



**HAL**  
open science

# Plateformisation, bancarisation des données de la recherche : un "mal" pour un "bien" ? Réflexions à partir du cas de beQuali

Guillaume Garcia

## ► To cite this version:

Guillaume Garcia. Plateformisation, bancarisation des données de la recherche : un "mal" pour un "bien" ? Réflexions à partir du cas de beQuali. Séminaire GERiiCO "Plateformisation de la science ouverte", GERiiCO, Jun 2022, Lille, France. hal-04030589

**HAL Id: hal-04030589**

**<https://sciencespo.hal.science/hal-04030589>**

Submitted on 15 Mar 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Platformisation, bancarisation des données de la recherche :  
un “mal” pour un “bien” ?  
Réflexions à partir du cas de beQuali**

*Guillaume Garcia, Centre de données sociopolitiques*

Séminaire GERiCO

*Platformisation de la science ouverte*

Université de Lille, vendredi 3 juin 2022

## L'arrière-plan de la plateformisation / bancarisation des données

Encore peu de travaux (largement diffusés) sur la question

Pendant longtemps, un idéal libertarien assez confidentiel en sciences sociales

Aujourd'hui, dans les sciences sociales pratiquant l'enquête de terrain :

- élan +/- spontané vers la science ouverte "institutionnelle"
- une communauté scientifique partagée entre indifférence et réticences faces à ce qui est perçu avant tout comme des logiques managériales & top down
- des conflits de métiers qui se superposent à ces tensions ?

Des discours parasités par des *a priori* normatifs

⇒ nécessité d'ouvrir la "boîte noire" de la construction et du fonctionnement des entrepôts de données

## Un point de vue situé

Celui d'un acteur de la diffusion des données d'enquête

Un positionnement spécifique (enquêtes de terrain en sociologie / science politique)

Une multitude d'autres points de vue possibles, de plateformes françaises...

- Cocoon
- Parchemins
- Phonothèque MMSH
- Odsas
- Ortolang
- Sawa / Watau
- Telemeta
- etc.

et bien sûr de plateformes à l'étranger

# Rappels sur beQuali

## A l'origine :

- un projet expérimental
- centré sur l'analyse secondaire
- à une échelle localisée (Sciences Po)

## Devenu un module d'une infrastructure de recherche

- multiples soutiens des tutelles (FNSP / CNRS)
- Equipex DIME-SHS (2011-2020)
- inscription au sein du CDSP & IR PROGEDO

## Élargissement des objectifs

- mise en avant des usages pour l'enseignement
- élargissement des usages pour la recherche

The screenshot shows the beQuali website interface. At the top, there is a navigation bar with the text "Banque d'enquêtes qualitatives en sciences humaines et sociales", the beQuali logo, and options for "English" and "Connexion". Below the navigation bar, there are menu items: "À PROPOS", "NOS SERVICES", "ENQUÊTES", "RÉUTILISER LES DONNÉES", "DÉPOSER UNE ENQUÊTE", "BIBLIOGRAPHIE", and "NEWSLETTER". The main content area features a large heading "Banque d'enquêtes qualitatives en sciences humaines et sociales" and a sub-heading "Banque d'enquêtes qualitatives en sciences humaines et sociales". Below this, there is a paragraph describing the platform's mission: "BeQuali propose un catalogue d'enquêtes de sciences humaines et sociales réalisées à partir de méthodes qualitatives (observations, entretiens, focus groups, etc.). Ce portail met à disposition de la communauté scientifique, sous réserve d'autorisation, des matériaux bruts et toute une documentation restituant le contexte de leur production. BeQuali s'inscrit dans une démarche de préservation des données de la recherche et propose des ressources et outils communs pour l'enseignement des méthodes et la production de nouvelles recherches." Below the text, there are three featured content blocks: "RECHERCHE AVANCÉE DE CORPUS" with a bar chart, "AFFICHAGE DYNAMIQUE DES ENTRETIENS" with a list of interviews, and "DECODAGE SCIENTIFIQUE DES ENQUÊTES" with a search interface. On the right side, there is a sidebar with a dark background and white text, containing several news items with dates: "Semaine DATA SHS" (01 Déc.), "beQuali à la journée d'études EPHN2020" (20 Sep.), "Nouvelle enquête en ligne" (22 Avr.), and "Enquête en cours de traitement" (08 Avr.).

