

Retour d'Expérience L'Écosystème TRF-GIS

Victor Gay

University of Toulouse 1 Capitole

Toulouse School of Economics

Institute for Advanced Study in Toulouse

Novembre 2020



Toulouse
School of
Economics

- L'écosystème TRF-GIS: un data paper et des données
- Pourquoi un data paper?
- La rédaction du data paper
- Les données FAIR en pratique
- La dissémination et la valorisation
- Quels rôles pour les fonctions support?

L'écosystème TRF-GIS: un data paper et des données

- Le TRF-GIS: un SIG pour la France de la Troisième République.

- Le TRF-GIS: un SIG pour la France de la Troisième République.
- Nomenclatures et shapefiles annuels (1870–1940)
 - Circonscriptions générales: départements, arrondissements, cantons, communes.
 - Circonscriptions spéciales: militaires, judiciaires, pénitentiaires, électorales, inspections du travail, ecclésiastiques.

L'écosystème TRF-GIS: des données

- Le TRF-GIS: un SIG pour la France de la Troisième République.
- Nomenclatures et shapefiles annuels (1870–1940)
 - Circonscriptions générales: départements, arrondissements, cantons, communes.
 - Circonscriptions spéciales: militaires, judiciaires, pénitentiaires, électorales, inspections du travail, ecclésiastiques.
- 16 bases de données
 - 901 nomenclatures, 830 shapefiles.
 - Matériel de reproduction: code source, données source, archives.

L'écosystème TRF-GIS: des données

- Le TRF-GIS: un SIG pour la France de la Troisième République.
- Nomenclatures et shapefiles annuels (1870–1940)
 - Circonscriptions générales: départements, arrondissements, cantons, communes.
 - Circonscriptions spéciales: militaires, judiciaires, pénitentiaires, électorales, inspections du travail, ecclésiastiques.
- 16 bases de données
 - 901 nomenclatures, 830 shapefiles.
 - Matériel de reproduction: code source, données source, archives.
- Accessible sur <https://dataverse.harvard.edu/dataverse/TRF-GIS>.

- “Mapping the Third Republic. A Geographic Information System of France (1870–1940).” Octobre 2020.

- “Mapping the Third Republic. A Geographic Information System of France (1870–1940).” Octobre 2020.
- Contenu:
 - Éléments de contextualisation
 - Description détaillée de la méthodologie technique et historique
 - Description détaillée des fichiers de données
 - Quelques conseils d'utilisation et usages potentiels

- “Mapping the Third Republic. A Geographic Information System of France (1870–1940).” Octobre 2020.
- Contenu:
 - Éléments de contextualisation
 - Description détaillée de la méthodologie technique et historique
 - Description détaillée des fichiers de données
 - Quelques conseils d'utilisation et usages potentiels
- Accessible sur <https://hal.archives-ouvertes.fr/hal-02951461>.

Pourquoi un data paper?

Pourquoi un data paper?

- Une réponse à la crise de reproductibilité en sciences (sociales).
 - État de la recherche empirique sur la France de la 3e République:
 - Quelques shapefiles départements et cantons, mais pas FAIR.
 - Pas de cadre commun (nomenclatures), pas de documentation.
 - Perte de temps, peu d'interopérabilité, peu de crédibilité.
- ⇒ Le TRF-GIS établit un cadre commun documenté et reproductible

Pourquoi un data paper?

- Une réponse à la crise de reproductibilité en sciences (sociales).
 - État de la recherche empirique sur la France de la 3e République:
 - Quelques shapefiles départements et cantons, mais pas FAIR.
 - Pas de cadre commun (nomenclatures), pas de documentation.
 - Perte de temps, peu d'interopérabilité, peu de crédibilité.
- ⇒ Le TRF-GIS établit un cadre commun documenté et reproductible
- Une réponse au manque de valorisation du travail sur les données
 - Peu d'incitations pour l'évaluation de la recherche.
 - Peu de valorisation des créateurs de données.
 - Pari sur la création d'un bien public avec externalités positives.

La rédaction du data paper

La rédaction du data paper

- Le data paper n'est pas qu'un codebook descriptif.
- Description détaillée sur plusieurs plans
 - Méthodologie de construction
 - Données source
 - Choix de catégorisation
 - Limites et valeur par rapport à l'état de l'art
 - Éléments techniques
 - Réutilisation

La rédaction du data paper

- Le data paper n'est pas qu'un codebook descriptif.
- Description détaillée sur plusieurs plans
 - Méthodologie de construction
 - Données source
 - Choix de catégorisation
 - Limites et valeur par rapport à l'état de l'art
 - Éléments techniques
 - Réutilisation
- Plusieurs modèles possibles, format évolutif (ici: *Scientific Data*).

