
Considérations préalables à la création d'un centre d'archives de données qualitatives

par Andreas WITZEL & Reiner MAUER

Introduction: Pourquoi instaurer un centre d'archives pour les données qualitatives?

En Allemagne comme dans bon nombre d'autres pays, il existe une profusion de cas d'analyse secondaire ou de réanalyse des données quantitatives, qui incluent par exemple une analyse interculturelle ou longitudinale de vastes jeux de données comparatives (e.g., *Eurobarometer*, *European and World Values Surveys*, *European Social Survey*, ALLBUS¹). En particulier, les *Archives Centrales des Sciences Sociales Empiriques* (ZA²) à Cologne, qui constituent le centre local des *Services Allemands d'Infrastructure pour les Sciences Sociales* (GESIS³), soutiennent et encouragent depuis plus de 40 ans cette tradition des sciences sociales et des études pluridisciplinaires en favorisant l'utilisation d'un vaste éventail de données de sciences sociales aux fins de l'analyse secondaire.

Il n'en va pas de même dans le domaine des études qualitatives en Allemagne. Il n'existe aucune culture répandue d'une utilisation secondaire du volume exceptionnel et fourni de données issues des études qualitatives disponibles — en particulier d'entretiens qualitatifs (re)transcrits. Et il est impossible de trouver une institution offrant un service de données orienté vers l'utilisateur pour les documents qualitatifs à l'échelle nationale. De plus, aucune étude scientifique systématique n'est menée sur les possibilités et les limites de la réutilisation et de la révision des informations qualitatives existantes. Ces lacunes sont étonnantes, au vu de l'importance croissante que prennent les méthodes qualitatives depuis les années 1970 et qui se traduit par l'augmentation constante du volume des données qualitatives, la croissance des applications de l'informatique aux sciences, ainsi que par les progrès des logiciels d'analyse de données qualitatives. Ceux-ci rendent possibles, grâce à l'existence de services appropriés, d'éventuelles utilisations secondaires dans le domaine de l'enseignement et de la recherche. Cette situation est également surprenante compte tenu des nombreux avantages que présente la réutilisation des données qualitatives⁴.

Actuellement, en Allemagne, il existe seulement deux centres d'archives décentralisées de matériaux qualitatifs spécialisés sur des thèmes spécifiques (psychothérapie, transitions dans la vie, culture politique, mouvements sociaux, documentation et manifestes politiques, mémoires de guerre et d'après-guerre, histoire naturelle et de protection de l'environnement, etc.). Il existe également des archives de différents types de données qualitatives — données d'histoire orale, lettres, photographies, journaux personnels, biographies, essais, correspondances et enregistrements audio et vidéo. Même s'il est généralement possible de consulter des catalogues électroniques dans les archives allemandes, jusqu'à présent la plupart des données ne sont pas accessibles sous format numérique ou lisible mécaniquement.

Au vu de cette situation, l'ALLF (*Life Course Archive*)⁵ de la *Graduate School of Social Sciences* (GSSS) de l'Université de Brême, se donne pour mission d'améliorer les conditions archivistiques et méthodologiques des données dans le cadre d'un projet national de développement des archives (Opitz & Witzel, 2005). Le centre d'archives ALLF est chargé du recueil d'entretiens qualitatifs, qui constituent la forme de données la plus utilisée pour les analyses secondaires selon l'expérience britannique. Ce centre a été créé selon la recommandation de la *German Research Foundation* (DFG)⁶ afin d'archiver les importants volumes de données du *Collaborative Research Centre 186 Status*

1. *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften*.

2. *Zentralarchiv für Empirische Sozialforschung / Central Archive for Empirical Social Research*.

Cf. <http://www.gesis.org/ZA/>

3. *Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen*.

4. Ce point est souligné par exemple par Corti & Thompson (2004). Contrairement à l'Allemagne, ces données sont déjà utilisées au Royaume-Uni grâce à *ESDS Qualidata* créé en 1994 (Corti, Witzel, & Bishop, 2005).

5. *Archiv für Lebenslauforschung*. Cf. <http://www.lebenslaufarchiv.uni-bremen.de>

6. *Deutsche Forschungsgemeinschaft*. Cf. <http://www.dfg.de>

Passages and Risks in the Life Course (1989-2001)⁷ de l'Université de Brême, et afin de les mettre à disposition des futurs utilisateurs. De nombreux entretiens recueillis dans le cadre de projets conduits par le centre (en partie des corpus de données longitudinales uniques traitant des transitions et des changements de statut au long de la vie des gens) ont été anonymisés, puis documentés et archivés sous format numérique. Il est remarquable que la plupart des projets réalisés aient utilisé des méthodes de recherche mixtes, ce qui rend possible l'accès à des entretiens qualitatifs et aux données quantitatives correspondantes.

La coopération avec les archives expérimentées est indispensable pour mener à bien le développement de ce type de centre d'archives : au niveau national, une coopération est ainsi menée avec le ZA qui jouit d'une longue expérience de l'archivage des données de sciences sociales et qui a encouragé l'analyse secondaire dans la recherche et l'enseignement dans le domaine des données quantitatives. A l'échelle internationale, il était évident de travailler en concertation avec *ESDS Qualidata*⁸, qui propose des procédures et des méthodes d'archivage ainsi qu'une analyse secondaire de recherche qualitative depuis sa création en 1994⁹.

La situation en Allemagne : résultats et conditions requises

Les considérations organisationnelles exposées ici, concernant le développement d'un centre d'archives national des données d'entretiens qualitatifs, sont issues d'une étude de faisabilité intitulée « archivage et utilisation secondaire des données d'entretiens qualitatifs »¹⁰. Cette étude principalement empirique — soutenue par la *German Research Foundation* et menée conjointement par les deux institutions ALLF et ZA — constitue le fondement scientifique de création d'un centre d'archives.

L'urgence de la création d'un centre d'archives

Il est urgent de créer un centre d'archives dans la mesure où un grand nombre de données risque de se perdre. Les résultats empiriques de l'étude de faisabilité¹¹ concernant la localisation des données de près de 1100 projets utilisant l'entretien qualitatif en Allemagne ne sont pas si catastrophiques à première vue : seules les données de 13% des projets identifiés ne sont plus disponibles. Mais si l'on considère que 60% de ces projets se sont achevés en 2003/2004 ou demeuraient en cours et que la période d'étude ne portait que sur les 10 dernières années, le volume des données irrécupérables est déjà considérable.