# Aujourd'hui : un catalogue d'une vingtaine d'enquêtes

<b>Choisir son école</b>	<b>Quand des Français, des Anglais et des Belges (francophones) parlent d'Europe</b>	<b>Les Français et la politique</b>
<i>VAN ZANTEN Agnès</i>	<i>Sophie DUCHESNE, Elizabeth FRAZER, Guillaume G...</i>	<i>SCHWEISGUTH Estienne, Enquêteurs professionnels</i>
France 1999-2005	France, Belgique, Royaume-Uni 2003-2006	France 1983

<b>Le Parlement français et l'Union Européenne : l'Europe saisie par les rôles parlementaires</b>	<b>Représentation du champ social, attitudes politiques et changements socio-économiques</b>	<b>Des femmes en politique</b>
<i>ROZENBERG Olivier</i>	<i>DONÉGANI Jean-Martin, MICHELAT Guy, SIMON Michel</i>	<i>SINEAU Martette</i>
France 1998-2005	France 1978	France 1984-1985

<b>Les infirmières et leur coordination</b>	<b>Avec ceux du FN</b>	<b>Les conditions de vie professionnelle des femmes enceintes</b>
<i>KERGOAT Danièle, IMBERT Françoise, LE DOARE ...</i>	<i>BIZEUL Daniel</i>	<i>DEVREUX Anne-Martin</i>
France 1988-1995	France 1995-2003	France 1985-1988

<b>Espace et temps du travail domestique (Publication à venir)</b>	<b>Panel de Coen (Publication à venir)</b>	<b>La formation des couples</b>
<i>Danièle Chabaud-Rychter, Dominique Fougeyrolles-Schweibel...</i>	<i>Claire Bidart (coord.)</i>	<i>Michel BOZON, François HÉRAN, Marie-Laurence...</i>
France 1977-1981	France 1995-2015	France 1985-1985

<b>La profession d'architecte : nouvelle théorie sociologique des professions</b>	<b>Comparaison des ministères de l'enseignement supérieur de France et d'Allemagne</b>	<b>De l'Afrique à la France, d'une génération à l'autre</b>
<i>CHAMPY Florent</i>	<i>MUSSELIN Christine, FRIEDBERG Erhard</i>	<i>BAROU Jacques (coord.), AOUICI Sabrina, ATTIAS-DONF...</i>
France, Etats-Unis 2004-2010	France, Allemagne 1987-1990	France, Afrique sub-saharienne (A... 2006-2010

<b>Dilapidation et prodigalité</b>	<b>Les réseaux économiques souterrains en cité de transit</b>	<b>Le sens de l'hospitalité</b>
<i>GOTMAN Anne</i>	<i>LAÉ Jean-François, MURARD Numa, DESHAYES Fo...</i>	<i>GOTMAN Anne</i>
France 1989-1995	France 1981-2010	France 1995-2001

<b>La gouvernance territoriale en Europe entre convergence et capacité</b>	<b>Operationalizing Programmatic Elites Research in America (Publication à venir)</b>
<i>Alistair COLE (coord.), Christian DE VISSCHER, Jean-B...</i>	<i>GENIEYS William (coord.), BEAUSSIER Anne-Laure, DARV...</i>
Belgique, Espagne, France, Royau... 2012-2014	Etats-Unis 2006-2012

## Caractéristiques du catalogue

des enquêtes essentiellement de sociologie et de science politique

réalisées plutôt par entretiens, et secondairement par observations ethnographiques + quelques enquêtes de méthodes mixtes (“ quanti / quali ”)

