

# Contemporary challenges and future perspectives for sustainable reuse of archived textual qualitative research data

**Reinhard Messerschmidt, Brigitte Mathiak**  
Forschungsdaten in den Geisteswissenschaften  
(FORGE 2016)

15.9.2016 - Universität Hamburg



Cologne Center for eHumanities (CCeH)  
University of Cologne



# Ausgangsfrage und -situation

- Wie können Methoden der Digital Humanities Annotationsverfahren der qualitativen Forschung erleichtern und verbessern?
- Software zur qualitativen Datenanalyse (QDA) wird in vielen Disziplinen immer öfter genutzt, wobei primär Texte mittels regelgeleiteter Codierstrategien (z.B. Grounded Theory) manuell annotiert werden



# z.B. für qualitative Interviews

The screenshot shows the MAXQDA 11 software interface. On the left, there are two panels: 'Liste der Dokumente' (List of Documents) and 'Liste der Codes' (List of Codes). The 'Liste der Dokumente' panel shows a tree structure of documents, including 'Interviews' and 'Bilder Klimawandel'. The 'Liste der Codes' panel shows a tree structure of codes, including 'WP - Größe Weltzerbreche' and 'WP - Größe'. The main window displays a document titled 'Interviews\B01 Jan' with a transcript of an interview. The transcript is numbered 1 through 17. Various codes are applied to different parts of the transcript, indicated by colored brackets and labels on the left side of the text. For example, 'WP - Größe' is applied to paragraphs 7, 8, 10, 11, 12, and 13. 'WP - Größe Weltzerbreche' is applied to paragraph 14. 'WP - Größe' is also applied to paragraph 15. 'WP - Größe' is applied to paragraph 16. 'WP - Größe' is applied to paragraph 17. The transcript text is as follows:

1 Interview mit Jan

2 I: Ok. Ja hallo.

3 B: Ja hallo, schönen guten Tag.

4 I: Schön, dass du Zeit für mich hast. Wir haben ja schon telefoniert und du weißt ja schon worum es geht.

5 B: Ja ich weiß so, du machst ein Interview. Und du willst mir ein paar Fragen stellen, wo ich was zu sagen soll.

6 I: zum Umweltverhalten. Ok, dann fang ich einfach mal mit der ersten Frage an. Und die lautet, was sind aus deiner Sicht die größten Probleme der Welt im 21. Jahrhundert generell?

7 B: Joa, das sind die jetzt die aktuellen Probleme die wir haben. Es gibt ja da die materiellen Probleme, dass wir halt Energieprobleme haben, wie regeln wir unsere Energieversorgung für die Zukunft? Wo gehen wir hin? Irgendwann sind die fossilen Sachen sind alle. Und welche Alternativen sprich diese Bioenergie und Solarenergie ähm... weniger Vor- und Nachteile. Wasserstoff ist noch nicht soweit, dass man sagen könnte dass es ausgereift ist. Das gibt es ja noch nicht aufm Markt. Das sehe ich als ein ganz großes Problem. Dann natürlich auch Klimaveränderungen, wobei da auch viel Hysterie drin ist, wir brauchen jetzt innerhalb von einem Jahr nicht alles runddrehen, was in den letzten 60 Jahren in die Atmosphäre geblasen worden ist, also das sollte man nicht so hektisch betrachten. Oder dann ist das natürlich so, dass wir auch ein emotionales Problem haben und zwar ist die Zeit zu schnelllebig geworden. Wir nehmen uns zu wenig Zeit für wichtige Sachen wie Gespräche für Familie und diverseres und die moderne Elektronik die es so gibt, zum Beispiel Handys, klingeln rund um die Uhr und wir kommen gar nicht mehr zur Ruhe. Das ist auch ein Problem, wo vielleicht auch die eine oder andere Krankheit irgendwo entsteht.

8 I: Gut, sonst noch irgendwelche Probleme was jetzt nicht mit dem Klimawandel zu tun hat, generelle Probleme, große Probleme die es auf der Welt gibt.

9 B: Ja gut, wir haben noch Hungersnöte, die Kriege, das sind auch große Probleme, die sind weiter weg für uns und nicht sichtbar, aber sie sind da. ....muss auch was tun, der Wechsel jetzt in Amerika vom Präsidenten, dass der die aggressive, das aggressive Verhalten gegenüber des Westens von den von den Ländern und wie das damals auch mit dem Anschlag war, auf die TwinTowers .... nicht mehr so aggressiv dann wie das war, dass das sich schon mal überlegt, es war ja nun so, dass die sich überall eingemischt haben....und....