Face à cette situation, il semble surprenant que les données de près d'un quart des projets soient déjà archivées. Toutefois, l'enquête auprès de ces archives et les entretiens d'experts menés dans le cadre de l'étude de faisabilité ont révélé que lors de l'utilisation des termes « remis aux archives » ou « archivés », bon nombre de répondants ne se représentaient rien d'autre qu'un simple stockage des données dans une salle, qui ne répond même pas aux normes élémentaires de l'archivage professionnel. Cela signifie le plus souvent que les bandes audio originales ou les textes d'entretiens partiellement retranscrits existent, que ce matériau n'est en général pas anonymisé, qu'il est conservé sans aucune mesure de protection physique, que le public ne peut y accéder ou qu'il n'existe ni documentation, ni catalogue.

Au-delà du risque de perdre d'importantes données empiriques, l'absence d'archives entrave le développement d'une culture de l'analyse secondaire dans la recherche sociale qualitative. Il n'est pas étonnant que la réutilisation spécifique des données externes (non produites par le chercheur) n'ait lieu

7. Les *Collaborative Research Centres* sont des centres de recherche universitaire à long terme dans lesquels les scientifiques et les universitaires effectuent conjointement des recherches interdisciplinaires ambitieuses.

8. Voir les chapitres de Louise Corti et de Libby Bishop dans cet ouvrage.

9. Les deux institutions soutiennent le développement d'un centre d'archives et nous adressons particulièrement nos remerciements à Louise Corti et Libby Bishop (*ESDS Qualidata*) ainsi qu'à Ekkehard Mochmann (ZA) pour nos discussions fructueuses et constructives.

10. Projet « *Archivierung und Sekundärnutzung qualitativer Daten – eine Machbarkeitsstudie* », 2003-2005. Equipe : Prof. Karl F. Schumann, Dr. Andreas Witzel, Irena Medjedović, Diane Opitz, Britta Stiefel (Bremen), Prof. Wolfgang Jagodzinski, Dr. h.c. Ekkehard Mochmann, Reiner Mauer (Cologne). Pour en savoir plus, voir le chapitre d'Irena Medjedović dans cet ouvrage.

11. Cf. Opitz & Mauer (2005) pour lesdits résultats de la partie quantitative de l'étude. Les résultats qualitatifs sont extraits du rapport de la DFG non publié pour le moment.

que très rarement, même si cela n'est pas totalement inhabituel. Par exemple, les résultats de l'étude de faisabilité révèlent que plus d'un tiers des répondants ont déjà réutilisé des données qualitatives mais que la plupart utilisaient leurs propres données (56%). A l'inverse, seuls 20% ont réutilisé les données externes. Dans la mesure où la plupart des répondants ont indiqué n'avoir aucune raison ni aucun motif particulier d'effectuer une analyse secondaire, on peut en déduire que la réutilisation des données qualitatives est fort peu exercée et que les avantages de l'analyse secondaire sont mal compris¹².

Malgré une incertitude bien réelle, un manque de connaissance et un scepticisme certain devant les opportunités et les avantages de la réutilisation des données qualitatives, 80% des chefs de projet interrogés étaient favorables à l'idée de créer une infrastructure pour l'archivage de leurs données et pouvant servir de source de données qualitatives en Allemagne. Une partie de l'étude de faisabilité a été consacrée à l'inventaire du stock de données qualitatives en Allemagne. Les analyses ont mis à jour un grand nombre de projets comportant des entretiens qualitatifs, pour lesquels 60% des chefs de projet avaient accepté en principe de transmettre les données à d'autres aux fins de l'analyse secondaire. De plus, 65% des répondants prévoyaient de réaliser des analyses secondaires dans l'avenir.

Performances attendues d'un centre d'archives

La demande de mise en place d'une infrastructure dans le domaine de données qualitatives est liée aux attentes en matière de performances d'un tel établissement. Les réponses aux questions ouvertes de notre étude indiquent que ces attentes émanent essentiellement des utilisateurs potentiels des données. Il s'agit surtout principalement d'utilisateurs qui attendent une bonne documentation du processus de recherche et un catalogue de données transparent (leurs mots-clé sont : *systématisation, catalogue, facilité d'accès, accès en ligne*). L'étude de faisabilité a établi que la documentation devait s'étendre au-delà des descriptions de projet que l'on peut lire ordinairement dans les publications et qu'elle devait inclure des informations précises sur la situation d'entretien mais aussi certaines parties de l'analyse de l'étude primaire.

Une organisation centrale ou locale ?

Il s'agit également de savoir si un futur centre d'archives doit être organisé sur un mode centralisé ou décentralisé, à l'instar du « modèle en étoile » (*hub and spokes model*) appliqué par *ESDS Qualidata*. Corti (2005) pense que le stockage décentralisé des données dépend d'un réseau de centres d'archives de données déjà existants et bien organisés. Toutefois, ce type de réseau d'archives fonctionnelles tel qu'il existe au Royaume-Uni n'existe pas en Allemagne, en particulier concernant les entretiens qualitatifs retranscrits et numérisés. Au mieux, on peut envisager une fonction de centre d'échange pour des centres d'archives spécialisés dans certains types de matériaux qualitatifs ou dans certains sujets d'étude. Néanmoins, cela est dépendant d'un mode de financement, non seulement pour créer un catalogue commun, mais aussi pour la numérisation des données, lesquelles n'existent actuellement que sur papier.

Une organisation centralisée permet de garantir des normes en termes de qualité, de conservation à long terme des données et de développement d'un accès contrôlé et durable aux données. Les experts interrogés s'expriment également dans ce sens. Les données des projets de recherche stockées au sein des instituts où elles ont été recueillies au départ ne répondent souvent pas aux normes de qualité que les experts eux-mêmes estiment essentielles pour une utilisation secondaire. Compte tenu des lacunes en termes d'exhaustivité de transcription, de numérisation, d'anonymisation et de documentation des données, du manque d'espace et de ressources financières pour la préparation et le catalogage des données, les éventuels déposants de données justifient la nécessité d'un archivage centralisé qui, grâce à la mise en commun des ressources, pourrait garantir les normes de qualité auxquelles ils prétendent. De plus, du point de vue des utilisateurs potentiels, la centralisation du catalogue et des données archivées permet un accès facile, complet et systématique aux données.

12. Irena Medjedović précise ce point dans son chapitre en présentant les résultats de l'enquête de faisabilité auprès des chercheurs.

