

***Enjeux à prendre en compte
pour rédiger et mettre en
œuvre un DMP -
infrastructures et services
offerts par FORS***

Atelier Ra&D Domaine TS, HES·SO
Lausanne, le 02.12.2019

Pablo Diaz
Alexandra Stam

Plan de la présentation

1. Présentation de FORS
2. Qu'est-ce que la gestion des données?
3. Préparer et mettre en œuvre un DMP

Présentation de FORS

Qui sommes-nous?

Définition

- Fondation pour la recherche en sciences sociales / Centre de compétences suisse pour les sciences sociales
- Infrastructure localisée à l'Unil mais de portée nationale
- Financée principalement par le Confédération et par le FNS
- Prestations destinées prioritairement à la communauté des sciences sociales en Suisse et à l'étranger

Tâches

- Production de données dans le cadre d'enquêtes nationales et internationales
- Archivage et distribution de données pour l'analyse secondaire
- Recherche méthodologique, notamment dans le domaine des enquêtes
- Conseils dans ses domaines de compétences
- Collaborations internationales avec les institutions similaires

Activités principales

- Recherche méthodologique
- Grandes enquêtes
- Service de données et d'information sur la recherche (DARIS)

DARIS

FORSbase

Archivage de données

Nouvelles exigences

Préservation à LT

Valorisation des données

Accès aux données

Accès direct à:

+ 500 jeux de données

+ 11'000 descriptions de projets

Gestion des données

Formations

Consultations

Développement de matériaux

Services en matière de gestion des données

Nous offrons:

- Un soutien en gestion des données pendant la durée de projet de recherche
- Une plateforme d'archivage pendant et à l'issue du projet de recherche

BIENVENUE DANS FORSBASE

Votre portail d'accès aux données et aux informations sur la recherche en Suisse.

FORSbase est la plateforme virtuelle de FORS. Dédicée aux sciences sociales en Suisse, elle permet d'obtenir des données de recherche, ainsi que des informations sur les projets de recherche. Vous pouvez y déposer la description de votre propre projet, mais également archiver et partager les données récoltées dans le cadre de celui-ci.

Choisissez une des options ci-dessous ou commencez par découvrir les différentes fonctionnalités de FORSbase en consultant notre guide **FORSbase : vue d'ensemble**.

Ressources utiles

- Préparation des données quantitatives pour le dépôt
- Préparation des données qualitatives pour le dépôt
- Liste des formats de fichiers acceptés
- Glossaire pour la documentation des données
- Contrat de dépôt
- Contrat utilisateur
- FAQs
- Contact

FORS ⁺

Consulter les données sur Nesstar

<p>Explorer le catalogue</p> <p>Découvrez plus de 10'000 projets en sciences sociales décrits dans notre catalogue, constamment actualisé et entrez en contact avec les chercheurs et chercheuses qui y ont contribué.</p> <p>Guide : Découvrir le catalogue</p>	<p>Obtenir des données</p> <p>Parmi des centaines de jeux de données, téléchargez gratuitement des données pour vos différents projets et commencez sans tarder vos propres analyses.</p> <p>Guide : Obtenir des données</p>	<p>Publier votre projet</p> <p>Saisissez votre description de projet dans notre base de données pour gagner en visibilité, élargir votre réseau et développer de nouvelles collaborations.</p> <p>Guide : Publier votre projet</p>	<p>Déposer vos données</p> <p>Partagez vos données et assurez leur préservation à long terme. Vous répondez ainsi aux exigences de vos mandants et/ou des revues scientifiques tout en augmentant les citations de votre recherche.</p> <p>Guide : Archiver vos données</p>
---	---	---	--

<https://forsbase.unil.ch/>

Qu'est-ce que la gestion des données

Définition:

La *gestion des données* comprend toutes les activités associées aux données à l'exception de la récolte proprement dite et de l'utilisation (analyse). Elle couvre la manipulation, organisation, documentation et optimisation des données, ainsi que leur pérennisation et leur partage.