10 I: Und das findest du gut, dass es so ist?

11 B: Dass sie sich einmischen?

12 I: Das ne...das wie es jetzt ist mit dem neuen Präsidenten.

13 B: Ich ... Ländern, die wo Streitigkeiten gibt.... dass man versucht, dass die sich auch untereinander einigen und dass man nicht sagt, wir sind jetzt die Starke, wir gehen dahin rücken die zurecht und sorgen dafür, dass es dann da ruhig ist. Wenn die das nicht selber packen, ist das mit Sicherheit nicht von Dauer und einer von denen, wenn nicht sogar Beide, ärgern sich, dass dann einer sich dann anmaßt, dahin zukommen ein bisschen was kaputtschießen, und dann wieder abzuhauen.

14 I: OK. Jetzt zur 2. Frage. Die bezieht sich auf die Erste, wie denkst du kann mit den Problemen umgegangen werden, sind sie prinzipiell überhaupt beeinflussbar und wenn ja, von wem?

15 B: Zum ersten Teil der Frage, mit den Problemen, denke ich kann umgegangen werden. Und das ist natürlich auch die Pflicht unserer Führung, Politik damit unzugehen, die Wirtschaft zu lenken, aber es ist vielmehr Aufgabe von unserer Weltwirtschaft von den Energieerzeugern, von den Versorgern, aus ethischer Sicht, schon aus moralischer Sicht, fair zu arbeiten, nicht nur gewinnorientiert zu arbeiten und nicht immer nur für die Rendite der Aktionäre zu arbeiten, weil viele sind ja dann auch Aktiengesellschaften und dass man dann einfach auf Langfristigkeit, Langlebigkeit setzt und sich teilweise auch selbst zurückstellt, dass man einfach sagt, man sieht das Große und Ganze und stellt diesen Gewinnzwang, den muss man zurückstellen, das muss man auch in die Köpfe reinkriegen von den großen Unternehmen, die es ja weltweit gibt, weltweit vernetzt sind, und da ist natürlich die Politik gefordert, die irgendwo in dem Maße zu lenken, dass hat uns jetzt ja auch die Finanzkrise gezeigt hat, dass die dann jetzt auch mit in die Vorstände einsteigen. Und dann auch kontrollieren, Rechenschaft haben wollen, was da passiert im Betrieb, wenn ihr Förderung kriegt.

16 I: Dass das vorher nicht passiert ist.

17 B: Das ist ja so, dass heutzutage der Hartz IV - Empfänger, der muss alles angeben, der geht betteln aufs Amt, und wenn du aber Subventionen haben willst, dann sagst du ich mache das und das und dafür gibt es dieses Förderungsprogramm, und da kriegst du die Förderung. Also ich weiß das jetzt aus der Landwirtschaft, da brauchst du nicht hingehen und sagen, mein Konto ist so und so voll oder so und so leer, ich mache jetzt 100 ha Landwirtschaft. Ich brauche Förderprämien, die kriegst du. Weil bestes Beispiel ist Prinz Charles, da weiß ich es. Ich glaub das weiß die ganze Welt, dass der Mann Geld hat. Und der hat 3500 ha Landwirtschaft, der kriegt für 3500 ha Landwirtschaft die pro-ha-Prämie und ein kleiner Landwirt hier in Deutschland zum Beispiel, der mit 40 ha arbeitet, davon vielleicht noch versucht seine Familie zu ernähren, was schon fast nicht mehr geht, wenn dann nicht noch einer aus der Familie arbeiten geht, kleine Maschinen kauft, und auch sich in den Dörfern mit einbringt, wenn was zu fahren ist, wenn was zu organisieren ist, sind die auch überall dabei. Dieser Mann kriegt genau prozentual die gleiche Prämie wie jetzt Prinz Charles oder wie natürlich auch andere große Genossenschaften hier in Deutschland, vor allen Dingen auch in den neuen Bundesländern wie Thüringen, was ja damals durch den Einfluss des Russen immens groß geworden ist, kassieren da riesen Prämien, aber brauchen die eigentlich gar nicht.