Cadre juridique de l'archivage des entretiens qualitatifs en Allemagne

Les chefs de projet interrogés ont cité le plus souvent, comme conditions préalables à la transmission de leurs données à des tiers, la *garantie de l'anonymat* et la *légitimité de la diffusion des données en matière de confidentialité et de consentement éclairé*. Paradoxalement, cet aspect est également le plus cité par ceux qui refusent la diffusion de leurs données. L'étude de faisabilité montre que les chercheurs analysent différemment le problème de la confidentialité selon le type de données et en fonction des risques d'influence ou de contrôle qui pèsent sur l'utilisation secondaire de leurs données.

Assurer la faisabilité d'un centre d'archives requiert un examen approfondi du cadre juridique, puisque l'optimisation de la qualité des données et l'accès à celles-ci à des fins d'analyse secondaire sont intimement liés aux exigences de la protection des données. Tout d'abord, les données qualitatives d'un futur centre d'archives ne peuvent être utilisées qu'à des fins de recherche¹³ et uniquement dans un objectif devant être confirmé par écrit¹⁴. De plus, en Allemagne, la liberté de la recherche scientifique garantie par la constitution¹⁵ s'affronte au droit individuel de l'auto-détermination informationnelle¹⁶. Ainsi, en 1983, par une décision fondamentale concernant une loi de réglementation des recensements, le Tribunal Constitutionnel Fédéral a explicitement indiqué que l'individu devait être protégé contre une conservation, un stockage, une utilisation et une diffusion illimités de données personnelles ; d'autant plus que le traitement moderne des données permet, comme jamais auparavant, de générer des situations ingérables aboutissant au contrôle et à des prises d'influence¹⁷. D'après cette disposition, les individus jouissent du droit fondamental de décider librement de la divulgation et de l'utilisation de leurs données personnelles¹⁸.

Puisque les lois de protection des données régissent les informations afférentes aux individus¹⁹, les données scientifiques peuvent être transmises à d'autres projets de recherche, ou bien à condition que la personne l'accepte (dans le cadre d'accords de « consentement éclairé ») ou bien à condition que les données soient anonymisées de telle façon que les données individuelles sur des situations personnelles ou objectives ne puissent être attribuées à une personne reconnaissable, du moins sans entraîner trop de temps, de frais et de main d'œuvre²⁰. Il est ainsi possible (par exemple, pour le ZA à Cologne) de diffuser des corpus quantitatifs sous réserve qu'ils aient été « factuellement anonymisés » (Müller, Blien, Knoche, Wirth, Beckmann, Bender, Helmcke, & Müller, 1991).

Toutefois, la question de savoir s'il existe les mêmes possibilités pour le stockage des données qualitatives reste floue, comme celle de savoir quelles formes d'anonymisation en seraient les conditions préalables. La sensibilité particulière des données qualitatives est le fruit d'une démarche scientifique qui, comparée aux méthodes standardisées, recourt à une enquête thématique plus détaillée à l'origine de réponses différenciées de la part des répondants. Ce faisant, les positions subjectives des personnes interrogées (acteurs en interaction avec d'autres acteurs et qui doivent être aussi protégés) demeurent le centre d'intérêt du chercheur. Dans les entretiens qualitatifs, on obtient des réponses différenciées à la fois en évitant les réponses prédéfinies par l'interviewer et en rendant possibles les réponses libres des interviewés ainsi que leurs clarifications. En fonction des diverses possibilités offertes par les diverses techniques d'entretien, les répondants reconstituent des actions, des expériences, des décisions, des connaissances et des points de vue, font référence à des conditions positives ou restrictives d'action et décrivent eux-mêmes les connexions qui les lient. Ainsi, bon nombre de données personnelles disponibles sous différents types de format, concernent le répondant lui-même ou des tiers.

Afin de procéder à l'anonymisation factuelle, il suffit souvent de supprimer des détails tels que le nom, le lieu, l'année ou — comme le propose le concept d'anonymisation de l'ALLF (Kluge & Opitz, 1999) — de les remplacer par des informations pertinentes selon les critères des sciences sociales. Cela se produit généralement au cours de la transcription des entretiens, conformément à la

13. § 40 para. 1. et 2, *Bundesdatenschutzgesetz* (BDS).

14. Par exemple, § 21, para. 1, *Bremisches Datenschutzgesetz*.

15. Art. 5 para. 3, *Grundgesetz* (GG).

16. Art. 2 para. 1, GG.

17. *Bundesverfassungsgericht* (BVerfGE), 65, S. 42 f; Simitis (1984).

18. BVerfGE, 65, S. 43.

19. §1 para. 1 et 2, BDSG.

20. § 3 para. 7, BDSG.

réglementation légale exigeant que l'anonymisation doit s'effectuer « le plus tôt possible en respect des fins de la recherche » (§ 40, para. 3, BDGS). De plus, il est possible d'exclure de l'archivage les échantillons trop petits et donc facilement identifiables. En ce qui concerne les entretiens biographiques contenant des *curriculum vitae* et de nombreuses descriptions d'événements de vie qui, ensemble, forment un contexte individuel complexe et distinct, la possibilité d'identifier les répondants ne peut souvent pas être évitée, même par la suppression supplémentaire de passages de texte entiers.

Toutefois, puisqu'il est difficile de procéder à d'ambitieuses analyses de ces entretiens sans connaître les détails biographiques, un conflit oppose les exigences de protection de la confidentialité et les conditions préalables à une recherche empirique valide : les chercheurs interrogés dans le cadre de notre étude de faisabilité craignent, d'une part, que les utilisateurs secondaires, en particulier lorsqu'ils utilisent des données biographiques, ne puissent traiter les données avec trop peu de confidentialité et de prudence. D'autre part, ces mêmes chercheurs posent comme condition préalable à l'utilisation secondaire non seulement la disponibilité des données initiales, des informations contextuelles sur l'entretien ou encore des mémos et descriptions (par exemple les résultats d'études de cas dans lesquels le contexte général est une condition préalable élémentaire à l'interprétation des textes), mais également la disponibilité des données démographiques, afin de mettre en relation les orientations et actions individuelles avec des caractéristiques socio-démographiques.

