QDR
QUALITATIVE DATA REPOSITORY

Transparency in Practice: Working with Annotation for Transparent Inquiry (ATI)

Colin Elman, Diana Kapiszewski,
and Sebastian Karcher, QDR

Qualitative Data Repository (QDR): QDR stores, preserves, and publishes digital data arising from qualitative and multi-method inquiry.

QDR is collaborating with **Hypothesis,** a software non-profit that develops open source software enabling the creation, storage, and sharing of web-standard annotations.

QDR

- Annotation for Transparent Inquiry (ATI) is a technique developed by QDR for achieving transparency in qualitative and multi-method research.
- ATI is based on the concept of “open annotation,” which allows for the generation, sharing, and discovery of digital annotations across the web.
- Holding two workshops to demonstrate and evaluate ATI. First involves authors of recently published articles; second authors of manuscripts in progress. Authors are drawn from health and social science disciplines.
- First workshop held, reporting some preliminary results.

QDR

QDR helps scholars to share their data to support three missions:

- to enable *secondary analysis*, i.e. so another scholar can analyze the data
- for *pedagogical purposes*
- to achieve *research transparency*

QDR

Research Transparency

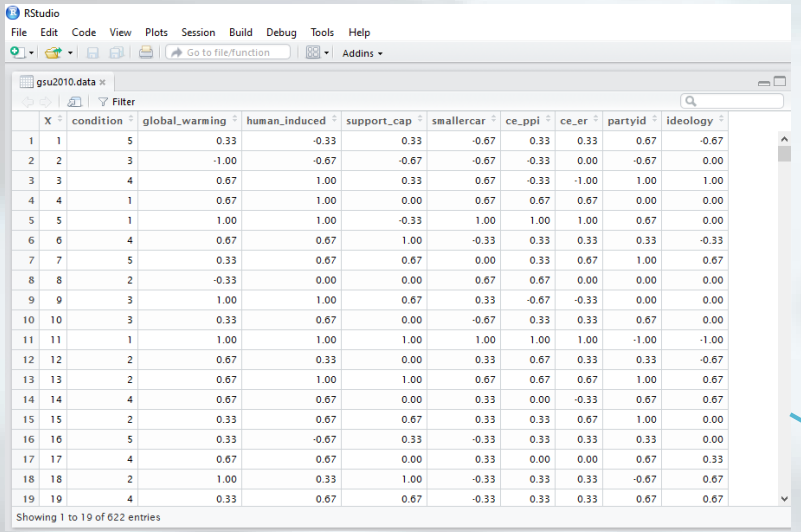
- Evidence-based inquiry is process-dependent, but so are almost all intentionally created outcomes.
- Not just that it is process-dependent, but that the central claim is that "the content is the method."
- To put it another way, the process is part of the outcome. If you can't see how the result was arrived at, you can't see all of the result.
- Principle applies to all rule-bound evidence based inquiry, but application needs to respect the challenges and opportunities that characterize various research traditions.
- Much better understandings and infrastructure for achieving transparency with quantitative than qualitative approaches to research.

QDR

Contrasting traditions of social inquiry

- Different ways to organize and typologize research.
- For example, could be by ontology (topic, puzzle, mechanism), discipline, episteme.
- From a transparency perspective, helpful to separate epistemically, by type of data and mode of analysis. In evidence-based inquiry focused on causal analysis, for example, can locate two central groupings: dataset and singular sources.

Quantitative Research: Matrix Data



	X	condition	global_warming	human_induced	support_cap	smallercar	ce_ppi	ce_er	partyid	ideology
1	1	5	0.33	-0.33	0.33	-0.67	0.33	0.33	0.67	-0.67
2	2	3	-1.00	-0.67	-0.67	-0.67	-0.33	0.00	-0.67	0.00
3	3	4	0.67	1.00	0.33	0.67	-0.33	-1.00	1.00	1.00
4	4	1	0.67	1.00	0.00	0.67	0.67	0.67	0.00	0.00
5	5	1	1.00	1.00	-0.33	1.00	1.00	1.00	0.67	0.00
6	6	4	0.67	0.67	1.00	-0.33	0.33	0.33	0.33	-0.33
7	7	5	0.33	0.67	0.67	0.00	0.33	0.67	1.00	0.67
8	8	2	-0.33	0.00	0.00	0.67	0.67	0.00	0.00	0.00
9	9	3	1.00	1.00	0.67	0.33	-0.67	-0.33	0.00	0.00
10	10	3	0.33	0.67	0.00	-0.67	0.33	0.33	0.67	0.00
11	11	1	1.00	1.00	1.00	1.00	1.00	1.00	-1.00	-1.00
12	12	2	0.67	0.33	0.00	0.33	0.67	0.33	0.33	-0.67
13	13	2	0.67	1.00	1.00	0.67	0.67	0.67	1.00	0.67
14	14	4	0.67	0.67	0.00	0.33	0.00	-0.33	0.67	0.67
15	15	2	0.33	0.67	0.67	0.33	0.33	0.67	1.00	0.00
16	16	5	0.33	-0.67	0.33	-0.33	0.33	0.33	0.33	0.00
17	17	4	0.67	0.67	0.00	0.33	0.00	0.00	0.67	0.33
18	18	2	1.00	0.33	1.00	-0.33	0.33	0.33	-0.67	0.67
19	19	4	0.33	0.67	0.67	-0.33	0.33	0.33	0.67	0.67

