Science & données ouvertes

Principes FAIR

Favoriser la reproductibilité de la recherche
Garantir des données trouvables, accessibles, interopérables et réutilisables tant pour les personnes que pour les machines.



Findable:	Accessible:	Interopérable:	Reusable:
<ul style="list-style-type: none"> (Méta)données facilement découvrable grâce à des métadonnées riches (Méta)données localisable grâce à un identifiant persistant (Méta)données indexées dans système de recherche interrogeable 	<ul style="list-style-type: none"> (Méta)données récupérables à l'aide d'un protocole de communication normalisé et ouvert (avec authentification si nécessaire) (Méta)données restent accessibles même si les données ne sont plus disponibles 	<ul style="list-style-type: none"> Données pouvant être combinées et utilisées avec d'autres données ou outils Utilisation de format ouvert, de standard de (méta)données et d'échange (Méta)données doivent référencées des vocabulaires FAIR 	<ul style="list-style-type: none"> (Méta)données suffisamment décrites pour pouvoir être réutilisées et/ ou combinées dans différents contextes Provenance détaillée Licence d'utilisation appropriée Format standard du domaine et auto-documenté

Enjeux, bénéfices, freins

Enjeux

- Capitaliser les données, éviter leur perte
- Accélérer les avancées scientifiques en encourageant les collaborations et la réutilisation des données (éviter la dispersion des efforts, la duplication)
- Transparence sur les résultats de la recherche (partage des données et des méthodes)

De nouvelles opportunités

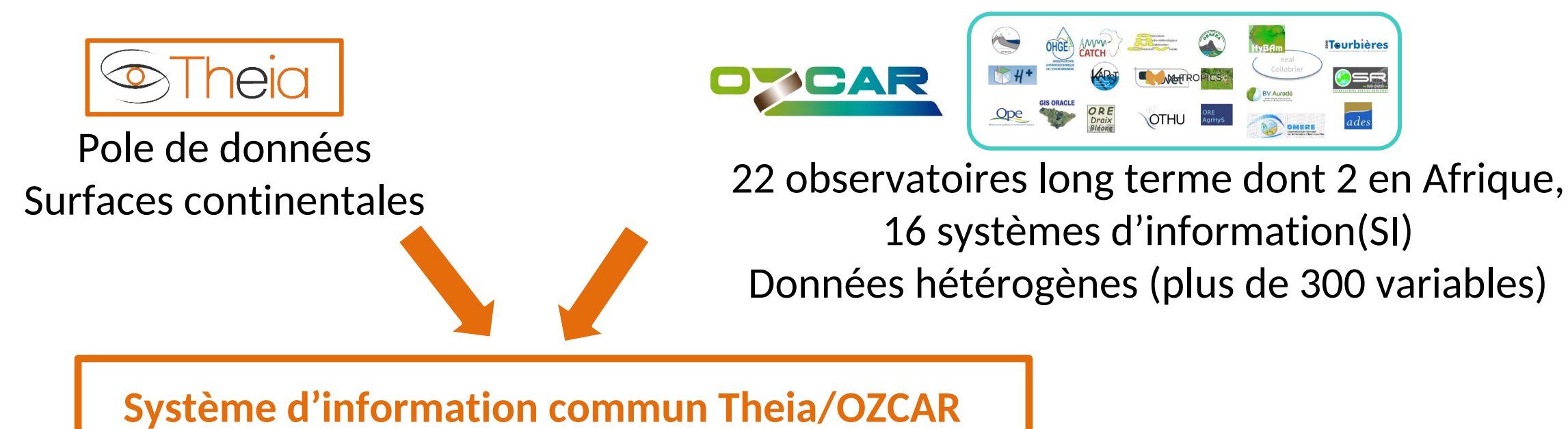
- Pour les chercheurs : plus de visibilité et de crédit, collaboration, participation à des projets de recherche interdisciplinaires et à des réseaux internationaux
- Pour la science : plus d'interdisciplinarité, reproductibilité de la recherche, renforcement de la qualité des données

Des freins

- Travail important pour bien décrire les données
- Manque de ressources humaines, de formation
- Critères d'évaluation de la recherche ne récompensant pas les efforts
- Réticence à partager les données si les auteurs ne sont pas crédités
- Crainte d'une mauvaise utilisation des données

FAIRiser les données in-situ des surfaces continentales

Le système d'information Theia/OZCAR



- Un portail unique pour l'ensemble des données in-situ des surfaces continentales
- Rendre ces données FAIR
- Utiliser les standards et thésaurus internationaux
- Etre interopérable avec les Infrastructures de Données et Services en construction au niveau national et international



Principaux défis :

