

Curating video data at the Research Data Centre for Education

Legal requirements and detailed documentations

Nadeshda Jung, Thomas Lösch, Sonja Bayer, Maike Porzelt, & Doris Bambey

CSPD2020
07. Dezember 2020

German Network of Educational Research Data (VerbundFDB)

- » Infrastructure partner for education sciences in Germany
- » Federated network
- » Cover the whole range of data types and disciplines



Qualiservice
data sharing

ApaeK
Arbeitskreis für pädagogische Statistik

MERCATOR
INSTITUT
für Sprachförderung
und Deutsch
als Zweitsprache

LifBi
LEIBNIZ-INSTITUT FÜR
BILDUNGSVERLÄUFE e.V.

dji

DIPF
Leibniz-Institut für Bildungsforschung
und Bildungsinformation

IQI
Institut zur Qualitätsentwicklung
im Bildungswesen

gesis
Leibniz-Institut
für Sozialwissenschaften

leibniz-psychology.org
Deutsches
Jugendinstitut

SOEP

fdz.DZHW.
Forschungsdatenzentrum
Deutsches Zentrum für Hochschul- und Wissenschaftsforschung

DiE

Research Data
Centre (RDC) for
Education:
Video data, e.g.
school lessons

Video Data in Psychology

Video Data in Psychology

- » Video data is well suited to measure a wide range of constructs
 - Actual behaviour as it unfolds
 - Spoken language, body posture, facial expression, gestures, dyadic or group interaction, etc.

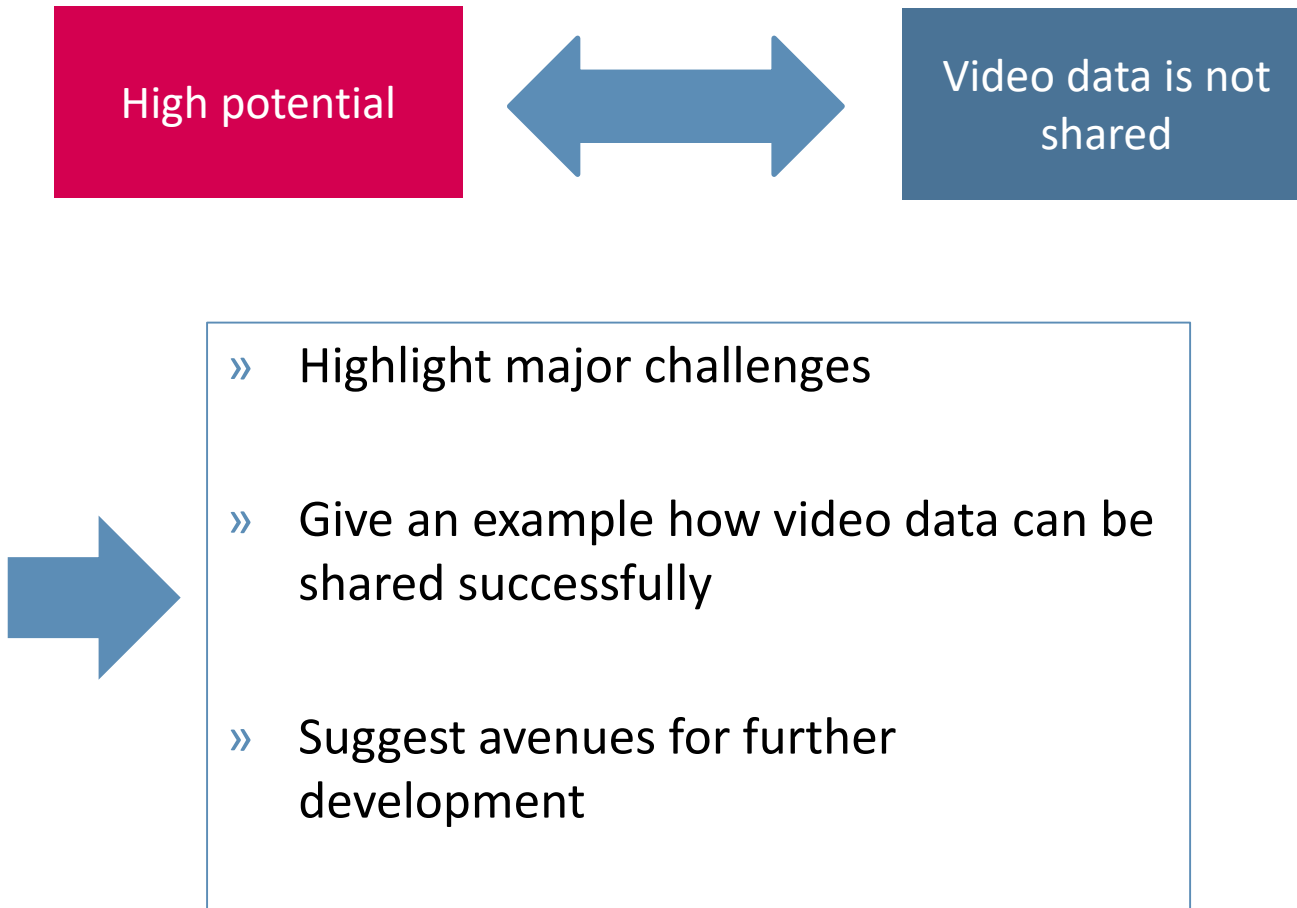
- » Examples:
 - Sánchez-Amaro et al. (2020): Economic game with children
 - Fauth et al. (2019): Teaching quality
 - Rentzsch & Löscher (2019): Emotion expression