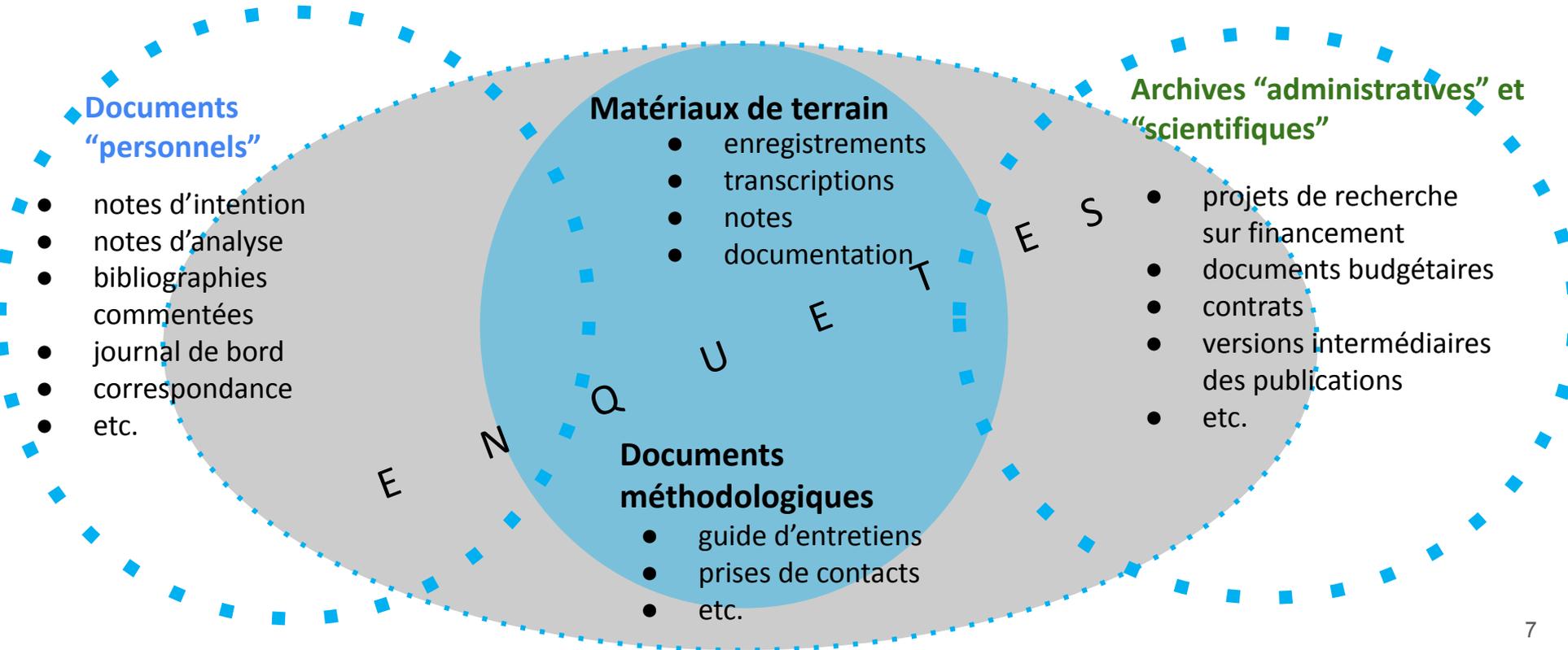
souvent (en partie) à numériser, plus rarement nativement numériques

certaines sont connues des spécialistes du domaine : *Choisir son école* (Van Zanten) ; *La formation des couples* (Bozon, Héran) ; *Avec ceux du FN* (Bizeul) ; *Panel de Caen* (Bidart), etc.

un “meta” corpus de 6 000 documents

- + 1 500 transcriptions d’entretiens
- + 150 notes d’observation

## Une conception large des corpus d'enquêtes



## S'engager dans un processus de plateformisation

Enjeux institutionnels / de financement difficilement résistibles... lorsqu'on veut dépasser le stade de l'expérimentation

Enjeux de pérennisation : le nécessaire recours

- à des outils informatiques maintenables
- à des standards / normes métiers
- à des protocoles de travail suffisamment stables

Enjeux de réassurance / mise en confiance des déposants

Avantages et limites des projets ad hoc, localisés & individualisés

- une flexibilité et une adaptabilité plus importantes à la spécificité des projets
- tout dépend des facteurs suivant : motivations, temps, ressources, compétences... et au-delà d'enjeux de carrière
- une probabilité de visibilité dépendant d'un capital de notoriété inégalement réparti
- un soutenabilité incertaine

# Protocoles et standards : un niveau minimal nécessaire

La nécessité d'obtenir des fichiers numériques sous des formats pérennes

Redonner de l'intelligibilité aux corpus

- reclassement et renommage des fichiers (pièce à pièce)
- le plus difficile : trouver le bon équilibre entre respect du corpus original et sa mise en compréhension pour des tiers

La description des corpus via des métadonnées suffisamment standardisées

- enjeux : moissonnabilité & interopérabilité

Documenter suffisamment le contexte de la recherche

- s'assurer de renseigner la préparation du terrain et l'analyse des matériaux
- via les archives de l'enquête & des "explications de texte"