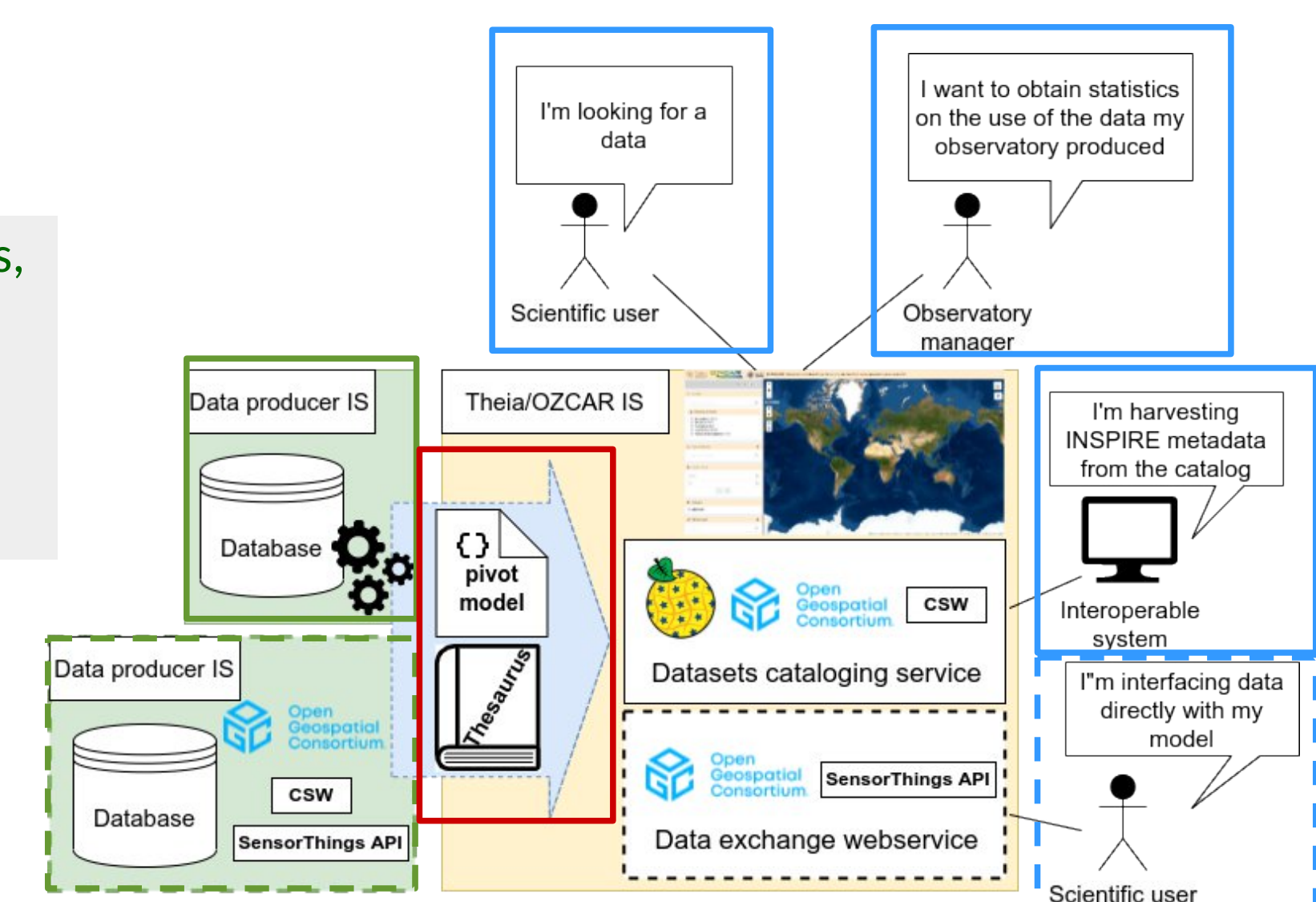
- Fédérer des données et des systèmes d'information hétérogènes pré-existants dans un système commun
- Rendre interopérable données et systèmes

Flux de données et services FAIR

Accessibilité des données à la fois pour les humains et les machines via

- un portail Web
- des services standardisés d'échange à la fois pour le catalogage de **jeux de données** (CSW) et pour l'échange de **données d'observation** (sensorThings API).

- Divers producteurs de données, différents vocabulaires, descriptions et formats, différentes maturités dans la gestion des données



- Modèle de données pivot + thesaurus Theia/OZCAR
 - Collecte et standardise les informations
 - Homogénéise les noms de variable
 - Est mis à jour en temps réel

3) Theia/OZCAR répond aux demandes des utilisateurs humains et machines

Comment FAIR ?