Concernant une utilisation secondaire des données qualitatives en particulier, aucun cadre juridique clair n'existe dans la législation allemande ni dans la directive européenne sur le respect de la vie privée²¹. L'urgence de trouver une solution ne naît pas seulement du besoin de créer un centre d'archives des données d'entretiens, mais également des recommandations de la DGF concernant les règles déontologiques de la pratique scientifique. La DFG rappelle que les données doivent être transmises à des tiers si ces derniers peuvent prouver un intérêt professionnel fondé pour leur réanalyse. De ce fait, les données devraient être conservées pendant dix ans à compter de la réalisation des analyses au sein des instituts où elles ont été produites — ou en tout autre lieu indépendant (considéré comme centre d'archives).

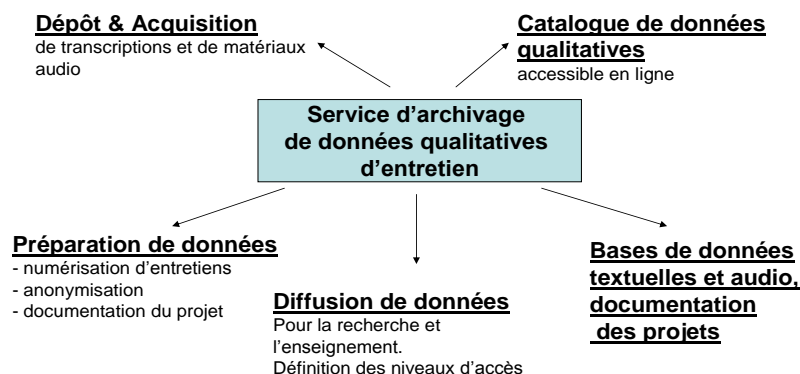
Pour résumer, tous ces propos sur la protection des données et la confidentialité relative aux données qualitatives révèlent la complexité du sujet. Un avis juridique est donc indispensable pour répondre à ces questions.

Les activités centrales d'une archive

Grâce à l'*Archive for Qualitative Interview Data* (AQUID) et suite à l'étude de faisabilité, nous nous efforçons de constituer un service national et central pour l'archivage et la diffusion de transcriptions numérisées d'entretiens. Le concept de ce centre d'archives inclut les travaux d'acquisition, de préparation, de documentation et de protection des données, la réalisation d'un catalogue centralisé de données disponible en ligne, ainsi que l'organisation d'un service de données orienté vers l'utilisateur. Ce qui est encore inexistant pour la communauté de chercheurs qualitatifs en sciences sociales figurera parmi les missions essentielles de ce centre d'archives : il s'agira ainsi de fournir des transcriptions numériques d'entretiens accompagnées d'une documentation. Ce service ne doit pas être seulement destiné aux utilisateurs potentiels qui font de la recherche universitaire, mais aussi à ceux intervenant dans le domaine académique plus large (Klingemann & Mochmann, 1975) et dans l'enseignement (cf. Graphique 1).

21 Directive européenne sur la Confidentialité et la Sécurité et les Communications Electroniques (12 juillet 2002), OJ EC L 201, 31.07.2002, p. 37.

Graphique 1 : L'organisation d'AQUID



Acquisition des fonds de données

Pour la mise au point des critères d'acquisition des fonds de données, l'AQUID s'appuie sur les expériences déjà réalisées en Grande-Bretagne. Nous avons défini des critères d'ordre méthodologique et thématique ; ils sont développés plus loin. En comparant les critères d'archivage et de réutilisation des données qualitatives du *Qualitative Data Archival Resource Centre (ESDS Qualidata)*, du *Medical Research Council* (Royaume-Uni) et du *Murray Research Center* (Université de Harvard, États-Unis), Corti et Backhouse (2005) concluent à l'importance des points suivants : caractère longitudinal des études avec possibilité de suivi dans le temps ; données remplissant les exigences de la recherche qualitative, étant largement diversifiées et de grande quantité ; potentiel d'analyse non entièrement exploité lors de l'étude primaire. Les auteurs insistent également sur le caractère essentiel de la qualité de l'entretien pour sa réutilisation à des fins d'analyse secondaire : « les matériaux issues d'entretiens approfondis constituent la meilleure source de données pour les analyses secondaires, sous réserve que les entretiens aient été adroitement menés par un enquêteur expérimenté et qu'ils couvrent un large éventail de thèmes »²² (Corti & Backhouse, 2005, paragraphe 22).

La création du centre AQUID pour l'archivage d'entretiens s'avère pertinente d'autant plus que la méthode d'entretien semble avoir actuellement un fort taux de diffusion en Allemagne. Ainsi, grâce à l'étude de faisabilité, on a pu faire l'inventaire des projets de recherche en sciences sociales utilisant la méthode d'entretiens. Plus de 1750 projets recourant aux entretiens qualitatifs ont pu être identifiés sur la période 1984-2003.²³

Suite à l'étude de faisabilité, l'AQUID utilisera au départ les critères méthodologiques suivants pour l'acquisition des données :

- Concentration sur des entretiens non structurés à semi-structurés, de préférence intégrant des parties narratives.
- Préférence pour des échantillons vastes. Néanmoins, dans certains cas il faudra vérifier si les projets à échantillons faibles peuvent également présenter une importance scientifique (par exemple, avec des acteurs de l'histoire contemporaine).

22. Voici la citation originale : "In-depth interview materials are the best source of data for secondary analysis, providing the interviews have been skillfully conducted by an experienced interviewer and cover a wide range of topics" (Corti & Backhouse, 2005, paragraphe 22).

23. La constitution de l'inventaire a été soutenue par le *Social Science Information Centre* (IZ, <http://gesis.org/IZ>), qui — comme les *Archives Centrales des Sciences Sociales Empiriques* (ZA) de Cologne — fait partie des *Services Allemands d'Infrastructure pour les Sciences Sociales* (GESIS). Source d'information : base de données FORIS (système d'information de recherche en sciences sociales contenant des descriptions de projets de recherche prévus, en cours et achevés en Allemagne, en Autriche et en Suisse). La description différente du flux de travail visée au centre *UK Data Archive* (2003) est exemplaire et utile pour l'AQUID.

- Concentration sur les données numérisées dans la mesure du possible, à peine la moitié des données potentiellement archivables existant sous format numérique. « Dans la moitié des cas, les données sont déjà numériques, 30% existent sur papier et 20% sous format audio. L'étude a indiqué que la disponibilité sous format électronique est désignée par les utilisateurs potentiels comme condition préalable importante à l'analyse secondaire. De même, la conversion sous format mécaniquement lisible a été posée comme service nécessaire d'une archive. »²⁴ (Opitz & Mauer, 2005, paragraphe 10).

