

Workshop Forschungsinfrastrukturen für die qualitative Sozialforschung
11. und 12. Oktober, Wissenschaftszentrum Berlin

Impulsreferat zur offenen Diskussion:
Wer entscheidet, welche Daten wie lange aufgehoben werden?
Andreas Witzel

Insbesondere die Frage „wie lange Daten aufgehoben werden sollen - verweist auf die Vorschläge der DFG zur Sicherung guter wissenschaftlicher Praxis (ursprünglich 1998 und dann 2009), die dafür plädieren,

Primärdaten nach Abschluss der Auswertung *für zehn Jahre*
- *entweder in der Institution, wo sie entstanden sind, aufzubewahren oder*
- *bei einer unabhängigen Stelle zu hinterlegen.*

Zum einen sollten mit der Aufbewahrung auch von sozialwissenschaftlichen Daten die *Transparenz* der Forschung und auch der *arbeitsökonomische Rückgriff* der Forschungsgruppe auf die Daten gewährleistet sein.

Zum anderen sollten mit der Aufbewahrung die Voraussetzungen für die Möglichkeit der *Überprüfung wissenschaftlicher Aussagen durch andere* geschaffen werden.

Es mag schon sein, dass bei diesen beiden Gründen für die Aufbewahrung von Daten – Transparenz und Kontrolle – die empfohlenen 10 Jahre der Aufbewahrung ausreichend sein könnten. Gedacht hat man damals allerdings ganz offensichtlich *nicht an die Möglichkeiten der systematischen Nachnutzung von Primärdaten in Forschung und Lehre*, die eine längerfristige, ja eher unbegrenzte Aufbewahrung voraussetzt. Dies spiegelt die damals noch weitgehend *fehlende Tradition der Forschungsstrategie der Sekundärnutzung*, zumindest der qualitativen Daten wieder, die das ALLF auch erst etwa ab dem Jahr 2000 in Deutschland publik gemacht hat.

Die Möglichkeiten der *Nachnutzung* von Primärdaten in Forschung und Lehre eröffnen nunmehr einen neuen Blick auf diese erste Initiative der DFG, Forschungsdaten aufzubewahren:

1. Entscheiden sich Institutionen wie z.B. Institute, in denen die Daten entstanden sind, für den Vorschlag der *dortigen längerfristigen Hinterlegung der Daten*, stehen sie vor der Frage, ob die vorhandenen Kompetenzen und Ressourcen überhaupt die geforderten arbeitsökonomischen Rückgriffsmöglichkeiten auf die Daten, deren Auffindbarkeit für potenzielle Nutzer in der scientific community sowie eine Datensicherheit über die Jahre erlauben. Die Machbarkeitstudie (2003 bis 2005) verweist darauf, dass die Institute nicht in der Lage sind, diese Funktionen wahrzunehmen.
2. Es ist nunmehr zu fragen: Ist nicht die erwähnte Alternative einer Hinterlegung der Daten bei einer *unabhängigen Stelle* so auszubuchstabieren, dass Primärforschende überzeugt werden sollten, ihre Daten einem Datenarchiv, besser: einem Datenservicezentrum zu übermitteln? Ist dieses nicht organisatorisch und technisch eher in der Lage, die Daten für den langfristigen Gebrauch aufzubereiten, zu verwalten, den Datenschutz und den Zugriff von Nutzern zweckbestimmt und verantwortungsvoll zu sichern?
3. Weiterhin: Befördert der in der Empfehlung enthaltene Kontrollgedanke, d.h. das Bezweifeln von Auswertungsergebnissen als so zentrale Begründung für die

Datenaufbewahrung nicht nur die *Angst von Primärforschenden, einer Karriere-schädlichen Konkurrenz ausgesetzt zu sein*? Mit der unerwünschten Folge, sich dafür zu entscheiden, keine Daten zur Verfügung zu stellen. Louise Corti nennt dieses Phänomen „fear of exposure“.

4. Und: Ist der *Kontrollgedanke nicht vielmehr positiv* zu wenden? Also als Anregung für die Verbesserung der Wissenschaft zu nehmen und gerade auch in der qualitativen Forschung offene, konstruktive Diskurse und Teamarbeit zu betonen? Sollten nicht gerade Datenservicezentren zu einer Kultur des datasharing beitragen? Wird dann nicht die Entscheidung, Daten zur Verfügung zu stellen, wesentlich erleichtert?
5. Oder sollte diese vernünftige Voraussetzung des data sharing doch durch *restriktivere Regeln* ersetzt oder ergänzt werden? Etwa durch die Verpflichtung Drittmittel finanzierter Projekte zu einer späteren Datenübergabe durch die Drittmittelgeber, wie dies etwa in England geschieht?
6. Zuletzt zur Frage, *welche Daten* aufgehoben werden sollten:
 - Einen Konsens gibt es wohl am Ehesten bei Daten, bei denen es *Zustimmung zur Nachnutzung sowohl von den Befragten wie von den Primärforschenden* gibt und die davon unabhängig mit der Sorgfalts- und Fürsorgepflicht eines Datenservicezentrums anonymisiert worden sind.
 - „Welche Daten“ heißt des Weiteren: *Nicht einfach nur Erhebungsdaten*, etwa Interviewtranskripte, sondern zusätzlich alle Formen von Metadaten, die in der Lage sind den Kontext der Befragung wie der gesamten Forschung optimal zu beschreiben.
 - Eine schwierigere Fragen ist die nach einer *selektiven Akquisition von Daten* auf der Basis von *Qualitätskriterien* (wie z.B. Größe der Stichprobe, Qualität der Interviewführung oder der Transkription). Haben Datenservicezentren hier nicht eher die Aufgabe, zur Verbesserung wissenschaftlicher Praxis in anderer Form beizutragen? Etwa zu der Entwicklung von Standards der Nachnutzung von Primärdaten (z.B. Forschungsdokumentationen, Anonymisierungsstrategien) und des Datenmanagements von Forschungsprojekten?