La gestion des données est constituée de différentes pratiques, telles que:

- Planification
- Considérations éthiques
- Préparation des données
- Documentation
- Anonymisation
- Organisation, stockage, sécurité
- Dissémination et copyright
- Partage des données

Cycle de vie de la recherche



Planification versus gestion au quotidien

Planification

- Vue d'ensemble
- Généralement bref
- Intentions de bonnes pratiques
- Accent particulier sur le partage
- « Problèmes » envisagés

ex. DMP

vs

Gestion des données au quotidien

- Mise en pratique de la gestion des données
- Stratégie détaillée
- Règles claires
- Accent sur les besoins immédiats en fonction du cycle de vie de la recherche
- Proposition de solutions

ex. formulation de règles, consentement

***Préparer et mettre en
oeuvre un DMP***

Le DMP comme ressource

- Doit être un instrument plus qu'une obligation
- Augmente la qualité (scientifique, matérielle et "éthique" des données)
- Permet d'engager la réflexivité sur des non-dits, des implicites et/ou des « routines »
- Permet de communiquer avec les bailleurs

Que sont les données ?

- Notion débattue
- « Données » VS « Obtenus »
- Les données sont toujours (co)construites

“Everything has the potential to be data, but nothing becomes data without the intervention of a researcher who takes note – and often makes note – of some things to the exclusion of others”

(Wolcott : 1994, pp. 3-4)



Définition des données de la recherche

“Les données de la recherche sont définies comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement **reconnus par la communauté scientifique** comme nécessaires pour valider les résultats de la recherche.”

(OCDE)

Comment classer/décrire les données ?

- Type
 - Degré d'intervention du·de la chercheur·e
 - Degré de traitement
 - Format
 - Nature juridique
- > ces éléments s'articulent de façon variable

Types de données générés par un projet

- Données de recherche

Matériaux utilisés pour administrer la preuve

- Données de contexte (documentation)

Informations nécessaires à comprendre/lire/utiliser les données de recherche « en connaissance de cause » (notes ethnographiques, codebook, etc.)

- Données administratives

Informations sur le projet en lui-même (PVs, contrats, mails, etc.)

Certains types de matériaux dépassent les typologies (ex: carnet de terrain)

Degré d'intervention du·de la chercheur·e

- Données “naturelles” (*naturalistic data*)

Traces/matériaux sur les activités humaines qui ne sont pas induits ni affectés par les actions des chercheur·e·s (journal intime, PV, article de presse)

VS

- Données « produites »

Traces/matériaux générés à travers l'intervention d'un·e chercheur·e (entretien, questionnaire, observation, etc.)

Niveau de traitement

- Données “brutes”

Matériaux tels qu’acquis/produits au moment du processus de recherche (enregistrements, questionnaires remplis, notes de terrain, etc.)

VS

- Données “traitées”

Données qui ont fait l’objet d’une transformation depuis leur production (transcriptions, données agrégées, tableaux, etc.)

-> On ne gère pas de la même manière un enregistrement qu’une retranscription

Format

- Numérique

Transcriptions, enregistrements vidéo/audio, scans, .docs, .pdfs, etc.

VS

- Physique

Photocopies, carnets, photos, livres/rapports, etc.

Nature « juridique »

- Données publiques
- Données personnelles
- Données sensibles
- Données classées secret défense
- Données soumises au droit d'auteur
- Etc.

Apprenez à connaître vos données !

- Les stratégies de gestion des données varient fortement en fonction du type de données collectées/produites/utilisées
- Les données personnelles/sensibles doivent être traitées avec soin
- Les données numériques sont plus « risquées » que les données physiques
- Les données digitales sont plus pérennes que les données physiques

Questions éthiques, légales et de sécurité

Deux précautions :

- Éthique \neq juridique

Ce qui est légal n'est pas forcément éthique (et vice versa)

- Protection \neq sécurité des données

On protège les gens et on sécurise les données

Éthique ou juridique ?