- » Video data adds information beyond self-reports (behavioural traces, sensing data, informant reports...)

Reusability and need for transparency

- » Multiple methods to rate video data
 - Observers (number, training, expertise)
 - Items (number, content, way of presenting)
 - Sequence to be rated (length, timing, number)
 - Number of parallel perspectives
- » Multiple theoretical backgrounds to investigate video data
- » High potential for reusability
 - Primary researchers cannot exploit the full potential
- » Transparency especially relevant?
 - High number of researchers' degrees of freedom
- » Costly to collect

Goals for the presentation



Curation of video data

Characteristics and recommendations for data protection

Information density of video data

- » Large analysis potential for re-use or further research questions

but

- » very complex in curating
- » contain plenty of personal data and bear the risk of disclosure

Personal data in videography

- » Direct personal and person-related attributes:
Names, names of places and streets, federal states, institutions and organisations (e.g. companies, schools)
- » Indirect but specific personal related attributes:
Characteristics of the test persons as well as those of third parties
- » Sensitive informations:
Information on ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, health or sex life

Anonymisation of audiovisual data

» Acoustic options:

- Covering names with a tone
- Noise signal to prevent voices recognition
- Pitch change (lowering or raising voices)

» Visual options:

- In video files, faces can be masked or blurred

Problems in the implementation of anonymisation

- » Expensive and time consuming
- » Destruction of the analysis potential for secondary analysis



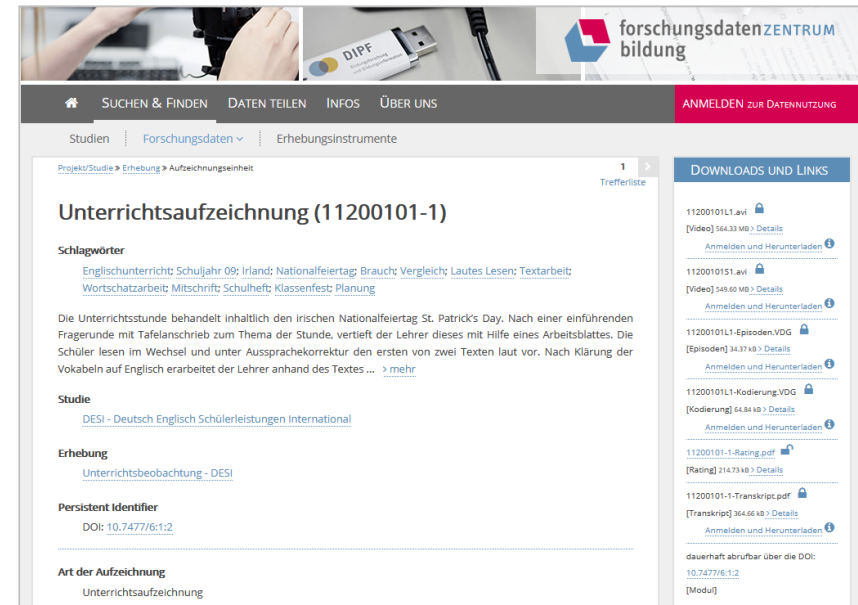
art_inthecity (o.D.): Diane Gougeon, Les iris, Art
souterrain 2014, Montréal. licensed under CC BY 2.0
<https://ccsearch.creativecommons.org/photos/1c93c849-29fc-42ce-8964-1e3f33c7bd21>

Alternatives to anonymisation

- » Obtain informed consent in advance so that the data can be used without anonymisation.
- » Avoid recording personal information during the collection process (e.g. by relocating students in the classroom).
- » Agree with the participants in advance not to mention identifiable characteristics – as long as this does not affect the informative value.
- » **Further step:** Provide access to this data via controlled access channels and only for certain user groups via a research data centre.