Se doter de repères pour anonymiser au mieux les matériaux

- pour se faciliter la tâche & garantir une certaine cohérence / rationalité au sein d'une même enquête, et d'une enquête à l'autre
- s'assurer de limiter les ruptures de confidentialité & la non falsification des données

Garantir la gestion pérenne des accès dans le temps

## Un entrepôt & des fonctionnalités

### Des objectifs à géométrie variable

- du simple “zip” aux fonctionnalités d’exploration...
- ... au sein d’un corpus voire entre plusieurs corpus de données

### beQuali : une voie médiane

- des outils d’exploration mis en place à l’intérieur d’une enquête
- dont le fonctionnement était conditionné par une mise en équivalence minimale des enquêtes
- via un plan de classement à la fois générique et adaptable à chaque enquête

### Le refus de détacher les documents du contexte de la recherche

⇒ impossibilité de faire des requêtes permettant de sortir des documents qui seraient détachés de leur corpus d’appartenance

# Arborescence du plan de classement beQuali

1e niveau Sous-fonds		2e niveau Dossier		3e niveau Sous-dossier		4e niveau Article	
[ADD] Compléments	[ADMI] Documents administratifs	[ANXE] Observations annexes		[ACTES] Actes de colloques			
[ANA] Analyse	[ARCHIVES] Archives	[BRCH] Brochures		[ARTCL] Articles			
[COL] Collecte	[DOCU] Documentation rassemblée en amont ou pendant terrain	[BUDG] Budget		[AUDIO] Enregistrements audio			
[PRE] Préparation	[ENTR] Entretiens	[CODE] Codage		[BIBLI] Bibliographie			
	[INTE] Documents intermédiaires	[COMM] Communications		[BRLN] Brouillons			
	[METH] Méthodologie	[DICO] Dictionnaire / Encyclopédie		[CARTE] Cartes			
	[OBSE] Observations	[ECHT] Recrutement et suivi de l'échantillon		[CHAPT] Chapitres			
	[PROD] Production scientifique	[ELAQ] Élaboration du questionnaire		[CMPPTA] Pièces comptables			
	[QNTI] Volet quantitatif	[ENRG] Enregistrements / documents multimédia		[CMPTR] Compte-rendu			
		[ENSG] Enseignement		[COURS] Cours			
		[EXQL] Exploitation des données qualitatives		[COUVT] Couverture			
		[EXQT] Exploitation des données quantitatives		[CPDCE] Correspondance			
		[FIGU] Documents figurés (images, graphique, histogrammes, tableaux etc)		[EXAME] Sujet d'examen			
		[FOCU] Focus-group		[EXPOS] Textes d'exposés			
		[FORM] Recrutement et formation des enquêteurs		[FICHE] Fiches			
		[INDV] Entretiens individuels		[GRAPH] Graphiques			
				[GUIDE] Guides / instructions / protocole / argumentaire / plan d'exploitation			
		[LGRI] Littérature grise		[INDEX] Index			
		[POST] Entretiens post-quant		[LISTE] Listes			
		[PRDQ] Presse / périodique		[MEMOI] Mémoires / thèses			
		[PREL] Entretiens préliminaires		[NOMCL] Nomenclature			
		[PROG] Programmation		[NOTES] Notes			
		[PUBL] Publications		[OFFIC] Texte juridique / réglementaire / publications officielles			
		[QFAF] Enquête par questionnaire en face à face		[OUVRG] Ouvrage			
		[QOUV] Relevés de réponses aux questions ouvertes		[PHOTO] Photographies			
		[TERR] Observation du terrain principal (journal, relevés)		[PLANS] Plans / sommaires			
				[PRCRM] Programme de colloque / conférences			
				[QUEST] Formulaires / questionnaires			
				[RESUM] Résumés			
				[RPPRT] Rapports / synthèses			
				[TABLE] Tableau de suivi			
				[TRACT] Prospectus			
				[TRANS] Transcriptions d'entretiens			

# Le moteur de recherche à facettes (sur l'enquête La formation des couples)

Explorer le corpus via le classement analytique ⓘ

1. Filtrer selon les étapes du processus de recherche ⓘ  

45	276 Collecte	156 Analyse
----	--------------	-------------
2. Filtrer selon le type d'activité ⓘ  

20	9	98 Méthodologie	24	181 Entretiens	9	66 Observations	70
----	---	-----------------	----	----------------	---	-----------------	----
3. Filtrer selon le type de documents  