Showing 1 to 19 of 622 entries

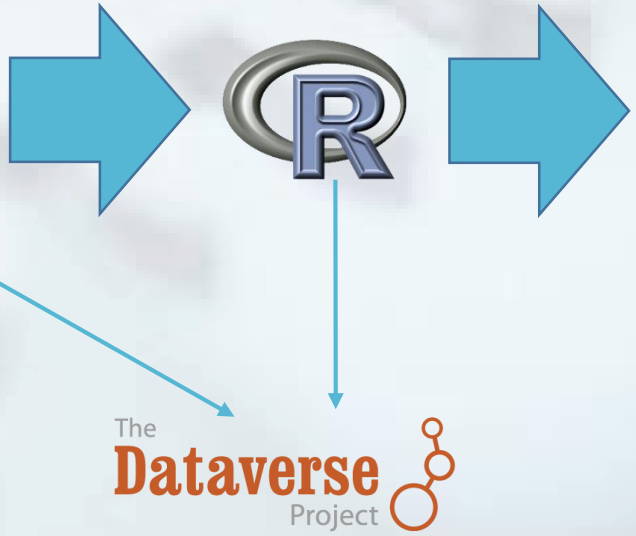


Table 2. Treatment Effects (Experiment 1)

	Perceived Personal Influence	Expected Reciprocity	Believe Global Warming is Happening	Believe Global Warming is Human Induced	Support Cap on Carbon Emissions	Willingness to Take Personal Action
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.38**	0.23**	0.37**	0.25**	0.28**	0.20**
	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
Pro Descriptive (D)	-0.04	0.02	0.05	-0.03	-0.06	-0.02
	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)	(0.09)
Pro Descriptive + Injunctive (D+I)	-0.08	0.07	0.09	0.06	-0.02	-0.05
	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)	(0.08)
Con D	-0.18**	-0.06	-0.06	-0.05	-0.16**	-0.17**
	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)	(0.09)
Con D+I	-0.10	0.11*	0.03	-0.06	-0.08	-0.06
	(0.08)	(0.07)	(0.09)	(0.09)	(0.09)	(0.09)
PartyID	0.21**	0.21**	0.28**	0.31**	0.21**	0.13*
	(0.08)	(0.08)	(0.11)	(0.09)	(0.10)	(0.09)
PartyID * Pro D	-0.05	-0.04	-0.10	-0.04	-0.05	0.04
	(0.11)	(0.11)	(0.14)	(0.13)	(0.13)	(0.13)
PartyID * Pro D+I	-0.10	-0.12	-0.06	-0.13	-0.14	0.04
	(0.11)	(0.11)	(0.13)	(0.12)	(0.13)	(0.11)
PartyID * Con D	-0.10	-0.19**	0.00	-0.10	0.03	0.11
	(0.11)	(0.11)	(0.14)	(0.13)	(0.13)	(0.13)
PartyID * Con D+I	-0.13	-0.21**	0.01	-0.15	0.01	0.04
	(0.11)	(0.10)	(0.13)	(0.12)	(0.13)	(0.11)
SER	.48	.47	.52	.53	.51	.54
n	612	612	612	611	612	612

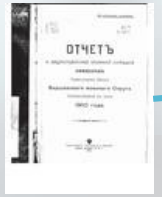
*p<.10, **p<.05, one-tailed test
Note: D = Descriptive, I = Injunctive. Cell entries are unstandardized linear regression coefficients with bootstrapped standard errors in parentheses. Baseline is the control condition. Party identification is coded with seven categories from Republican (-1) to Independent (0) to Democrat (+1). SER is the Standard Error of the Regression (Beck, 2010).

While we expected norms promoting action to influence both beliefs (Hypothesis 1a) and behavioral intentions (Hypothesis 2a), we find little support our hypotheses. Table 2 reports regression results with indicators for each treatment condition (relative to a control condition), including a control for party identification. Looking at the rows for the Pro Norm conditions, we do not see any effect on any outcome. Indeed for several outcomes (perceived personal influence and support for carbon emissions caps), we find effects in the opposite direction of our

Toby Bolsen, Thomas J. Leeper, and Matthew Shapiro. 2014. "Doing What Others Do: Norms, Science, and Collective Action on Global Warming." *American Politics Research* 42(1): 65–89.

QDR

Qualitative Research: Granular Data



Analysis

Analysis

Analysis

Analysis

DANILOV'S DEFENSIVE PLAN OF 1910

The Russian war plan of 1910 was more defensive and more oriented toward Germany than any other Russian plan made between 1880 and 1914. General Danilov, the principal draftsman of the plan, allocated 53 divisions to the German front and 19 to the Austrian front. His deployments abandoned Warsaw, the forward fortresses, the Narew River barrier, and indeed most of Poland. Forces deployed against Germany and Austria were to advance after completing their concentration, if circumstances permitted. Rail and road improvements were, however, planned with strictly defensive operations in mind. The abandoned forward zones were to be stripped of all supplies.[1]

Danilov gave two reasons for concentrating forces far from the frontier and primarily against Germany. One was Russian military weakness after the Russo-Japanese War, including the continuing lag in the speed of Russian deployments, and the other, the possibility of a large-scale German attack against Russia and the uncertainty of adequate French support.