« Un jour... une belle journée ensoleillée, oui... en juillet 98, je me suis retrouvée entre les mains de policiers. Ils sont arrivés soudainement, ils m'ont fait monter brutalement dans une voiture et ils m'ont emmenée pour l'interrogatoire. Ils ont saisi tous mes documents de travail et ont commencé immédiatement à m'interroger sur cette recherche et sur les gens à qui j'avais parlé. Je leur ai dit que je travaillais beaucoup avec des groupes d'exclus et qu'un tel travail reposait sur le respect de l'anonymat des personnes interviewées, et qu'en ce sens il m'était impossible d'informer la police. Pour cette recherche, je m'étais d'ailleurs engagée expressément auprès des personnes interviewées, leur garantissant de ne pas révéler leur identité. **C'était un problème éthique pour moi** et une condition de l'enquête sociologique : **je ne pouvais pas donner les noms** ».

Sylvain Laurens & Frédéric Neyrat, Entretien avec Pinar Selek : "Je n'allais pas donner les noms, c'est une question d'éthique", in S. Laurens et F. Neyrat, Enquêter de quel droit : menaces sur l'enquête en sciences sociales, Editions du Croquant, 2010, pp. 235-242.

Éthique VS juridique

Cas Shamus Kahn:

« Several years before (...), on the advice of a lawyer, I had moved almost all of my field notes out of the country to a property I did not own. The logic was that these notes were potentially valuable to my future writing but contained a lot of material that I may not want to share. If they were not in my possession, not in the country, and not under my control, I was told, I could not be compelled to produce these materials »

(Khan, 2018)

Considérations légales

- C'est la loi du pays où la collecte est faite qui prime
- En UE : RGPD
- En Suisse : lois fédérales et cantonales : LPD, LPrD, LRH, etc.
 - **Données personnelles** : « toutes les informations qui se rapportent à une personne identifiée ou identifiable »
 - **Données sensibles** : « les données personnelles sur : les opinions ou activités religieuses, philosophiques, politiques ou syndicales, la santé la sphère intime ou l'appartenance à une race, des mesures d'aide sociale, des poursuites ou sanctions pénales et administratives ».
 - **Consentement**: obligatoire avec exceptions pour la recherche

Consentement et gestion des données

- Combien d'information sera donnée aux participants ?
 - Comment l'information sera-t-elle donnée ?
 - Comment le consentement sera obtenu ?
 - Quelle sera la portée du consentement (total ?)
- > Si pas de consentement il faut le justifier (ex. codes)
- > Le consentement doit être + « éthique » que « légal »

Anonymisation

- Anonymisation ≠ codage
 - L'anonymisation est définitive (mais difficile)
 - Le codage implique l'existence d'une clé
- Il n'y a pas de limitation concernant la nature et le caractère des données personnelles. Celles-ci peuvent être objectives ou subjectives.
 - Visage, opinions, style d'écriture, démarche, etc.

Les identifiants

- Identifiants **indirects** : informations qui suffisent à elles seules à identifier une personne. Ex : nom, adresse mail nominale, identifiants biométriques, etc.
- Identifiants **indirects forts** : informations qui facilitent grandement l'identification d'une personne. Ex : adresse, numéro de téléphone, numéro de compte, adresse IP, adresse mail (sans nom), titre de travail inhabituel, maladie rare, poste unique, distinction, etc.
- Identifiants **indirects faibles** : informations qui, à elles seules ne suffisent pas à identifier une personne mais qui combinées peuvent aider à déduire l'identité. Ex : âge, formation, statut, composition du ménage, revenu, situation matrimoniale, langue, origine, etc.

