

# Codebook

## Codebook abbreviations

API - Application programming interface

bsd - Big social data

bsr - Big social research

IP - Intellectual property

IRB - Institutional review board

Refs - References

qual - Qualitative

Name	Files	Refs
comparability - complexity of qual data	2	2
comparability - documentation and metadata	9	11
comparability - interoperability - formats, metadata, language, etc	11	14
comparability - matching of different datasets	5	6
comparability - more data = stronger conclusions	10	10
consent - as it applies to whole communities	2	3
consent - biggest challenge of the six	2	2
consent - bsd archiving for historical record or unknown future use	4	8
consent - concern that consent issues would affect participation	3	3
consent - consent language and procedures	15	28
consent - don't know what future uses might be	5	6
consent - harm analysis	3	5
consent - human subject vs historical data	2	2
consent - IRB	22	33
consent - little guidance	3	4
consent - participant review and redaction of transcripts	4	4
consent - participants may not understand nuances of consent form	4	5
consent - public vs. private	8	9
consent - re-consent	2	4
consent - repository terms of use	2	3
consent - research ethics education and literature	4	6
consent - sensitivity of data	5	5
consent - social media terms of service include consent	6	6
consent - some bsr platforms are more public than others	3	4

consent - taking care with direct quotes	7	9
consent - tiered consent	3	5
consent - uncommon for big social data	5	6
context - big social data - interface and features provides context	6	8
context - can't ask follow up questions of bsd or qual reuse participants	3	3
context - description, metadata, documentation to support context	13	21
context - different disciplinary expectations	3	3
context - different research design and methods provide different contextual info	4	4
context - filling gaps in hashtag data by retroactively collecting timeline data	2	2
context - good documentation is time consuming	7	9
context - in tension with privacy	10	12
context - in which data was posted or collected vs in which it will be used	4	4
context - including related materials with data	9	13
context - involve original authors for reanalysis	2	2
context - key to understanding reused data	3	3
context - look to existing literature for guidance	3	3
context - may be difficult to ascertain with bsr	7	8
context - misinterpretation may be inevitable	3	3
context - providing enough, but not too much information	3	3
context - representativeness of data	5	7
context - reproducibility	4	5
context - researchers, reusers, curators have different backgrounds	4	5
context - some data have more inherent context	4	5
context - trust in data creators	2	2
curation - collaborating with curators and repositories	7	8
curation - considering reusers needs when publishing data	2	2
curation - cost and time	10	11
curation - data sharing requirements	2	2
curation - for transparency	4	6
curation - good enough metadata is sometimes as good as it gets	2	2
curation - planning for data sharing makes it less of a hurdle	5	6
curation - qualitative data reuse is rare and hard to track	3	3
curation - repository SEO, data findability	2	2
curation - researcher reluctance to share data	3	3

curation - technical requirements of bsr and data reuse	4	8
curation - value of bsr and qual data sharing	10	18
domain differences - bsr collaboration with social scientists	4	4
domain differences - data sharing values and norms	12	19
domain differences - skills, training, and background	8	8
domain differences - research practices and standards	9	12
IP - checking with participants and organizations involved	3	3
IP - citation to support IP	5	7
IP - data licensing	6	7
IP - data sovereignty and ownership	7	11
IP - fair use	2	2
IP - lack of clarity about IP laws	5	5
IP - of social media users	2	2
IP - platform or data provider terms of service	13	28
IP - purchasing or using commercially-available data	8	10
IP - repository terms of use	3	3
IP - sharing a subset of data	2	2
privacy - assembling a lot of big data can threaten privacy	2	3
privacy - care to make sure quotes aren't identifiable	3	3
privacy - check back with participants	2	2
privacy - considering potential harms	10	14
privacy - data security	6	6
privacy - datasets provided by social media companies	2	2
privacy - deidentification	18	36
privacy - deletion requests	3	5
privacy - partial sharing to support privacy	2	3
privacy - participant expectations	10	13
privacy - research design	8	14
privacy - research ethics education and training	3	3
privacy - restricted access	11	25
privacy - sensitivity of data	11	11
quality - bsr data loss over time	2	2
quality - combining datasets to support quality	2	2
quality - curator review	4	6

quality - data completeness	10	15
quality - description, metadata, documentation support data quality	18	23
quality - issues with large-scale and automated collection	7	9
quality - look to existing literature for guidance	2	2
quality - representativeness of data	5	6
quality - researcher bias	3	3
quality - spam and bots	6	10
quality - trust in data creator	3	4
quality - understanding transcripts or videos as opposed to in person	3	5
strategies for responsible practice - appropriate research questions and scope	5	8
strategies for responsible practice - considering power dynamics of research	4	4
strategies for responsible practice - discussions with colleagues and collaborators	13	19
strategies for responsible practice - ethical guidelines for bsr	3	4
strategies for responsible practice - risk-benefit analysis	17	32