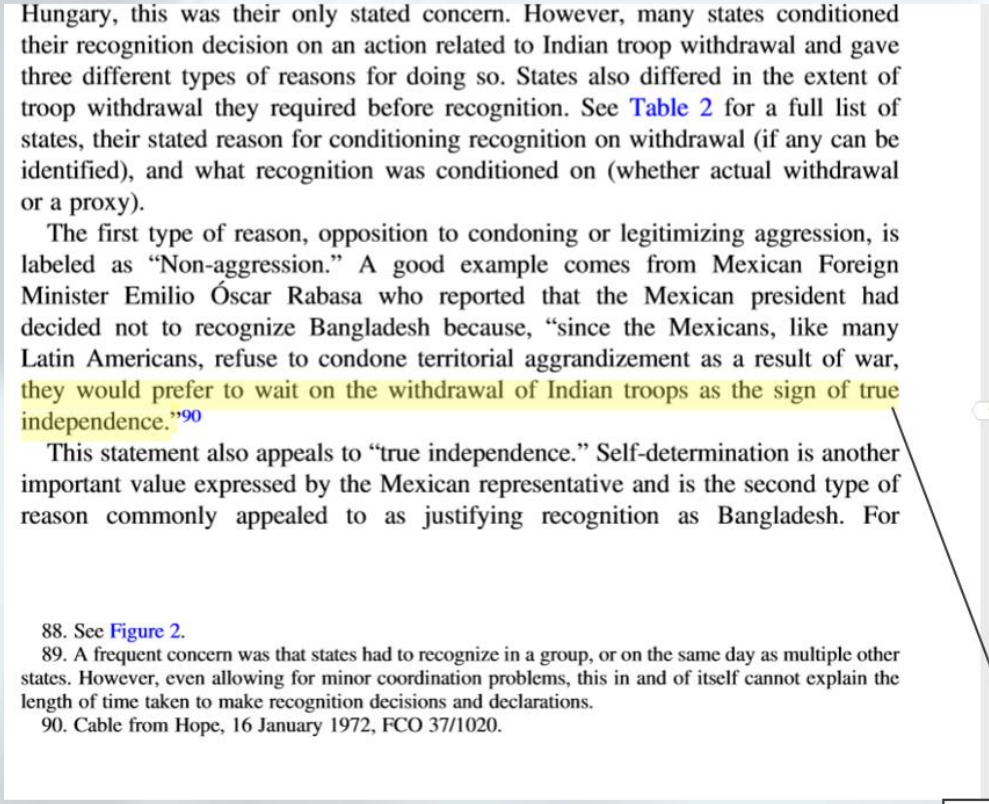
The Russo-Japanese War and its attendant political and economic difficulties had a ruinous effect on the Russian army. Matériel stocks were drawn down and not replaced because of Russia's financial crisis, which lasted almost until 1910. Official targets for the stockpiling of equipment and supplies would not even be approached until 1913 or 1914.[2] Between 1906 and 1908 as many as one-third of Russia's soldiers were used for internal security duties, which sapped morale and interfered with training.[3] Moreover, Russia's defeats in Manchuria had called into question the fitness of existing military doctrines, organizational forms, and commanders. The first few years after the war were spent debating the doctrinal and institutional implications of the defeat. Significant reform and reorganization were delayed until 1908 and 1910, when funding was increased and policymaking powers were concentrated in the hands of the war minister.[4]

Russian military figures warned at several times during this period that the army's devastated condition made the use of force dangerous [end of 166] against even Turkey, not to mention Austria or Germany.[5] Such was the state of disorganization even as late as 1909 and 1910 that it would have been "utterly impossible to make war," according to General A. Brusilov, who commanded the Russian Eighth Army. In Brusilov's high-priority corps in the Warsaw military district, supplies were barely sufficient even for peacetime requirements, and the transport service was in chaos.[6]

QDR

ATI at a glance

Any digitally published manuscript can be annotated using ATI (here: an article in *International Organization* published by Cambridge University Press)



Hungary, this was their only stated concern. However, many states conditioned their recognition decision on an action related to Indian troop withdrawal and gave three different types of reasons for doing so. States also differed in the extent of troop withdrawal they required before recognition. See Table 2 for a full list of states, their stated reason for conditioning recognition on withdrawal (if any can be identified), and what recognition was conditioned on (whether actual withdrawal or a proxy).

The first type of reason, opposition to condoning or legitimizing aggression, is labeled as "Non-aggression." A good example comes from Mexican Foreign Minister Emilio Óscar Rabasa who reported that the Mexican president had decided not to recognize Bangladesh because, "since the Mexicans, like many Latin Americans, refuse to condone territorial aggrandizement as a result of war, they would prefer to wait on the withdrawal of Indian troops as the sign of true independence."[90]

This statement also appeals to "true independence." Self-determination is another important value expressed by the Mexican representative and is the second type of reason commonly appealed to as justifying recognition as Bangladesh. For

88. See Figure 2.
89. A frequent concern was that states had to recognize in a group, or on the same day as multiple other states. However, even allowing for minor coordination problems, this in and of itself cannot explain the length of time taken to make recognition decisions and declarations.
90. Cable from Hope, 16 January 1972, FCO 37/1020.