RDC for Education: Secure access

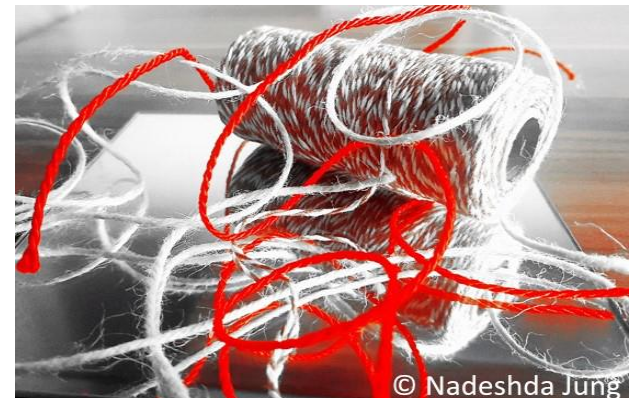
- » Data is stored in a protected directory
- » After registration an additional application for use is required (online form)
- » Individual examination of each application, PostIdent procedure
- » Agreement of use: for a maximum of three years with the possibility of renewal, strictly bound to contractual conditions
- » Authorised for non-anonymised data:
 - » only for proven qualified scientists
 - » only for defined research purposes
 - » condition: research is dependent on specific data
 - » NOT for teaching



Access levels of the RDC for Education

Practical tips for using video data

- » Be aware of the high content of personal data in videos
- » Use informed consent (examples <https://www.forschungsdaten-bildung.de/einwilligung>)
- » Perform anonymisation during the evaluation process, mark anonymised and sensitive passages (protocol)
- » Access via protected distribution channels




© Nadeshda Jung

Curation of video data

Context awareness

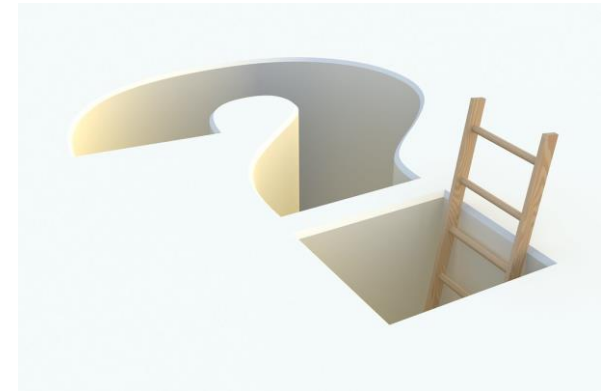
Reusability and interoperability of video data

- » Data management includes data protection as well as contextualisation.
 - » For the quality of the data, it is essential to document the contextualization in addition to the actual collection.
 - » Audiovisual data are framed according to the situation and are the result of interactions and mutual interpretations.
-  An elaborated and detailed documentation allows to understand the perspective of the primary researchers and thus to interpret the data.

Problems of contextualisation

In many cases, the relevant contextual information is taken for granted for the primary researchers and is therefore not documented.

- ➔ Data and project descriptions are often incomplete
- ➔ Reconstruction of the research process and thus the re-use of the data is either difficult or not possible



3D Problem Solving (2012): Chris Potter licensed under CC BY 2.0 .
<https://ccsearch.creativecommons.org/photos/01da9e81-d484-4ded-8f91-18ac0a1c7d3c>

Contextualisation on different levels

Level Study (Macro)	Level Setting (Meso)	Level Objekt (Micro)
<p>Background of the study (project managers, academic staff etc.)</p> <p>Description of the methodology and process of data collection (survey instruments, sample selection/field access, place and time of the survey etc.)</p> <p>Legal aspects (confidentiality, access rules and conditions of use, declarations of consent)</p>	<p>Meta information about the observation or the setting</p> <p>Information on the population (age, gender, education etc.)</p>	<p>Data editing (e.g. procedure for transcribing and anonymising)</p> <p>Modifications to the data; versioning</p> <p>Number and relationships of files (e.g. records) among each other</p> <p>Information about the file itself (file name, format, size etc.)</p>

Data published at the RDC for Education

- » Data from 12 studies, e.g. history of education "Audiovisuelle Aufzeichnungen von Schulunterricht in der DDR" or large, national studies such as "Pythagoras: Unterrichtsqualität, Lernverhalten und mathematisches Verständnis"
- » approx. 2100 Video and audio files
- » approx. 2500 text documents (transcripts of teaching observations, codings, episodes, ratings)
 - [Classroom Observation: DESI](#): German and English lessons in ninth grade; 213 observations; collected between 2003 – 2004; questionnaire survey and performance measurement are also available
 - [Classroom Observation: KuL](#): German and Mathematic lessons in primary school; 41 observations; collected in 2014; performance measurement is also available

Towards a culture of reusing video data

How to improve sharing of video data?