Trouver l'équilibre par les risques

- Déterminer les **impacts potentiels** sur la vie privée des personnes concernées
- Estimer la **gravité** potentielle, c'est-à-dire l'ampleur du risque (négligeable, limitée, importante, maximale).
- Estimer la **vraisemblance**, à savoir la possibilité qu'un risque se réalise (négligeable, limitée, importante, maximale). Elle dépend principalement des **vulnérabilités** des supports et des **sources** de risque.

<https://www.cnil.fr/fr/outil-pia-telechargez-et-installez-le-logiciel-de-la-cnil>

Stockage et préservation des données

- Importance des formats (court VS long-terme)
 - .txt, .pdfa, .csv, .tif, .wav, etc.
- Bien choisir les emplacements de sauvegarde (Suisse)
- Bien choisir les plateformes de partage (interne)
- Réfléchir à l'arborescence

Sécurité des données

- Beaucoup d'instruments vont à l'encontre des droits des participant.e.s
 - Collecte (téléphones, questionnaires en ligne, etc.)
 - Traitement (transcript en ligne, pdf merger, etc.)
 - Stockage (cloud, etc.)
- On distingue trois types de mesures (à combiner) :
 - Les mesures portant spécifiquement sur les données
 - Les mesures générales de sécurité du système de traitement
 - Les mesures organisationnelles (gouvernance)

Partage et réutilisation des données : FAIR

Findable

Accessible

Interoperable

Reusable

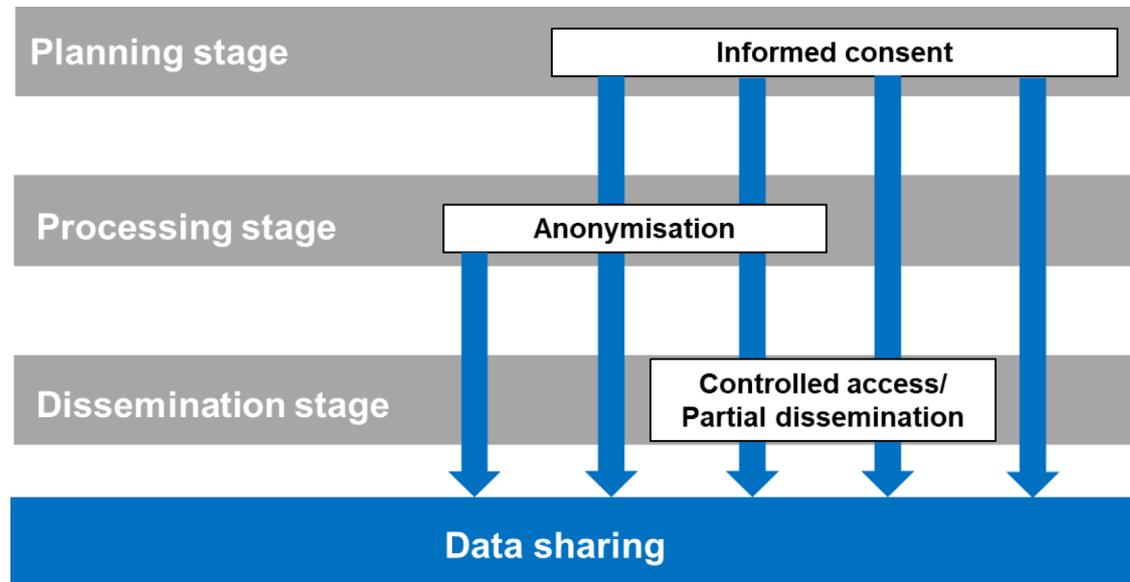
“This is a key, but often misunderstood, element of FAIR. The ‘A’ in FAIR does not necessarily mean ‘open’ or ‘free’. Rather, it implies that one should provide the exact conditions under which the data may be accessed. Hence, even heavily protected and private data can be FAIR”.