ATI Annotation: Displayed alongside article. Created by author, curated by QDR, hosted and served by Hypothesis, displayed on publisher's web site

Elements of an ATI annotation: One or more of the following:

- Analytic Note
- Source Excerpt
- Source Excerpt Translation
- Link to Data Source

Link to data source housed in QDR

Any passage in the text or in notes of a manuscript can be annotated using ATI

QDR

Two ATI examples

- O'Mahoney, Joseph. 2017. "Making the Real: Rhetorical Adduction and the Bangladesh Liberation War." *International Organization* 71 (2): 317–48.
- Gans-Morse, Jordan. 2017. "Demand for Law and the Security of Property Rights: The Case of Post-Soviet Russia." *American Political Science Review* 111 (2): 338–59.

QDR

ATI Initiative

- Includes the ATI Pilot Working Group
- Authors from various health and social science disciplines who employ multiple analytic approaches, use diverse types of qualitative data, and study different world regions and temporal periods.
- Reviewers, charged with evaluating authors' use of ATI according to a pre-determined set of criteria.

The ATI Pilot working group by the numbers:

37 Scholars (20 authors, 17 reviewers)

17 Articles

7 disciplines (Anthropology, Health, Public Health, Linguistics, Political Science, Psychology, Sociology)

379 Annotations

159 Data source files

QDR

Some of the issues the ATI working group is trying to address

- Does ATI help make qualitative research more transparent?
- For what articles, methods, disciplines does it work particularly well or particularly poorly?
- Does ATI have benefits/uses other than transparency?
- Could the structure of ATI be improved?
- How well is the technology we use suited for the purpose and how could it be improved?
- How can the documentation be improved?

Learning about ATI from the working group

- Authors
  - Were asked to keep logbook while annotating
  - Were asked to complete questionnaire
- Reviewers
  - Were asked to read/assess/evaluate original (i.e. pre-annotation) article, and compare with annotated version.
- Workshop
  - Thematic panels plus group discussion of all 17 annotated articles

QDR

Logbook questions for the annotating authors

We asked authors to keep a logbook while they annotated, answering questions that included:

1. Why are you choosing particular passages, citations, or footnotes of the manuscript for annotation?
    a. Are you annotating particular types of claims (e.g., descriptive, or causal, or controversial)?
    b. Do you find you're drawn to annotating claims in some sections of your article more than others?
2. Where you are including analytic notes, what is their main function (e.g., do they offer additional context, reflections, or interpretation; evaluate sources; discuss how sources were produced or analyzed; or elucidate links between evidence and claims)?

QDR

Logbook questions for the annotating authors

3. Why are you choosing to link underlying data sources (i.e., deposit a file with QDR and link to it in the annotation) to particular passages, citations, or footnotes of the manuscript? Where you choose not to link in underlying data sources, what are some of the reasons?
4. How did the way you organized your notes and files impact your ability to annotate your article?
5. Are there any aspects of our instructions for creating an ATI Data Supplement that are unclear, or any processes that can be simplified?