- Une attention particulière doit être prêtée aux données longitudinales, particulièrement intéressantes pour la réutilisation (Corti & Wright, 2002). Les expériences britanniques montrant le potentiel à long terme des données longitudinales et leur pertinence pour l'analyse secondaire ont été déjà confirmées par les interrogations fréquentes du catalogue de l'ALLF. Celui-ci recense en majeure partie des entretiens issus d'études longitudinales sur les trajectoires de vie. Les fonds de données de l'ALLF seront donc complétés par d'autres données d'entretiens issus d'études longitudinales.

Parallèlement aux critères méthodologiques, les critères thématiques suivants seront appliqués à l'acquisition des données :

- Etudes sur les trajectoires de vie : d'une part, les études longitudinales à orientation interdisciplinaire doivent être privilégiées pour compléter ainsi sur le plan thématique les fonds de l'ALLF. D'autre part, l'acquisition d'études transversales complémentaires débouche sur une diversité thématique. L'objectif est d'archiver des études diverses présentant des variations en termes d'âge, de région, de caractéristiques socio-démographiques, et de date de recueil. Elles devraient ainsi se compléter et permettre la réalisation d'analyses secondaires avec de nouvelles questions de recherche ciblées sur des comparaisons (entre autres à caractère historique).

- Acquisition d'études présentant une importance sociétale générale et/ou une importance thématique avérée pour la communauté scientifique. L'acquisition des données doit être conduite de manière progressive dans des domaines scientifiquement et socialement importants.

Pour préserver l'utilité des données et leur utilisation large dans le cadre d'analyses secondaires, il est important d'ancrer cette fonction d'acquisition de données au sein de la communauté scientifique. Une acquisition active et thématique des données doit donc inclure les études dont la pertinence scientifique et sociale a été évaluée par un comité consultatif composé d'experts spécialisés dans différentes disciplines et différents secteurs.

Préparation et conservation à long terme des données

En premier lieu, l'AQUID traitera des données textuelles — principalement des transcriptions d'entretiens qualitatifs — qui devront en majeure partie être disponibles sous format numérique. La préparation des données²⁵ recouvre toutes les activités nécessaires à l'archivage et à la gestion des données susceptibles de permettre aux utilisateurs secondaires de localiser et d'utiliser les données archivées. Au-delà du contrôle lié au caractère exhaustif et à la qualité des données, il est à mener un travail technique (par exemple, la conversion en formats appropriés, la numérisation des données, etc.) particulièrement quant à la production de méta-données²⁶. Ce dernier travail consiste essentiellement à compiler des documentations liées aux projets de recherche. Selon nos répondants, elles doivent inclure une documentation complète sur les données ainsi que sur le processus de recherche.

24. Voici la citation originale : "In half of the cases the data material is already digitalized, further 30% are available in paper format and the remaining 20% in audio format. The survey showed that availability in an electronic format is named as an important prerequisite for secondary analyses by potential users. Likewise the conversion into a machine-readable format was formulated as demanded service of an archive." (Opitz & Mauer, 2005, paragraphe 10).

25. L'AQUID prend comme modèle la description détaillée des tâches de travail par le centre *UK Data Archive* (2002).

26. Ces dernières années, la norme internationale *Data Documentation Initiative* (DDI, cf. <http://www.ddialliance.org>) a été largement acceptée dans le domaine des données quantitatives. Le ZA joue un rôle actif dans la formulation et le développement de ces normes.

Plus de 40% des chefs de projet ont indiqué que les entretiens réalisés dans le cadre de leurs projets n'ont pas été retranscrits ou ne l'ont été que trop peu, et que les données n'ont pas été anonymisées suffisamment pour être transmises à des tiers. Seuls 45% des données existent sous format numérique. Dans un quart des cas, aucune documentation pertinente n'est disponible. Les entretiens que nous avons menés montrent que l'existence du contexte de recherche est une condition nécessaire pour rendre les données utilisables pour des analyses secondaires. Il s'ensuit que l'AQUID devra déployer un effort considérable pour la préparation et la documentation des données, notamment pendant la phase de l'acquisition des fonds.

Sur la base de nos propres recherches (Medjedović & Witzel, 2005), la préparation des données en vue d'une analyse secondaire doit également inclure les grilles de codage et de catégories issues des analyses assistées par ordinateur de l'étude primaire. L'existence de grilles de catégorisation, comportant des catégories larges et heuristiques, ne devrait pas toutefois entraver, lors l'analyse secondaire, un codage plus ouvert avec la définition de codes *in-vivo* — conformément à la tradition de la *Grounded Theory* (Strauss & Corbin, 1996). Par ailleurs, afin de pouvoir effectuer également l'utilisation secondaire de ces codages avec d'autres programmes d'analyse textuelle, les interfaces avec tous les logiciels de type CAQDAS doivent être prévues. Sur ce point, selon *ESDS Qualidata* (Corti, 2003b) il faudrait encourager le développement d'un format de données non dépendant d'un logiciel. Dans ce processus, les formats XML ouverts, fiables et flexibles (Muhr, 2000) jouent un rôle central.

Comme pour la conservation numérique à long terme des données, au-delà de la simple protection contre les pertes physiques et de la garantie de l'intégrité des données, il est particulièrement important de réduire les risques liés aux changements technologiques. Cela induit, entre autres : la création d'une structure cohérente de répertoire utilisant des conventions de désignation standardisées, le choix de formats appropriés d'acquisition, de stockage et de distribution, le stockage dans des conditions optimales (par exemple, à une température et une humidité optimales) ; l'application de systèmes de sécurité qui réduisent les risques de corruption et de perte des données dus aux dangers extérieurs (incendie, vol) ; un changement régulier de support de stockage ; une migration ponctuelle des données vers des formats ou des supports de stockage courants.

Hormis la dotation d'un matériel et de logiciels appropriés, de systèmes de sauvegarde redondants (également séparés dans l'espace), de scanners, etc., il est important de respecter les normes scientifiques techniques et d'information de la gestion des données. La traçabilité et l'exploitabilité à long terme des données archivées ne peuvent être garanties que par ce moyen.