20	417	61 Publications	58	55	146 Entretiens post-quantis	7	7	60	42	Codage	9	9	18
----	-----	-----------------	----	----	-----------------------------	---	---	----	----	--------	---	---	----
4. Filtrer selon le type d'articles  

35	18	54 Notes	11	15	15	9	78 Fiches	100	76	Compte-rendu	15	21	14	6
----	----	----------	----	----	----	---	-----------	-----	----	--------------	----	----	----	---

Explorer le corpus via la localisation géographique ⓘ

Inclure aux résultats les fichiers non localisés

[-] [+]

[Réinitialiser les filtres](#)

477/477 Documents

[Exporter la sélection](#)

Brouillon pour le projet d'enquête

[PDF](#) [PARIS](#)

Réflexion sur l'enquête de l'Ined de 1959 dans la perspective d'une nouvelle enquête sur ce thème.

[PDF](#) [06/1982](#) [PARIS](#)

Programme des enquêtes statistiques pour 1983.

[PDF](#) [1983](#) [PARIS](#)

## Explorer le corpus via le classement analytique ?

### 1. Filtrer selon les étapes du processus de recherche ?



### 2. Filtrer selon le type d'activité ?



### 3. Filtrer selon le type de documents



### 4. **54/54 documents Notes (notes)**



[Réinitialiser les filtres](#)

54/477  
Documents

[Exporter la sélection](#)

Notes sur la programmation  
des entretiens non directifs

[PDF](#) [PARIS](#)

Notes à propos des remarques  
de B. Riandey sur l'enquête

[PDF](#) [PARIS](#)

## La nécessité de se doter d'une équipe pluri-métier

### Nécessité de compétences diversifiées & de personnels qualifiés

- en archivistique
- en documentation
- en informatique
- en recherche en science sociales

A beQuali on est passé progressivement de 1 à 7 personnes entre 2011 et 2019... avant de redescendre rapidement à 1 ETP (en 2022)

### Des alignements d'intérêt nécessaires mais difficiles

- des processus frictionnels...
- ... mettant parfois en jeu des conflits de métier
- aboutissant à des effets de composition & une production nécessairement "hybride"

## Quelles évolutions d'un tel dispositif ?

Anticiper l'obsolescence programmée de l'outil beQuali

⇒ le passage progressif à Dataverse

S'adapter à la diminution des ressources RH

⇒ investir la formule de l'accompagnement au dépôt

Une moindre "normativité" dans les protocoles de traitement des enquêtes

- persistance d'exigences élevées en matière de métadonnées descriptives
- fin du reclassement & du renommage "imposés"
- accompagnement des déposants pour le classement / l'anonymisation / la documentation des données

Un recentrement sur les enquêtes "non (ou moins) problématiques"

⇒ corpus plus "modestes", plutôt nativement numériques, ne posant pas de problèmes majeurs de confidentialité des données

## Bilan : un “mal” pour un “bien” ?

Est-il possible d'échapper à la plateformesation ?  
dans une certaine mesure sans doute, mais à une échelle large cela paraît douteux...

Un débat à consolider

- un niveau élevé d'abstraction, des positions souvent de principe, des idéaux assez confus
- un flottement terminologique et des figures repoussoirs (la “bancaisation”, la “standardisation”...)

Une communauté de recherche “hétéroclite” face aux enjeux de la science ouverte

Que peuvent nous apprendre les exemples étrangers ?

- l'auto-dépôt ne fonctionne pas (ou très peu)
- le risque de surinvestir les réutilisations pour la recherche / à des fins de reproductibilité, au détriment d'autres formes tout aussi légitimes et utiles, mais considérées comme moins “nobles” scientifiquement parlant
- une tension entre logique extensive et de standardisation (faire du nombre, ce qui implique de standardiser beaucoup les processus, à niveau de ressources constant) *versus* logique intensive et artisanale (traiter peu de corpus, le plus possible à façon)

Une confusion entre les contraintes intrinsèques de l'archivage des données & les contraintes imposées par les “infrastructures” ?

Un point d'achoppement majeur : faire en sorte que les “bonnes pratiques” ne soient pas définies à l'avance...  
ce qui est contraire au principe même d'établissement des bonnes pratiques

**Merci de votre attention !**

**Contacts :**

[guillaume.garcia@sciencespo.fr](mailto:guillaume.garcia@sciencespo.fr)

[info.cdsp@sciencespo.fr](mailto:info.cdsp@sciencespo.fr)