QDR

Questionnaire Results

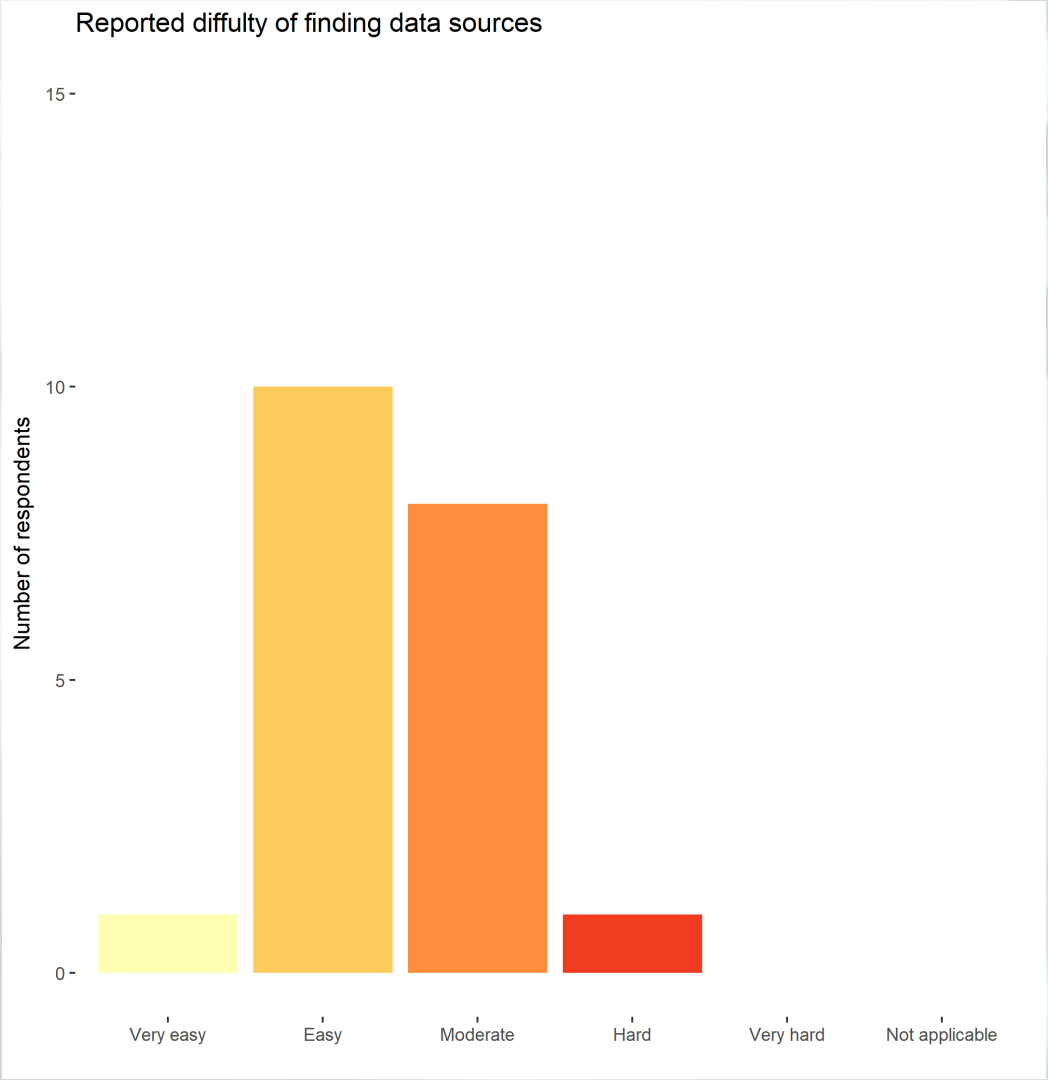
QDR

Tracking down data sources

- How hard/easy was it to find data sources on which claims were based on computer/in records?
- Based on retrospective annotation of a previously (albeit, recently) published article. May look different for contemporaneous annotation.

QDR

Difficulty Finding Data Source



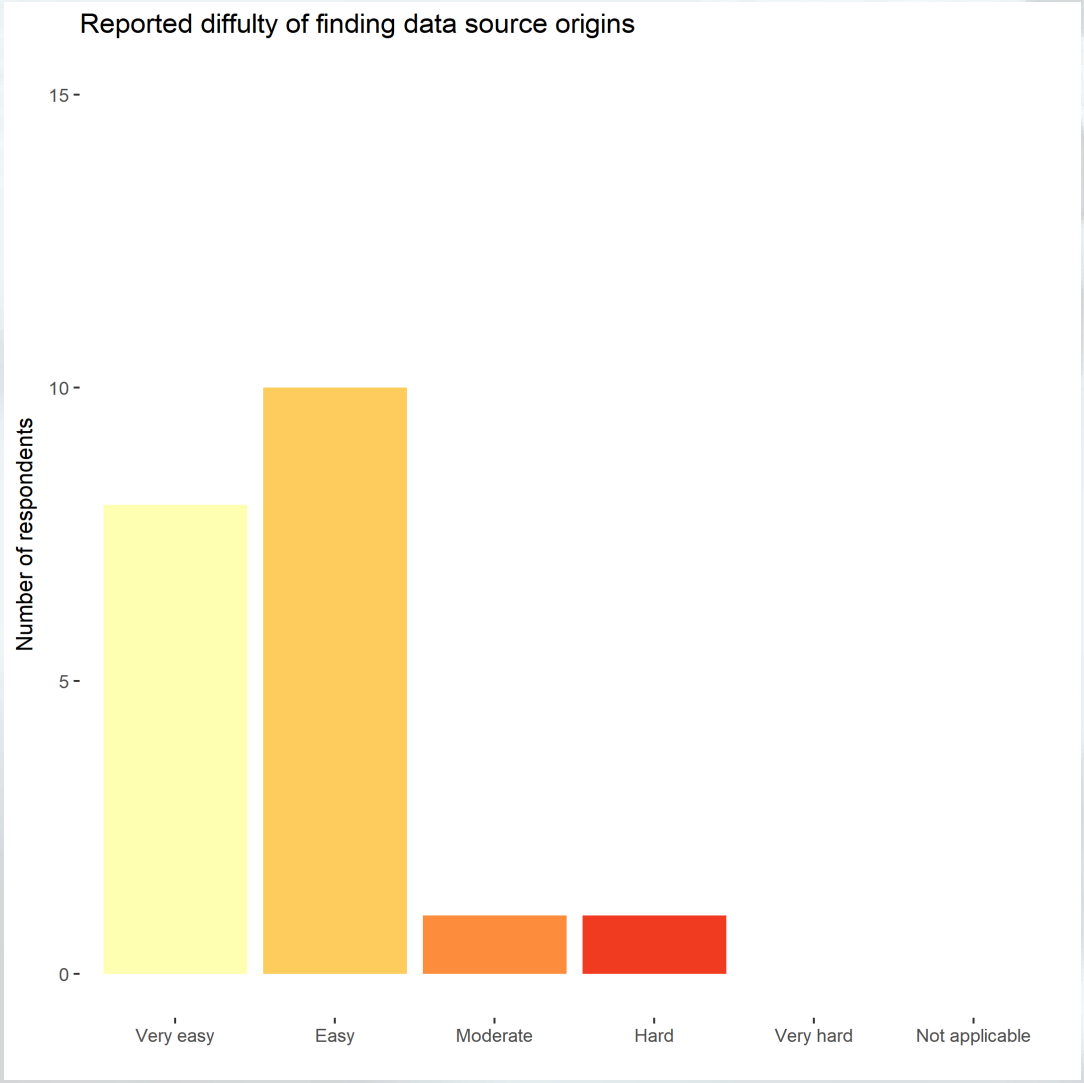
QDR

Tracking down *origin* of data sources

- How hard/easy was it to find the origin of the data sources on which claims were based?
- Again, based on retrospective annotation of a previously (albeit, recently) published article. May look different for contemporaneous annotation.