18 I: Also das findest du nicht so gut.

19 B: Das finde ich nicht gut, ich bin der Meinung, dass ... und genauso, wenn man Förderung haben will, Rechenschaft ablegen muss, sich selbst offenbaren muss, und vom Prinzip her wie ein Hartz IV - Empfänger auch, wer was haben will vom Staat, der muss dem sagen, das hab ich, das bin ich, und der muss dann auch ... sich offen legen. Und jemand, der Geld hat, der braucht keine Förderung, weil der kann ja sein Geld erstmal einsetzen, dass ist natürlich jetzt einfach gesagt, man muss natürlich dann das so umsetzen, dass es dann auch wieder keine Ungerechtigkeiten gibt. Nicht wenn sich jemand, ein gewisses Grundkapital erarbeitet hat, was er braucht zur Sicherung seines Unternehmens, dass der das erstmal aufzählen (aufbringen?) muss. Da muss man Grenzen setzen, aber mit Sicherheit ist ... kann das nicht sein, dass wirklich wenn da riesen Gewinne gemacht werden, und es Größenordnungen sind, wo ein Betrieb selbständig arbeiten kann, da braucht er keine Subventionen. Und dann kann man da auch einsparen.

# ...qualitative Inhaltsanalysen

The screenshot shows the MAXQDA 11 software interface. On the left, there are two panels: 'Liste der Dokumente' (List of Documents) and 'Liste der Codes' (List of Codes). The 'Liste der Codes' panel shows a hierarchical code system with various categories and their frequencies. On the right, the 'Dokument-Browser' (Document Browser) displays an interview transcript with three numbered questions. The text is in German and discusses the development of bibliometric services at the University of Loughborough. The questions are:

1. You set up the discussion list Lis-Bibliometrics in 2010 together with Jenny Delasalle. Are you still actively promoting and using it? Are you happy with its development in the past years?
2. Setting up that list clearly showed your high professional interest in bibliometrics. Where did you receive your training in bibliometrics and at what point in your career?
3. Before your secondment to the universities Change Project team, you worked as an Academic Services Manager and Research Support Coordinator in the library of the University of Loughborough. How and when did the bibliometric services at University of Loughborough come into being?

The interview text includes responses to these questions, such as: 'I wouldn't say I'm actively promoting it - partly because bibliometrics is such a small part of my part-time role - and actually I'm on secondment to another part of the University at the moment too. I would like to see it better used (I set up Lis-Copyseek which is heavily used) but I think bibliometrics has quite a competitive element and people don't want to share too much about how they are using them.' and 'When I started (1996) academics were interested in Journal Citation Reports (then on microfiche) I would say we started teaching bibliometrics about 6 years ago when they started being of more mainstream interest and tools like the WOS Researcher ID came online.'

# ...oder Diskursanalysen

D:\Diss\_Korpus\_Analyse\2015-02-17 FINAL incl. IZA.mx5 - MAXQDA 11 (Release 11.1.0)

Projekt Bearbeiten Dokumente Codes Summaries Memos Variablen Analyse Mixed Methods Visual Tools Fenster MAXQDico Symbolleisten ?

Liste der Dokumente

Liste der Codes

Dokument-Browser: BILD\2006-03-15

**Bild.de**

## BABY-SCHOCK! DIE WENIGSTEN GEBURTEN SEIT 1945

### In 12 Generationen sind wir Deutschen ausgestorben

15.03.2006 — 00:11 UHR

Baby-Schock in Deutschland! Nie zuvor wurden in der Bundesrepublik weniger Kinder geboren als heute: Nur 676 000 Kinder kamen 2005 zur Welt, der niedrigste Stand seit dem 2. Weltkrieg, meldet das Bundesamt für Statistik.

Experten warnen: Wir Deutschen sterben aus, in kaum mehr als 12 Generationen!

Die Fakten: In den nächsten 44 Jahren (bis 2050) wird die Zahl der Deutschen um 12,5 Mio. sinken – von heute 82 auf knapp 70 Mio. Menschen. „Der Trend setzt sich fort, wenn nichts geschieht“, warnt Deutschlands bekanntester Bevölkerungswissenschaftler Prof. Herwig Birg: „2100 wird die Zahl der Deutschen auf 46 Mio. geschrumpft sein, 2300 liegen wir bei 3 Mio., also kurz vorm Aussterben“ – und das in nur 12 Generationen! Die Deutschen, eine bedrohte Spezies? Dabei leisten wir uns ein Sozialsystem, das jedes Jahr 100 Milliarden Euro in die Förderung von Kindern und Familien steckt:

34 Mrd. gibt der Staat allein für das Kindergeld aus
44 Mrd. für Schulen und Lehrer
3,5 Mrd. für Mutterschafts- und Erziehungsgeld
Immerhin 3,6 Mrd. für Kinderzuschläge an Beamte.