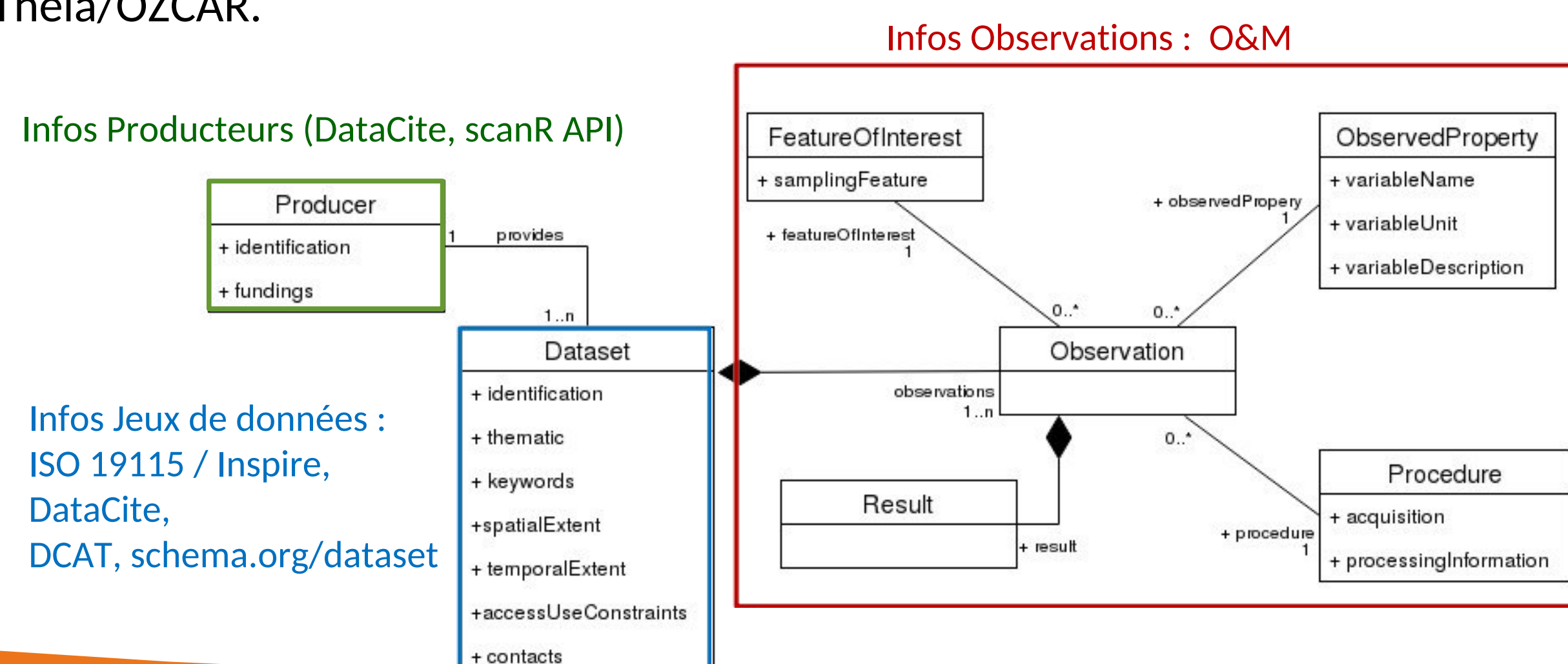
Un modèle de données pivot basé sur des standards

Objectifs : harmoniser et standardiser la description des (méta)données

Métadonnées **riches** avec les informations requises pour implémenter :

- Critères de recherche sur le portail: noms de variables, financeurs, ...
- Services standardisés d'échange (CSW: ISO 19115/Inspire, SensorThingsAPI : basé sur O&M)
- DOI pour les jeux de données (DataCite)

La description des variables observées référencent les termes du thesaurus Theia/OZCAR.



Autres éléments

- Portail Web Theia/OZCAR** : recherche par facette sur des critères utiles pour la communauté (variable, climat, géologie, financeur): <https://in-situ.theia-land.fr>
- Un vocabulaire contrôlé commun FAIR** : le thesaurus Theia/OZCAR publié en ligne (Voir poster associé)
- Format** pour la diffusion des données : ouvert, normalisé et auto-documenté (CSV, NetCDF)
- Un Plan de Gestion des Données (PGD)** rédigé en 2020 et mis à jour régulièrement, publié sur opidor : https://dmp.opidor.fr/public_plans
- Formation de la communauté au FAIR
 - Granularité des jeux de données
 - Recommandation de déclarer un DOI sur l'observatoire et les jeux de données
 - Recommandation d'utiliser ue licence **Creative Common CC-BY 4.0** + phrase de remerciement

Bénéfices

- Montée en qualité des données : meilleure description pour une réutilisation par soi-même et par les autres
- Visibilité et réutilisation accrue des données : plus de crédit pour les équipes qui collectent les données
- L'effort sur les vocabulaires permet d'aller vers l'interdisciplinarité
- DOI sur les jeux de données : données citables dans des publications scientifiques
- Licence sur les données : formalise les conditions de réutilisation des données ; offre un cadre juridique (pas les data policy). Protège les auteurs tout en garantissant les droits d'auteurs
- Fédère une communauté de gestionnaires et de producteurs de données autour des données de la zone critique