QDR

Difficulty Finding Data Source Origins



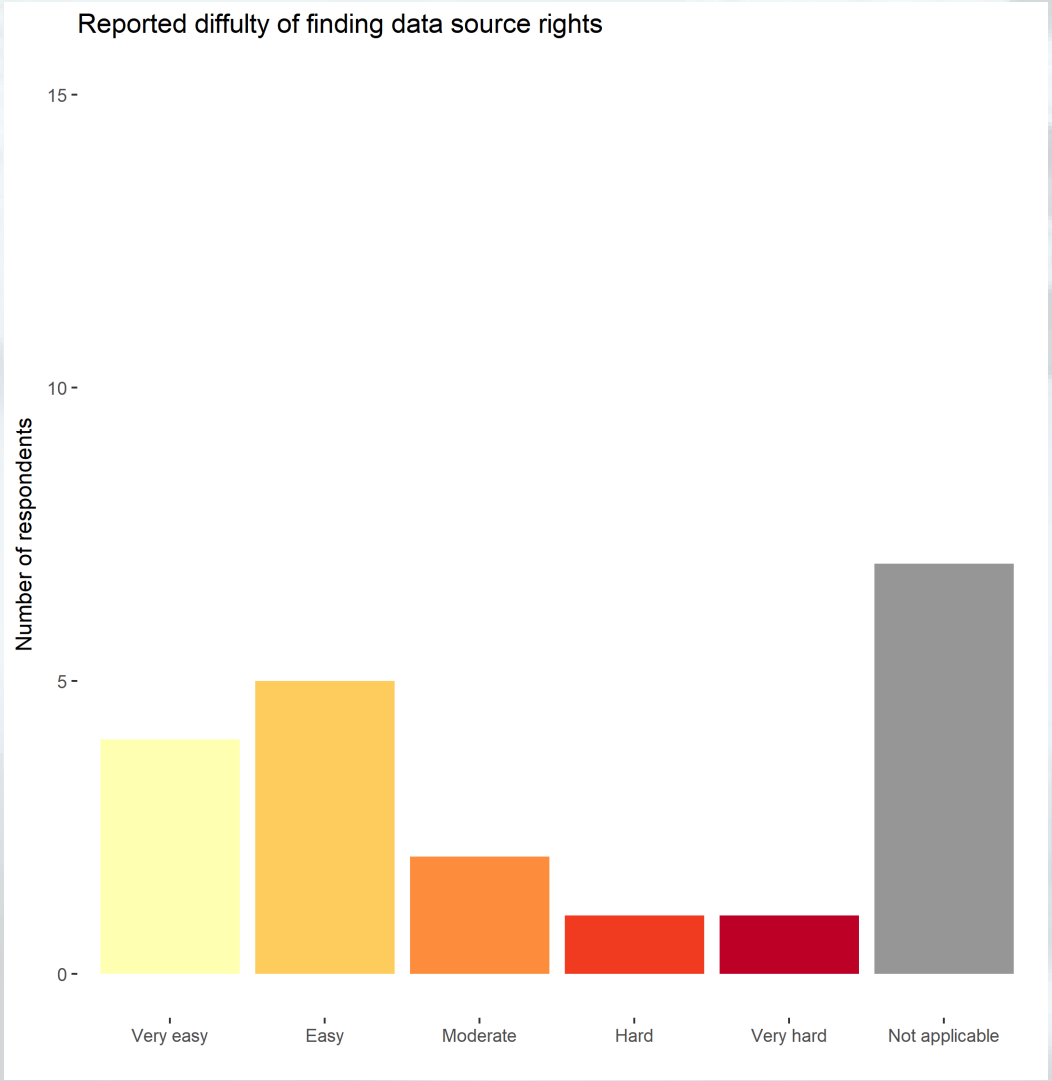
QDR

Determining source rights (copyright and/or license) of data sources:

- How hard/easy was it to determine whether the data sources were subject to proprietary constraint?

