Projet Open OPSE

# Ecrire des règles de partage de données : quand, quoi, comment ?

## Séance participative

Laurence FLEURY et toute l'assistance !
IRD
laurence.fleury@ird.fr



CC BY SA

# Différents acteurs du partage de données

- Producteurs de données
- Utilisateurs de données
- Gestionnaires de données
- Juristes / Commissions nationales
- Etablissement / Employeur du producteur de données
- Bailleur de la collecte de données

# Différents acteurs = différents besoins

- Producteurs de données
  - Etre au courant des utilisations du jeu / y collaborer
  - **Etre cité / cosigner** les publications
- Utilisateurs de données
  - Disposer de données de **bonne qualité**, **le plus tôt possible**
  - Bénéficier de l'**expertise** du producteur (métadonnées, contact…)
- Gestionnaires de données
  - Disposer de règles de partage claires et **éviter les conflits**
- Juristes / Commissions nationales
  - Faire **respecter** la **loi** et les réglementations à divers niveaux
- Etablissement / Employeur du producteur de données
  - Etre **visible**
  - Ne pas encourir de poursuites
- Bailleur de la collecte de données
  - Etre **visible**

# D'autres acteurs ??? SOS

- Populations d'étude
  - Respect des termes du consentement
  - Respect de la dignité humaine et droits des personnes
- Grand public, enseignant, société civile
  - Besoin d'information
- Décideurs, Ministères
  - Besoin d'information ciblée
- '' Prédateurs ''
  - Réponse : ils ont les mêmes besoins que les utilisateurs mais contreviennent à l'éthique dans les usages. Pas de nouveaux besoins.

# D'autres besoins ??? SOS

- Bailleur de la collecte de données
  - Besoin de valorisation des données

# But des règles de partage de données

- Satisfaire les besoins de tous ces acteurs
  → Définir les droits et devoirs des différents acteurs
- Permettre un partage de données efficace
- Ces textes n'ont pas force de droit / sont dites non contraignantes. Ce sont plutôt des recommandations, des règles de bonnes pratiques...

# Mais c'est pas l'Open Data alors ?

**Quand** écrire des règles d'échange de données ?

- **PROJET** : projet en cours, données préliminaires, non publiées
  Exemple : **The DACCIWA Data and Publication Policy (2014)**

- **PORTAIL** : données finalisées,
  Exemple : **Règles condition d'accès aux jeux de données de la collection OPSE sur ANADS**

- **SENSIBLE** : partage de données personnelles dans le cadre de projets scientifiques.
  Exemple : ???

- **SOS ???**

# Jouons un peu …
# Quelles conditions dans quel cas ?

- Devoirs du producteur de données
  - Fournir les données
  - Fournir les données dans un délai fixé
  - Fournir la dernière version des données (en terme de qualité, de date,etc.)
  - Documenter chaque version des données
  - Fournir des données anonymisées

- Définition du/des cercles d'utilisateurs
  - Période d'embargo
  - Cercle de collaborateurs
  - Procédure de demande spécifique