100 000 000 000 Euro, das sind Jahr für Jahr 6750 Euro für jedes Kind in Deutschland, über 120 000 Euro für ein Kinderleben bis zum 18. Lebensjahr!

Trotzdem verharren die Deutschen im Gebärstreik: Im Schnitt bekommen deutsche Paare gerade mal 1,38 Kinder. Nur Österreich, Italien, Spanien und Griechenland schneiden noch schlechter ab. Dirk Konietzka vom Max-Planck-Institut für Bevölkerungsforschung warnt bereits vor einer „Kultur der Kinderlosigkeit“: „Rund ein Viertel aller westdeutschen Frauen bleibt kinderlos. Das ist Weltspitze.“

Um den schleichenden Untergang der Deutschen zu bremsen, „reicht Geld vom Staat nicht aus“, glaubt Dr. Reiner Klingholz (52), Chef des Berlin-Instituts für Bevölkerung und Entwicklung. Klingholz nennt die wichtigsten Gründe für den Rückgang der Geburtenraten:

# Ziele

- a) methodologische Forschung: Untersuchung von Datenstrukturen und der komplexen Entstehung von Codesystemen im Zusammenhang mit der Spezifität von Disziplin, Methode und Daten
- b) Potenziale und Grenzen zukünftiger Tools zur Unterstützung der zeitintensiven Arbeit qualitativer Forscher/innen bzw. Ermöglichung größerer Korpora



# (zunächst) geplantes Vorgehen

1. **Aufbau interdisziplinärer Sammlung von Korpora**
2. *wenn möglich, Standardisierung der Formate*
3. *Entwicklung datengetriebener Typologie bzgl. Struktur und Methoden bzw. Codesystem*
4. *Untersuchung diachroner „Evolution“ innerhalb von Codesystemen (dazu Zwischenversionen nötig)*
5. *Testen möglicher Automatisierbarkeit repetitiver Kodierung nach Erreichen des Sättigungspunkts*
6. *Testen computerlinguistischer Verfahren*



# Online-Archiv Recherche (1)

- Deutschsprachig: Qualiservice, FORS
  - nur Datensätze ohne Codesystem, oft Interviewtranskripte
- Europa: DANS (NL), FSD (Finnland)
  - ebenfalls nur Datensätze ohne Codesysteme verfügbar
- International:
  - Australien (ADA): nur 5 relativ alte qualitative Studien
  - USA: trotz 511 Treffer beim Harvard Dataverse Network und 364 im ICPSR Michigan oft nur links zu Uni-Archiven oder Autorenkontakt + generell **IRB-approval** nötig (ebenso bei Qualitative Data Repository der Syracuse University etc.)





# Online-Archiv Recherche (2)

## – UK/IRL:

- UK Data Service: 888 qualitative oder mixed methods Datensätze zunächst sehr vielversprechend, aber keine Suche von Dateitypen, also manuelle Selektion + Download nötig > am Ende lediglich 4 Nvivo Datensätze, davon nur 3 mit Codesystem
- Irish Qualitative Data Archive (IQDA) > im Aufbau
- ARK Qualitative Archives (Nordirland) > lediglich Interviewtranskripte



# Zwischenfazit

- Trotz stetig zunehmender Anwendung qualitativer Forschungsmethoden in einer Vielzahl von Disziplinen sind in deutsch- und englischsprachigen Archiven bislang kaum annotierte Datensätze, sondern in der Regel nur Transkripte und Rohdaten archiviert.
- vgl. „3 challenges“ (Bischoff et al. 2014):
  - accessibility, findability, and interpretability ???
- open data/ methodology ?