Permettre l'accès aux fonds via un catalogue de données

Les participants à l'étude de faisabilité insistent sur la mise en place d'un système d'information transparent sur les corpus de données potentiellement utiles (le plus probablement par Internet). Il doit être facile à utiliser et doit, selon différents critères de recherche, permettre une bonne vue d'ensemble des études importantes, en informant sur les contenus, les chercheurs primaires, les méthodes appliquées, les matériaux disponibles, l'état des données, la réglementation de leur usage, etc. On suppose alors que la construction d'une base de données informative et d'un moteur de recherche convivial — de préférence en ligne — incitera davantage de chercheurs à vérifier s'il existe déjà des données utiles à leurs études avant d'entreprendre leurs propres études. Le format XML répondant à la norme DDI, dont nous avons parlé précédemment, peut être utilisé pour les descriptions d'études.

Diffusion des données

Pour près d'un quart des répondants, le contrôle sur une nouvelle utilisation des données constitue une condition importante de leur volonté de transmettre leurs données à des archives. Différentes craintes « d'abus » par des tiers sont ainsi associées au dépôt de données. Ce contrôle comprend entre autres des informations sur l'utilisateur des données, l'objectif de son utilisation, un agrément d'utilisation donné par le déposant des données pour chaque utilisation de ses données, ou encore pour la publication des résultats de l'analyse secondaire. Pour répondre à cette exigence, la diffusion des données s'effectue sur la base de critères et de restrictions réunis dans un contrat passé avec les déposants de données et peut inclure des possibilités de contact avec les chercheurs de l'étude primaire. Ces critères de diffusion des données seront élaborés sur le fondement de l'expérience du ZA

(relative à la protection des données)²⁷ et d'*ESDS Qualidata*, ainsi que des résultats des entretiens qualitatifs de l'étude de faisabilité. Dans la mesure où l'AQUID archive principalement des données numérisées, l'apport des données peut s'effectuer sans trop d'effort au moyen d'un support approprié (Cd-ROM, DVD) ou encore par téléchargement sur FTP. Dans l'avenir, la diffusion des données pourrait également être organisée au moyen d'une inscription sur Internet et d'un système de téléchargement.

Extension des fonctions du service

L'expérience d'*ESDS Qualidata* et la situation en Allemagne, révélée par les résultats de notre étude de faisabilité, suggèrent la création d'un site d'approvisionnement plus vaste de données qualitatives dans le futur, à savoir le développement de l'AQUID au-delà des fonctions centrales des archives, vers un *Centre de Service de Données Qualitatives (QualiService)* (cf. Graphique 2).

Graphique 2 : *Centre de Service de Données Qualitatives (QualiService)*

<i>QualiService</i>				
<p>Archive pour les données d'entretiens</p> <p style="text-align: center;"><u>Activités centrales</u></p> <ul style="list-style-type: none"> • Dépôt et acquisition • Préparation des données • Diffusion des données • Catalogue transparent des données • Bases de données textuelles et audio et documentation des projets <p style="text-align: center;"><u>Activités supplémentaires</u></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%; vertical-align: top;"> <p>Recherche & Développement</p> <ul style="list-style-type: none"> • Méthodes d'analyse secondaire • Exemples d'analyse secondaire • Codage des données audio • Standards pour les métadonnées • Amélioration des méthodes d'enseignement </td> <td style="width: 50%; vertical-align: top;"> <p>Conseil & Formation</p> <ul style="list-style-type: none"> • Protection des données • Analyse secondaire • Méthodes qualitatives • Enseignement à distance et modules de formation </td> </tr> </table>		<p>Recherche & Développement</p> <ul style="list-style-type: none"> • Méthodes d'analyse secondaire • Exemples d'analyse secondaire • Codage des données audio • Standards pour les métadonnées • Amélioration des méthodes d'enseignement 	<p>Conseil & Formation</p> <ul style="list-style-type: none"> • Protection des données • Analyse secondaire • Méthodes qualitatives • Enseignement à distance et modules de formation 	<p style="text-align: center;">Méiateur pour les matériaux qualitatifs d'autres centres d'archives germanophones</p> <ul style="list-style-type: none"> • Références aux fonds des autres centres d'archives de données qualitatives
<p>Recherche & Développement</p> <ul style="list-style-type: none"> • Méthodes d'analyse secondaire • Exemples d'analyse secondaire • Codage des données audio • Standards pour les métadonnées • Amélioration des méthodes d'enseignement 	<p>Conseil & Formation</p> <ul style="list-style-type: none"> • Protection des données • Analyse secondaire • Méthodes qualitatives • Enseignement à distance et modules de formation 			

Fonction de médiation pour les données qualitatives

L'élargissement des fonctions centrales du centre d'archives pourrait se faire de deux manières : d'une part, en alimentant les fonds avec de nouveaux types des données qualitatives (par exemple, groupes de discussion, textes, documents) ; d'autre part, en améliorant l'accès aux données provenant d'autres archives germanophones, au moyen d'une fonction de médiation centralisée entre ces archives et les chercheurs en quête de données pertinentes.

Par ailleurs, un système d'information central sur les données issues de méthodes mixtes, associant méthodes qualitatives et quantitatives, sera mis à la disposition des utilisateurs potentiels des archives. Cette fonction pluridisciplinaire et pluriméthodologique de *QualiService* correspond à celle d'un « courtier », telle qu'elle est exercée par *ESDS Qualidata* en Grande-Bretagne.

La coopération avec d'autres archives ou d'autres initiatives d'archivage de données germanophones sert à améliorer l'accès aux fonds de données qualitatives jusqu'alors très éparpillés. Outre l'Allemagne, on pourrait également envisager la coopération avec les initiatives autrichiennes et suisses afin que *QualiService* puisse contribuer à la médiation des données qualitatives en langue allemande.

27. Cf. http://www.gesis.org/en/data_service/order/usage_regulations.htm

Conseil, formation et relations publiques

Les répondants souhaitent obtenir des conseils et des services (par exemple dans le domaine de la planification et de la mise en place de projets). Ils souhaitent également des ateliers sur le recueil ou l'analyse des données. Une autre fonction importante de *QualiService* devrait donc tenir au conseil à l'utilisateur (personnes, projets, institutions) sur des questions relatives à l'archivage, à la protection de la confidentialité et de la sécurité des données, au choix des données qualitatives et à la réanalyse ou l'analyse secondaire des données qualitatives. Les données numérisées peuvent être aussi proposées en association avec des programmes d'analyse textuelle pour l'enseignement des méthodes (voir par exemple, Gläser & Laudel, 2001). L'innovation pourrait venir de la dispense d'une formation *via* plusieurs combinaisons d'éléments en ligne et hors ligne pour l'amélioration de l'enseignement.