# Quelles conditions dans quel cas ?

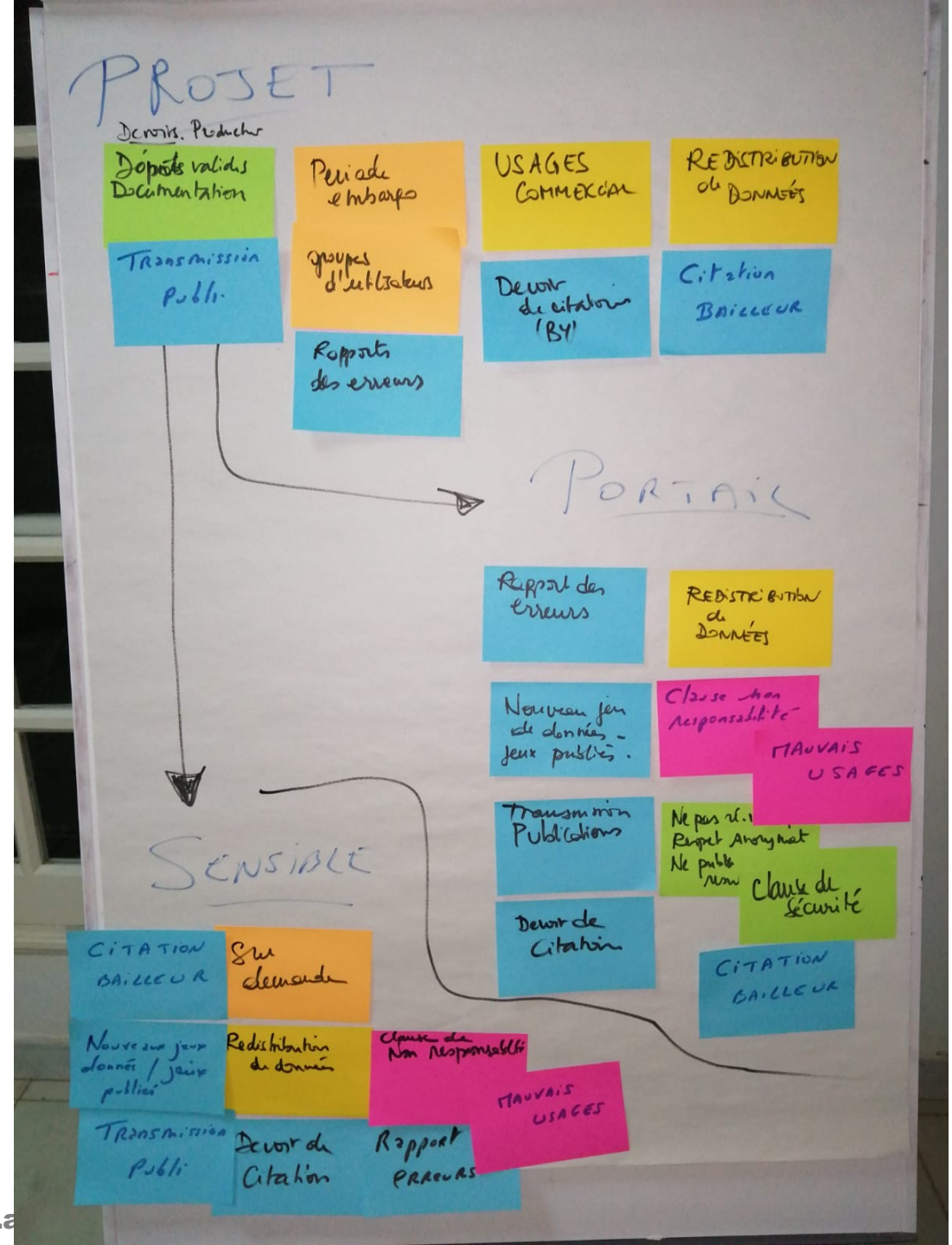
- Définition des usages autorisés des données
- Interdiction de redistribution des données
- Devoir de citation du producteur de données (cf CC BY)
  - Prise de contact
  - Proposer collaboration
  - Co-signature
  - Remerciements au minimum
- Citation du bailleur de la collecte de données
- Restitution des résultats de l'utilisateur
  - Transmission des publications → alimentation des métadonnées

# Quelles conditions dans quel cas ?

- Rapport des erreurs du jeu de données
- Nouveaux jeux de données produits
  - Distribution des jeux de données créés sur le même portail (cf. licence contaminante CC SA - ShareAlike)
  - Traitement des jeux multi-sources de données
  - Liens vers les jeux publiés en data paper → alimentation des métadonnées
- Clauses institutionnelles
  - Non responsabilité de l'institution
  - Action légale en cas de mauvais usage
- Données personnelles : clause de sécurité
  - Interdiction de ré-identifier
  - Respect de l'anonymat
  - Ne publier que des résultats agrégés

# D’autres conditions ??? SOS

# Projet Open OPSE



# Et quid de la rédaction des conditions

**SOS** : l'avis des juristes ?

- Préciser la juridiction dont relève le partage de données dans le document.
- Définir les termes utilisés dans le texte
- Quid si les données sont transférées à l'étranger ?
- Pourquoi ne pas faire un document contraignant ?
  - Réponse : licences CC contraignantes
  - Difficulté de contrôler les usages / d'engager une action en cas de mauvais usage

# Un exemple d'ouverture sous CC dans le cas portail

## Ouverture des données DACCIWA en fin de projet

LICENSE

This work is licensed under a **Creative Commons Attribution 4.0 International License**.

In case the data are used in research and publications we would appreciate when the following statement is added: "The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007-2013) under grant agreement n° 603502."

We would welcome data users to offer co-authorship to the data producers in case that the data make a substantial contribution to a publication. Please contact the main data PI to discuss this.

# D'autres conditions pour … le partage de données personnelles IDENTIFIANTES ???

**SOS**

- La ré-identification est la raison d'être d'un observatoire comme OPSE
- Partager des données nominatives est utile à de nombreux projets
- La position actuelle est à ma connaissance l'interdiction, alors qu'un encadrement et des recommandations sont nécessaires.

# Perpective 1 : finaliser les conditions d'accès à la collection OPSE du portail ANADS

- Merci de vos suggestions et réactions d'aujourd'hui
- Comment avancer ?
  - Groupe de travail
  - Espace d'écriture collaboratif en ligne
- Comment finaliser ?
  - Approbation par le comité de suivi OPSE (13 institutions partenaires représentées)

# Perpective 2 : comment partager des données personnelles identifiantes

- ????