▪ <https://www.go-fair.org/fair-principles/a1-2-protocol-allows-authentication-authorisation-required/>

Partage et réutilisation des données

- Anticiper l'archivage
- Anticiper les (ré)utilisations possibles (documentation)
- Bien choisir le référentiel
- Identifier ce qui est (ou non) partageable (on ne peut pas aller contre la loi !)
- Bien gérer les données non partageables (cas Salzbrunn)

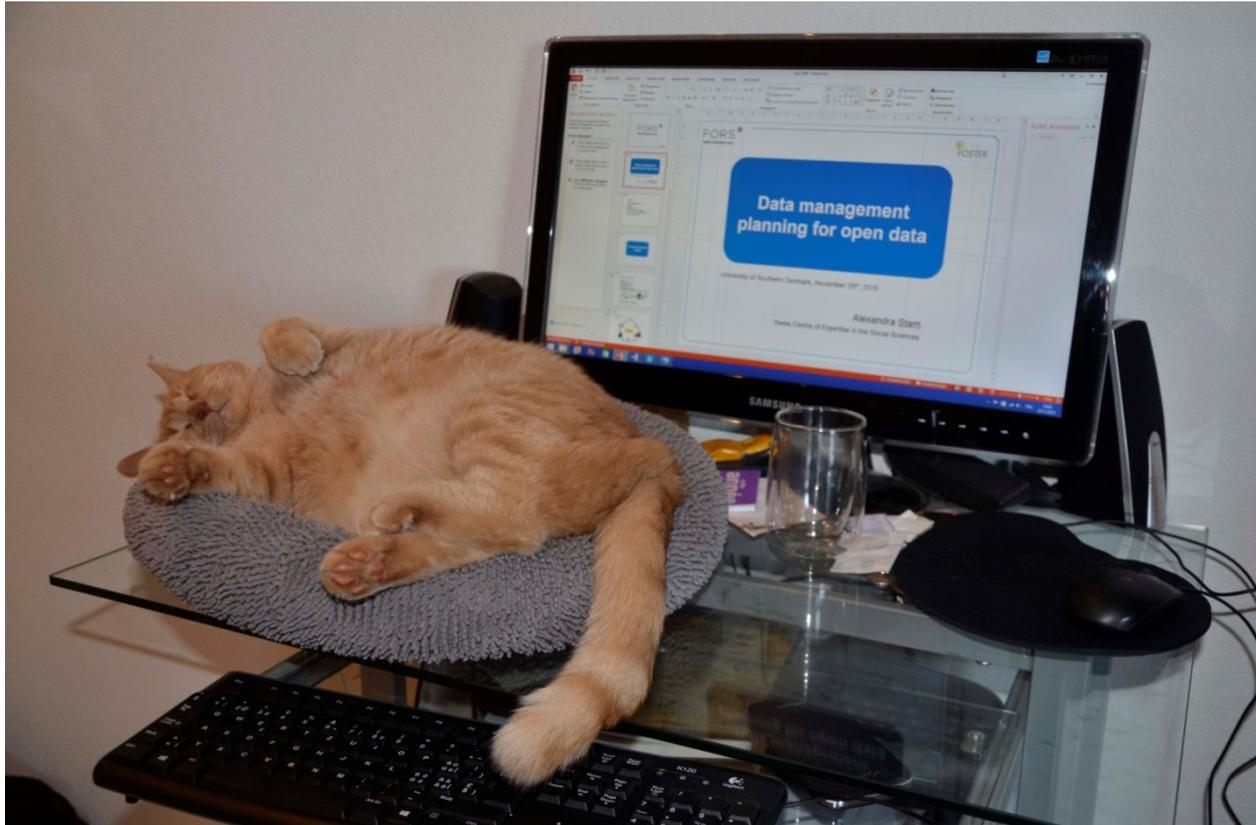
Adopter une approche par couches



Conclusion

Il s'agit de trouver un équilibre entre:

- Protection & partage
- Ouverture & accessibilité
- Science & bien-être des participants



Merci pour votre attention

alexandra.stam@fors.unil.ch

PabloAndres.DiazVenegas@unil.ch