La rédaction du data paper

- Le data paper n'est pas qu'un codebook descriptif.
- Description détaillée sur plusieurs plans
 - Méthodologie de construction
 - Données source
 - Choix de catégorisation
 - Limites et valeur par rapport à l'état de l'art
 - Éléments techniques
 - Réutilisation
- Plusieurs modèles possibles, format évolutif (ici: *Scientific Data*).
- La conception des données et du data paper doit être simultanée.
- Un travail conséquent!

Les données FAIR en pratique

- Quelques éléments FAIR du TRF-GIS
 - Métadonnées riches (DDI, Dublin Core, DataCite, ...).
 - Vocabulaires contrôlés pour mots-clés (LCSH).
 - Identifiants persistents pour chaque fichier (DOI).
 - Formats pérennes et ouverts (csv, txt, shapefiles).
 - Compatibilité avec standards répandus (COG INSEE, GEOFLA IGN).
 - Compatibilité avec données historiques disponibles (CRH).
 - Licence appropriée (CC-BY 4.0).
 - Versionnage précis.
- Beaucoup d'éléments gérés par le dépôt de données (Dataverse).
- De nombreux choix à faire en amont.

La dissémination et la valorisation

La dissémination et la valorisation: le data paper

- Indissociable des données.

La dissémination et la valorisation: le data paper

- Indissociable des données.
- Un dépôt central et pérenne: HAL sous license CC-BY 4.0.

La dissémination et la valorisation: le data paper

- Indissociable des données.
- Un dépôt central et pérenne: HAL sous license CC-BY 4.0.
- Le document de travail:
 - Séries institutionnelles: TSE, IAST, UT1.
 - Séries disciplinaires: MPRA, AFC, EHES.
 - (Séries génériques: SSRN, Academia.)
 - Moissonneurs: Google Scholar, IDEAS/RePEc, ORCID, Isidore.

La dissémination et la valorisation: le data paper

- Indissociable des données.
- Un dépôt central et pérenne: HAL sous license CC-BY 4.0.
- Le document de travail:
 - Séries institutionnelles: TSE, IAST, UT1.
 - Séries disciplinaires: MPRA, AFC, EHES.
 - (Séries génériques: SSRN, Academia.)
 - Moissonneurs: Google Scholar, IDEAS/RePEc, ORCID, Isidore.
- La revue à comité de lecture: format, lecteurs, portée?
 - *Scientific Data*
 - *Historical Methods*
 - *Research Data Journal for the Humanities and Social Sciences*
 - *Histoire et Mesure*

La dissémination et la valorisation: les données

- Indissociable du data paper.

- Indissociable du data paper.
- Choix du dépôt de données: critères, besoins, usages?
 - Respect (et gestion) des principes FAIR.
 - Espace de stockage (13K fichiers, 6GB).
 - Ergonomie producteur et utilisateur (structure complexe).
 - Visibilité.
 - Générique, disciplinaire, interne?
 - Possibilités: Harvard Dataverse, Nakala, Figshare,...

La dissémination et la valorisation: les données

- Indissociable du data paper.
- Choix du dépôt de données: critères, besoins, usages?
 - Respect (et gestion) des principes FAIR.
 - Espace de stockage (13K fichiers, 6GB).
 - Ergonomie producteur et utilisateur (structure complexe).
 - Visibilité.
 - Générique, disciplinaire, interne?
 - Possibilités: Harvard Dataverse, Nakala, Figshare,...
- La dissémination: plusieurs vecteurs
 - Portails et agrégateurs: data.gouv.fr, PROGEDO.
 - Logiciels: Khartis, Magrit, Stata, R.

Quels rôles pour les fonctions support?

Quels rôles pour les fonctions support?

- Large spectre de connaissances nécessaires.
- Rôle primordial de la formation
 - Pas d'enseignement à l'école doctorale.
 - Suivi de formations continues (URFIST, MSHS-T, DoRANum).

Quels rôles pour les fonctions support?

- Large spectre de connaissances nécessaires.
- Rôle primordial de la formation
 - Pas d'enseignement à l'école doctorale.
 - Suivi de formations continues (URFIST, MSHS-T, DoRANum).
- Pas d'appui local (UT1: économie et droit très en retard).
⇒ Besoin démarche proactive des deux parties.

Quels rôles pour les fonctions support?

- Large spectre de connaissances nécessaires.
- Rôle primordial de la formation
 - Pas d'enseignement à l'école doctorale.
 - Suivi de formations continues (URFIST, MSHS-T, DoRANum).
- Pas d'appui local (UT1: économie et droit très en retard).
⇒ Besoin démarche proactive des deux parties.
- Un travail conjoint chercheurs / fonctions support?
 - Spécificités scientifiques disciplinaires: usages, diffusion,...
 - Des structures d'incitation et responsabilité incompatibles?

- TRF-GIS Dataverse
<https://dataverse.harvard.edu/dataverse/TRF-GIS>
- “Mapping the Third Republic”
<https://hal.archives-ouvertes.fr/hal-02951461>
- Contact:
 - Email: victor.gay@tse-fr.eu
 - Site personnel: <https://www.victorgay.me>
 - Twitter: @victorgayeco