QDR

Difficulty Finding Source Rights



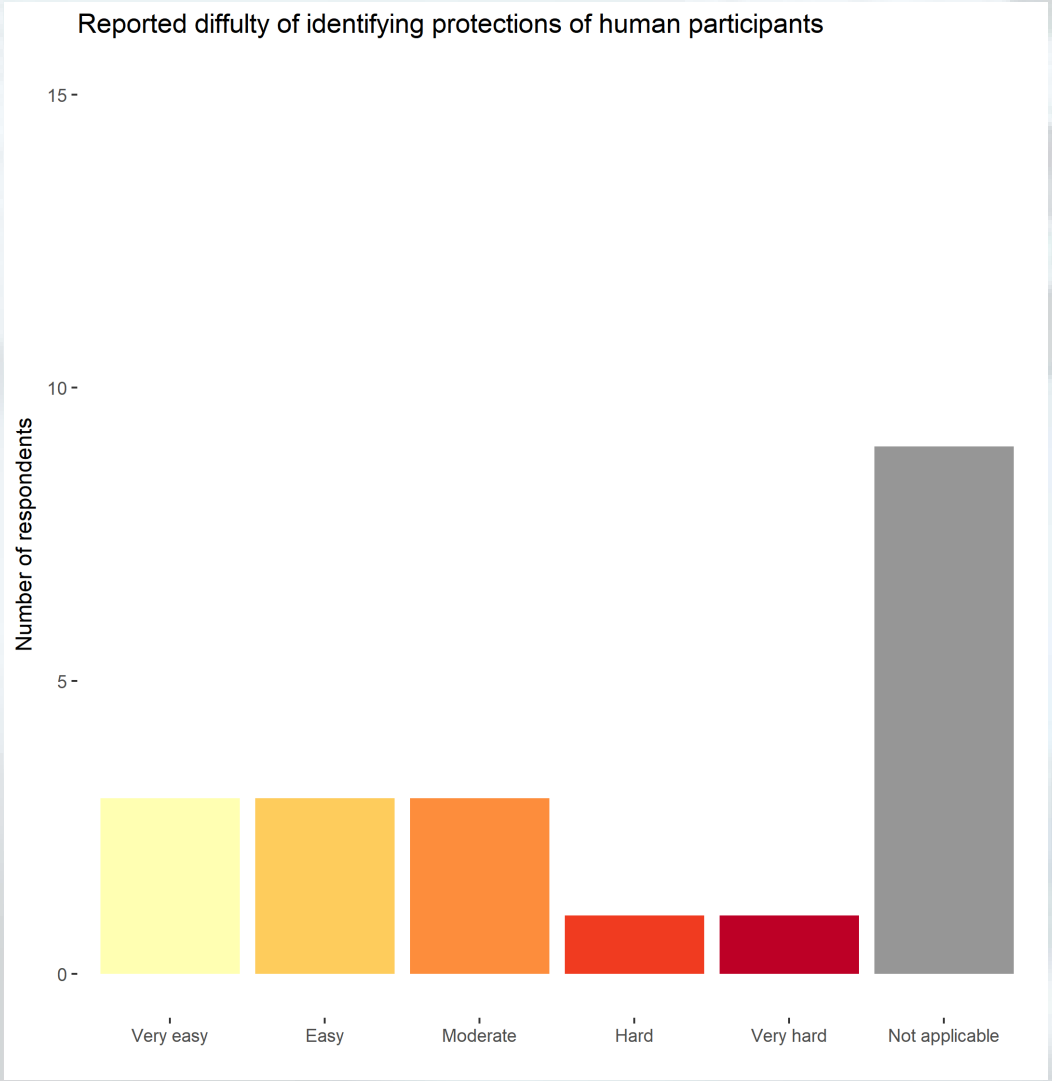
QDR

Determining whether impact of human subjects protections?

- How hard/easy was it to determine whether guarantees offered to human participants affected sharing data?