Pour palier le manque de notoriété des méthodes d'analyse secondaire (Corti, Witzel, & Bishop, 2005) et les expériences fort peu satisfaisantes de cette pratique, il est nécessaire de sensibiliser la communauté scientifique au potentiel de réutilisation des données en recourant aux relations publiques. Une plateforme en ligne d'échanges d'informations par des listes de diffusion ainsi que par des revues et lettres d'information, sur le modèle du portail Internet www.qualitative-research.net qui collabore déjà avec l'ALLF, constitue un moyen efficace.

Recherche et développement

Corti (2005) souligne que les activités de recherche en interne dans les centres d'archivage rendent plus efficaces tant l'acquisition des fonds que leur utilisation scientifique²⁸. Ces activités de recherche ont ainsi porté leurs fruits en Grande-Bretagne, en Scandinavie et aux Etats-Unis. Concernant un *QualiService* allemand, divers sujets scientifiques pourraient être étudiés : ils devraient permettre la mise au point des normes méthodologiques et normes de qualité pour l'archivage des données qualitatives.

De plus, un accès en ligne innovant à destination des utilisateurs de données doit être mis en place. Pour encourager une « culture de l'analyse secondaire » dans le domaine de la recherche qualitative, il est indispensable d'en étudier de manière systématique les avantages et les inconvénients. On citera comme exemples l'étude du potentiel analytique secondaire de différentes techniques d'entretien, la considération du contexte de recueil des données, et les analyses secondaires réalisées à titre d'exemple (cf. Medjedović & Witzel, 2005).

Enfin, des activités de recherche devraient porter sur les nouveaux développements de logiciels de type CAQDAS. Grâce au logiciel d'analyse ATLAS.ti, il est par exemple possible de préparer une table des matières de données audio numérisées, à savoir le système de codage (C-TOC) (Hauptmann, 2005) dont les éléments peuvent être utilisés sans devoir passer par l'étape intermédiaire des transcriptions comme base de l'analyse, en particulier pour l'extraction des données. Afin de garantir l'anonymisation des fichiers audio en vue de leur diffusion à des tiers, de nouvelles méthodes techniques (Pätzold, 2005) devraient être testées ; grâce à elles les sources audio devraient être suffisamment brouillées pour éviter ou compliquer leur reconnaissance.

Perspectives

Un scientifique en quête de données qualitatives à usage secondaire en Europe ou aux Etats-Unis peut trouver des centres d'archives de données tels que le *US Murray Research Center*, le centre *Finnish Social Science Data Archive* (FSD) ou encore le *UK Economic and Social Data Service* (ESDS)²⁹, qui proposent des données tant qualitatives que quantitatives. Pour son projet de création progressive d'un centre de service de données qualitatives (AQUID), l'ALLF coopère avec *ESDS Qualidata* en Grande-Bretagne et avec le ZA en Allemagne, spécialiste de l'archivage et de la diffusion de données principalement quantitatives depuis plus de 40 ans. La création de l'AQUID relève d'un projet non dogmatique sur le plan méthodologique qui vise à dépasser le schisme méthodologique particulièrement visible dans le partage de la *German Society for Sociology* (DGS) en deux sections méthodologiques, qualitative et quantitative. La coopération avec le ZA et *ESDS*

28. Sur cette problématique, voir également la contribution de Libby Bishop dans le présent ouvrage.

29. Cf. <http://www.radcliffe.edu/murray> (*US Murray Research Center*), <http://www.fsd.uta.fi/english/index.html> (FSD), <http://www.data-archive.ac.uk> (ESDS).

Qualidata consiste dans un premier temps à utiliser leurs connaissances méthodologiques, organisationnelles et techniques pour organiser un service national et central pour l'archivage et la diffusion de transcriptions numérisées d'entretiens qualitatifs.

L'élargissement des fonctions de service devra être réalisé uniquement dans un deuxième temps avec la transformation de l'AQUID en *QualiService*. Les étapes nécessaires à court terme de la construction de l'AQUID concernent les aspects suivants :

- Relations publiques

Afin de parvenir à réaliser l'archive envisagée et répondre au besoin de clarification sur les potentiels de la réutilisation des données qualitatives au sein de la communauté scientifique, une initiative intensive de relations publiques est nécessitée sous la forme de conférences, de publications et de forums en ligne. Les organismes de financement doivent être encouragés à veiller à ce que les projets de recherche intègrent parmi les tâches planifiées les questions de l'anonymisation, de la protection à long terme et de la documentation. Cela ne vient pas seulement étayer la transparence de la recherche et de la protection à long terme des données mais contribue également à rendre la recherche primaire plus efficace ; ainsi, une telle préparation des données rend plus simples voire possibles le processus analytique itératif, la réutilisation pour des nouvelles questions de recherche et la formation des nouveaux utilisateurs. L'analyse de la partie qualitative de l'étude de faisabilité se poursuivra pour approfondir les arguments autour de l'utilité de la réutilisation des données.

- Analyses secondaires à titre d'exemples

Il est également prévu de réaliser des analyses secondaires à titre d'exemples dans le domaine des relations familiales (par exemple, relations entre sexes / relations de couple) en vue de démontrer les potentiels de l'analyse secondaire ainsi que ses spécificités et ses problèmes méthodologiques.

- Sauver les études qualitatives importantes et anciennes

Afin d'apporter des données adéquates aux utilisateurs potentiels, il est important de sauver d'une perte certaine les anciennes études présentant une importance thématique et méthodologique sur les plans sociétal et scientifique. Une opération de sauvetage des données de « projets-phares » est indispensable, à l'exemple de la mission première d'*ESDS Qualidata* à l'occasion de sa création en 1994 en Grande-Bretagne. Les études sélectionnées devront être acquises, préparées et documentées. L'acquisition pourrait progressivement se porter sur des thèmes tels que les relations familiales et les relations entre sexes, les questions liées au métier et à l'emploi, la migration, la formation continue, le vieillissement, la jeunesse, l'environnement, les problèmes est-ouest, les médias, ou les modes de vie, et prolonger enfin le centre d'intérêt actuel des études de l'ALLF sur le cours de vie.