Infrastructure

- » Repository with secure data access required
- » Determine focus of a repository with regard to (sub-)disciplines
- » Documentation standards covering the wide range of material

Researchers

- » Awareness that it is legally possible to share video data
- » Awareness for funding opportunities
- » Knowledge of GDPR
- » Provide material, e.g. consent forms (that does not require data deletion)
- » Do researchers want to share video data?
- » Rewards for sharing

Thanks for your attention!

Questions?



Contact

Do you have further questions or suggestions on this topic?

German Network of Educational Research Data/
Verbund Forschungsdaten Bildung

(VerbundFDB)

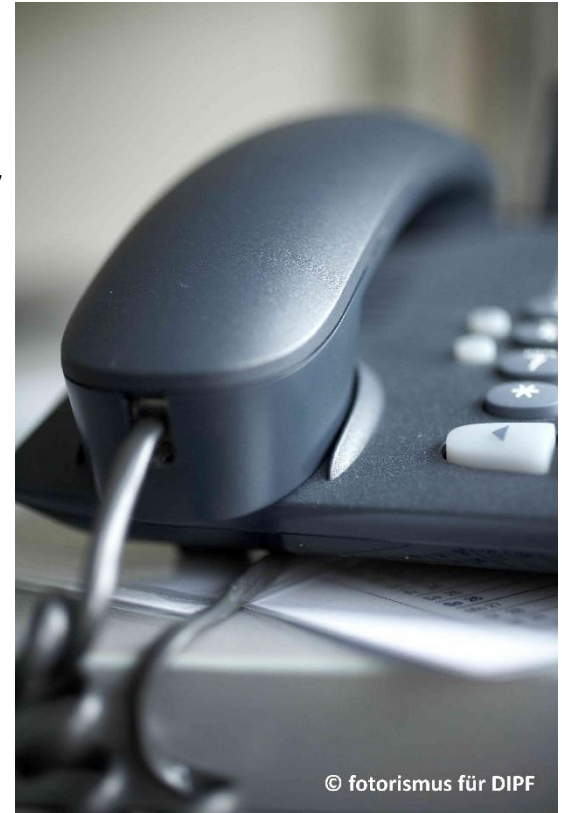
E-Mail: verbund@forschungsdaten-bildung.de

Phone: +49 (0)69 - 24708 - 300

Contact:

Nadeshda Jung
jung@dipf.de

Thomas Lösch
loesch@dipf.de



© fotorismus für DIPF

References

- Fauth, B., Decristan, J., Decker, A.-T., Büttner, G., Hardy, I., Klieme, E., & Kunter, M. (2019). The effects of teacher competence on student outcomes in elementary science education: The mediating role of teaching quality. *Teaching and Teacher Education*, 86, 102882. <https://doi.org/10.1016/j.tate.2019.102882>
- Metschke, R. & Wellbrock, R. (2002). Datenschutz in Wissenschaft und Forschung (Materialien zum Datenschutz.28). <https://www.hu-berlin.de/de/datenschutz/einwilligung/datenschutz-in-wissenschaft-und-forschung>
- Meyermann, A. & Porzelt, M. (2014). Hinweise zur Anonymisierung von qualitativen Daten. https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert-nr-1.pdf
- Jung, N., Porzelt, M., Kretzer, S., & Berdelmann, K. (2019). Kontextualisierungsbogen von qualitativen Daten. Leitfaden zur Kontextualisierung für Befragungsmethoden. Version 1.0, fdbinfo Nr. 9. https://www.forschungsdaten-bildung.de/files/fdbinfo_9_Kontextualisierung-von-qualitativen-Befragungsdaten_v1.0.pdf
- Rentzsch, K., & Lösch, T. (2019, July 30). PESI (Personality and Emotions in Social Interactions). <https://osf.io/2uf7a>
- Sánchez-Amaro, A., Duguid, S., Call, J., & Tomasello, M. (2020). Do 7-year-old children understand social leverage? *Journal of Experimental Child Psychology*, 199, 104963. <https://doi.org/10.1016/j.jecp.2020.104963>