# Schneeballverfahren?

- Alternativ: persönliche Kontakte + Mailingliste des Diskursnetz (4692 registrierte User)
- Feedback im unteren zweistelligen Bereich, damit nicht nutzbar



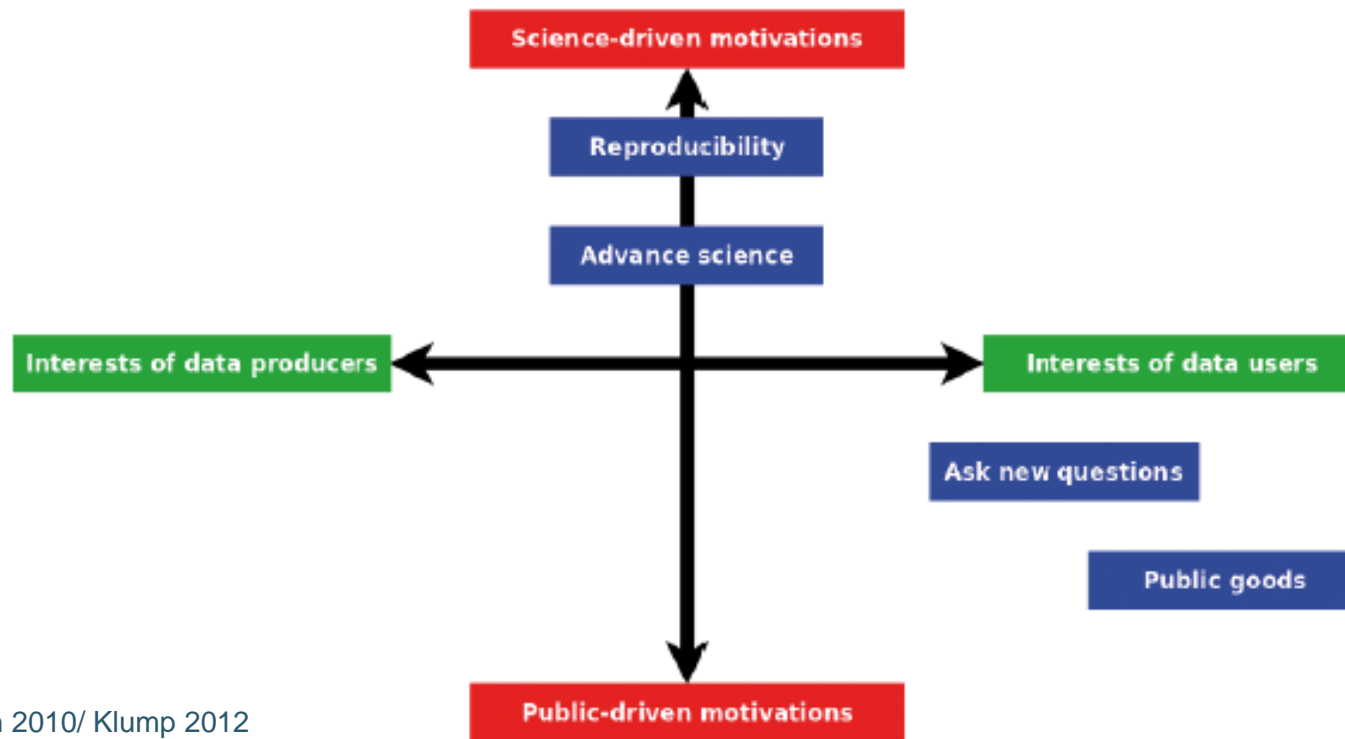
# Erklärungen (Literatur)

- mangelnde Incentivierung
- Vermeidung epistemologischer und methodologischer Kontroversen (QUAL vs. QUAN)
- mangelnde Nachnutzung für Sekundäranalysen (auch da Themen/ Sampling hochgradig spezifisch)
- Hauptzweck bislang immer noch Lehre, dafür Annotationen nicht zwangsläufig nötig



# Erklärungen (schematisch)

Motivations and interests in sharing data



Quelle: Borgman 2010/ Klump 2012



Cologne Center for eHumanities (CCeH)  
University of Cologne



# Konsequenzen

- unser Forschungsvorhaben ist in zunächst geplantem methodologischem Rahmen bislang nicht umsetzbar
- Alternative: Fokus auf methodische Unterstützung einzelner Forschungsprojekte
- Big Data picture bleibt so lange unvollständig, bis möglicherweise generelle Veränderungen der Wissenschaftskultur in Richtung open science zu veränderter Archivierungspraxis führen



# Vielen Dank für die Aufmerksamkeit!

## Kontakt:

Jun.Prof. Dr. Brigitte Mathiak [b.mathiak@uni-koeln.de](mailto:b.mathiak@uni-koeln.de)

Reinhard Messerschmidt, M.A. [reinhard.messerschmidt@uni-koeln.de](mailto:reinhard.messerschmidt@uni-koeln.de)



Cologne Center for eHumanities (CCeH)  
University of Cologne