- Construire une infrastructure

Pour éviter les problèmes dans le développement de l'infrastructure du centre d'archives et, en conséquence, tout travail et dépenses supplémentaires, il est prévu de chercher un financement pour pouvoir doter l'infrastructure des technologies de pointe sur le plan informatique. Pour la définition des travaux nécessaires et du personnel on s'appuie sur les expériences du ZA et d'*ESDS Qualidata*. Le développement de l'infrastructure est une des conditions préalables essentielles à la réalisation de premiers essais pour l'acquisition et la préparation des données dans le cadre d'exemples d'études secondaires et du sauvetage des données.

Un autre fondement important du centre d'archive envisagé consiste à trouver des solutions juridiquement solides aux questions de la protection des données qui se posent tout au long du processus d'archivage (à savoir dès l'origine des données, avec le chercheur primaire qui dépose les données dans l'archive, jusqu'au centre d'archives qui devient distributeur des données aux utilisateurs secondaires). Un rapport de recherche sur ces questions juridiques complexes est nécessaire.

Références bibliographiques

- Corti, L. (2000, December). Progress and Problems of Preserving and Providing Access to Qualitative Data for Social Research—The International Picture of an Emerging Culture [58 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [Online Journal], 1(3). Available at: <http://www.qualitative-research.net/fqs-texte/3-00/3-00corti-e.htm>
- Corti, L. (2003): Infrastructure Services and Needs for the Provision of Enhanced Qualitative Data Resources. In: *International Social Science Journal*, 417-432.
- Corti, L. (2005): Re-using Archive Qualitative Data – Where, How, Why? (unpublished)
- Corti, L., & Wright, M. (2002): Consultants' Report to the Medical Research Council on the MRC Population Data Archiving and Access Project. UK Data Archive, University of Essex, UK.
- Corti, L.; Thompson, P. (2004): Secondary Analysis of Archive Data. In: C. Seale, G. Gobo, J. F. Gubrium & D. Silverman (Eds.), *Qualitative Research Practice* (pp.327-343). London: Sage.
- Corti, L., & Backhouse, G. (2005): Acquiring Qualitative Data for Secondary Analysis. In: M. Max Bergman & Thomas S. Eberle (Eds.). *Qualitative Inquiry: Research, Archiving, and Reuse*. *Forum Qualitative Sozialforschung/ Forum: Qualitative Social Research* [On-line Journal], 6(2), Art. 36. Available at: <http://www.qualitative-research.net/fqs-texte/2-05/05-2-37-e.htm>
- Corti, L., Witzel, A., & Bishop, L. (2005). Sekundäranalyse qualitativer Daten / Secondary Analysis of Qualitative Data. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* [On-line Journal], 6(1). Retrieved Aug. 3, 2005, from <http://www.qualitative-research.net/fqs/fqs-d/inhalt1-05-d.htm>
- Gläser, J., & Laudel, G. (2001): Re-Analyse als Vergleich von Konstruktionsleistungen. In: *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [Online Journal], Vol. 1, Nr. 3, verfügbar über <http://qualitative-research.net/fqs/fqs.htm>
- Hauptmann, S. (2005): Das "C-TOC" zur Gliederung von Audiodaten. Ein Beispiel für die qualitative Analyse am Rohmaterial. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], 6(1), Art. 33. Verfügbar über: <http://www.qualitative-research.net/fqs-texte/1-05/05-1-33-d.htm>.
- Klingemann, H.-D.; Mochmann, E. (1975): Sekundäranalyse. In: van J. Koolwijk; M. Wieken-Mayer (Hrsg.): *Techniken der empirischen Sozialforschung*, Bd. 2, München, 178 – 194.
- Kluge, S.; Opitz, D. (1999): Die Archivierung qualitativer Interviewdaten: Forschungsethik und Datenschutz als Barrieren für Sekundäranalysen? In: *Soziologie, Mitteilungsblatt der DGS*, Heft 4, 48-63.
- Medjedović, I., & Witzel, A. (2005). Sekundäranalyse qualitativer Interviews. Verwendung von Kodierungen der Primärstudie am Beispiel einer Untersuchung des Arbeitsprozesswissens junger Facharbeiter. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], 6(1). Retrieved 02.03.2006, from <http://www.qualitative-research.net/fqs-texte/1-05/05-1-46-d.htm>
- Müller, W., Blien, U., Knoche, P., Wirth, H. et al. (1991): Die faktische Anonymität von Mikrodaten. *Schriftenreihe Forum der Bundesstatistik*, Band 19, Stuttgart: Metzler Poeschel.
- Muhr, T. (2000): Increasing the Reusability of Qualitative Data with XML. In: *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [Online Journal], 1(3). Available at: <http://www.qualitative-research.net/fqs-texte/3-00/3-00muhr-e.htm>
- Opitz, D., & Mauer, R. (2005). Erfahrungen mit der Sekundärnutzung von qualitativem Datenmaterial – Erste Ergebnisse einer schriftlichen Befragung im Rahmen der Machbarkeitsstudie zur Archivierung und Sekundärnutzung qualitativer Interviewdaten. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], 6(1). Retrieved June 8, 2005, from <http://www.qualitative-research.net/fqs-texte/1-05/05-1-43-d.htm>
- Opitz, D., & Witzel, A. (2005): The Concept and Architecture of the Bremen Life Course Archive. In: M. M. Bergman; T. S. Eberle (Eds.). *Qualitative Inquiry: Research, Archiving, and Reuse*. *Forum Qualitative Sozialforschung/ Forum: Qualitative Social Research* [On-line Journal], 6(2), Art. 37. Available at: <http://www.qualitative-research.net/fqs-texte/2-05/05-2-37-e.htm>.
- Pätzold, H. (2005): Sekundäranalyse von Audiodaten. Technische Verfahren zur faktischen Anonymisierung und Verfremdung. *Forum Qualitative Sozialforschung / Forum: Qualitative Social*

Research [On-line Journal], 6(1), Art. 24. Verfügbar über: <http://www.qualitative-research.net/fqs-texte/1-05/05-1-24-d.htm>.

- Simitis, S. (1984): Die informationelle Selbstbestimmung. Grundbedingung einer verfassungskonformen Informationsordnung. In: Neue Juristische Wochenschrift, 37. Jg., Heft 8, 398-405
- Strauss, A. L.; Corbin, J. (1996): Grounded Theory: Grundlagen Qualitativer Sozialforschung. Weinheim: Beltz, PVU.
- UK Data Archive (2003). Qualidata Process Guide. Qualitative Data Processing