QDR

Difficulty Identifying Participant Protections



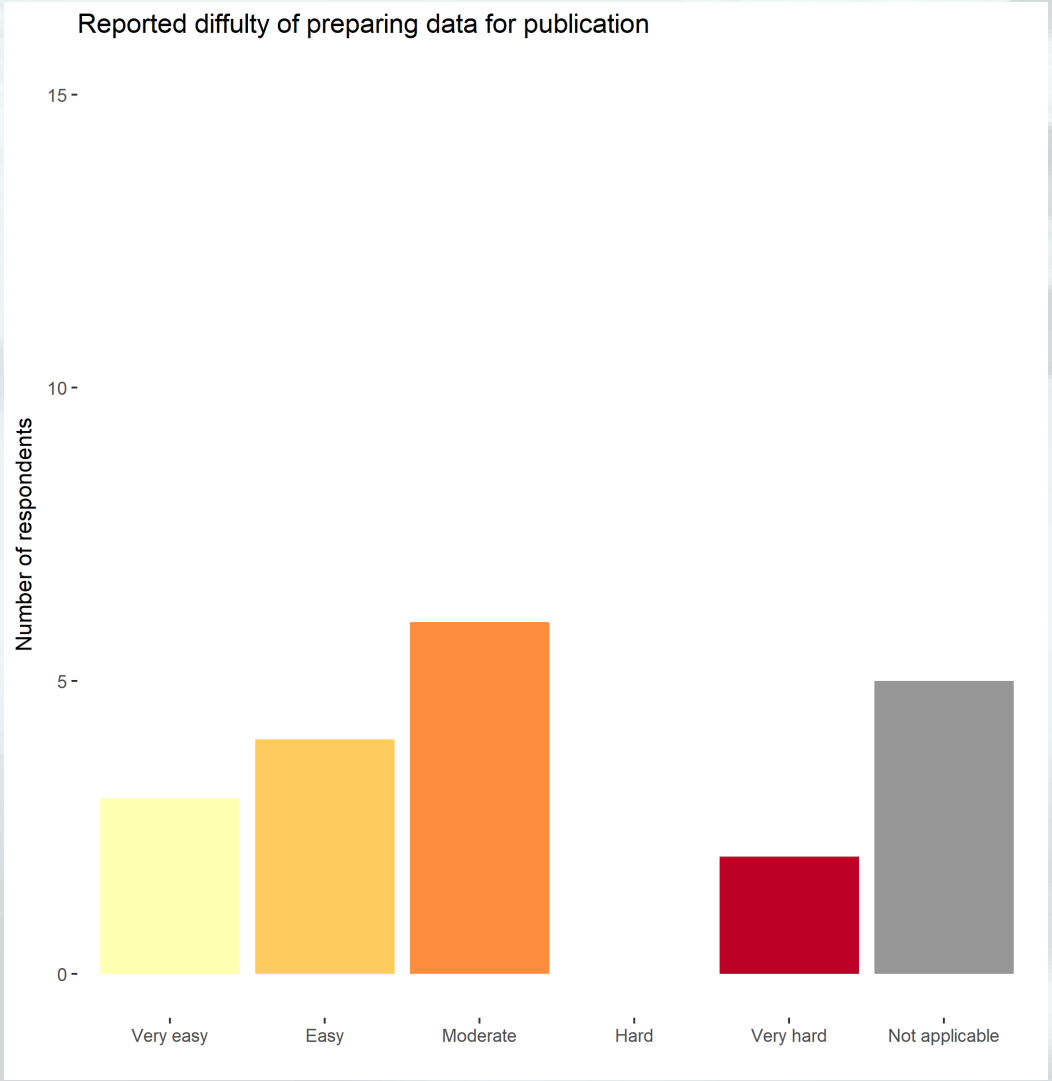
QDR

Preparing data for publication?

- How hard/easy was it to prepare data for archiving (e.g. converting paper to digital, generating a pdf, scanning a document, etc.)?

QDR

Difficulty preparing data for publication



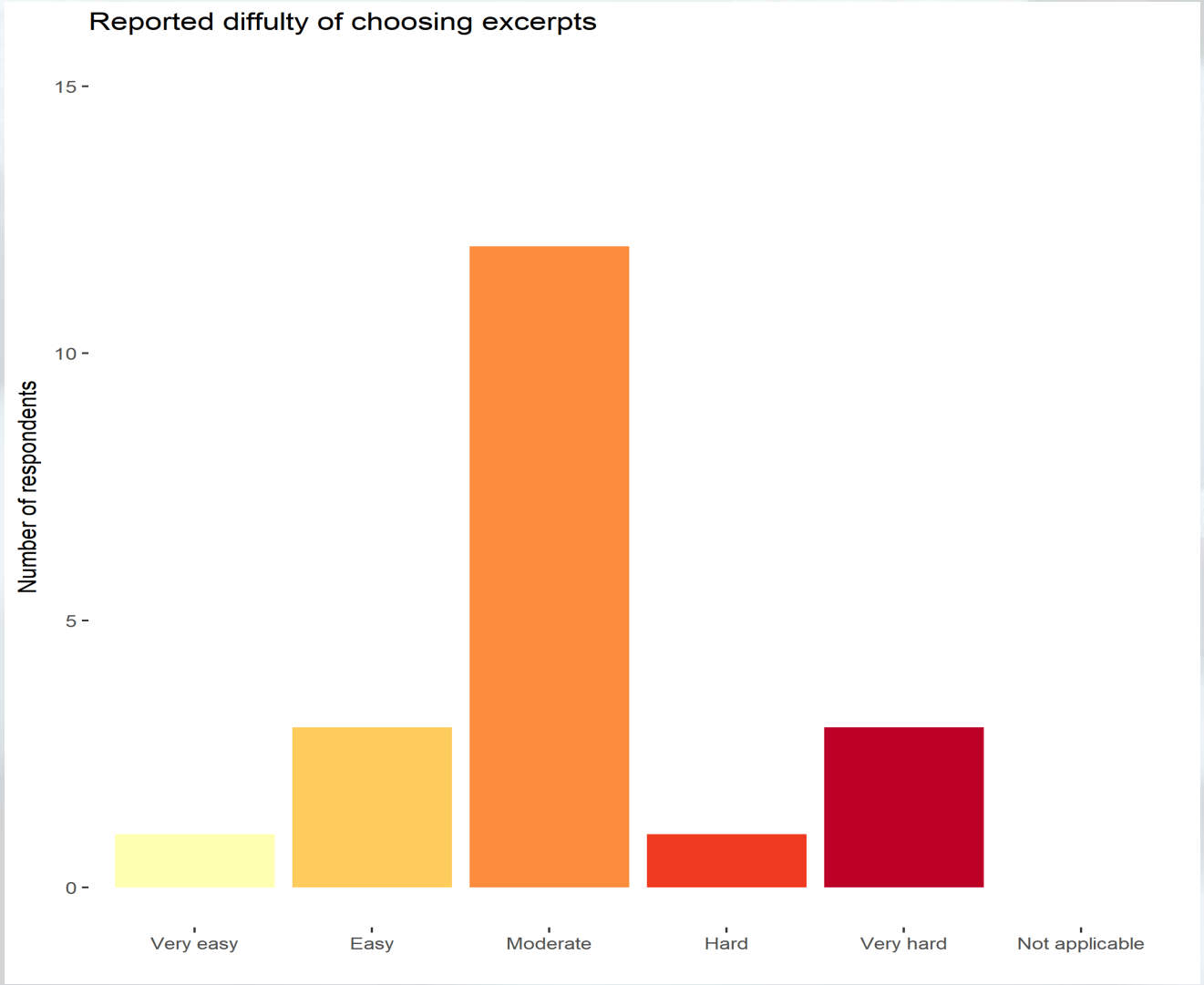
QDR

Choosing what to excerpt in annotations?

- How hard/easy was it to choose relevant excerpts of data sources to include in annotations?

QDR

Difficulty of Choosing Excerpts



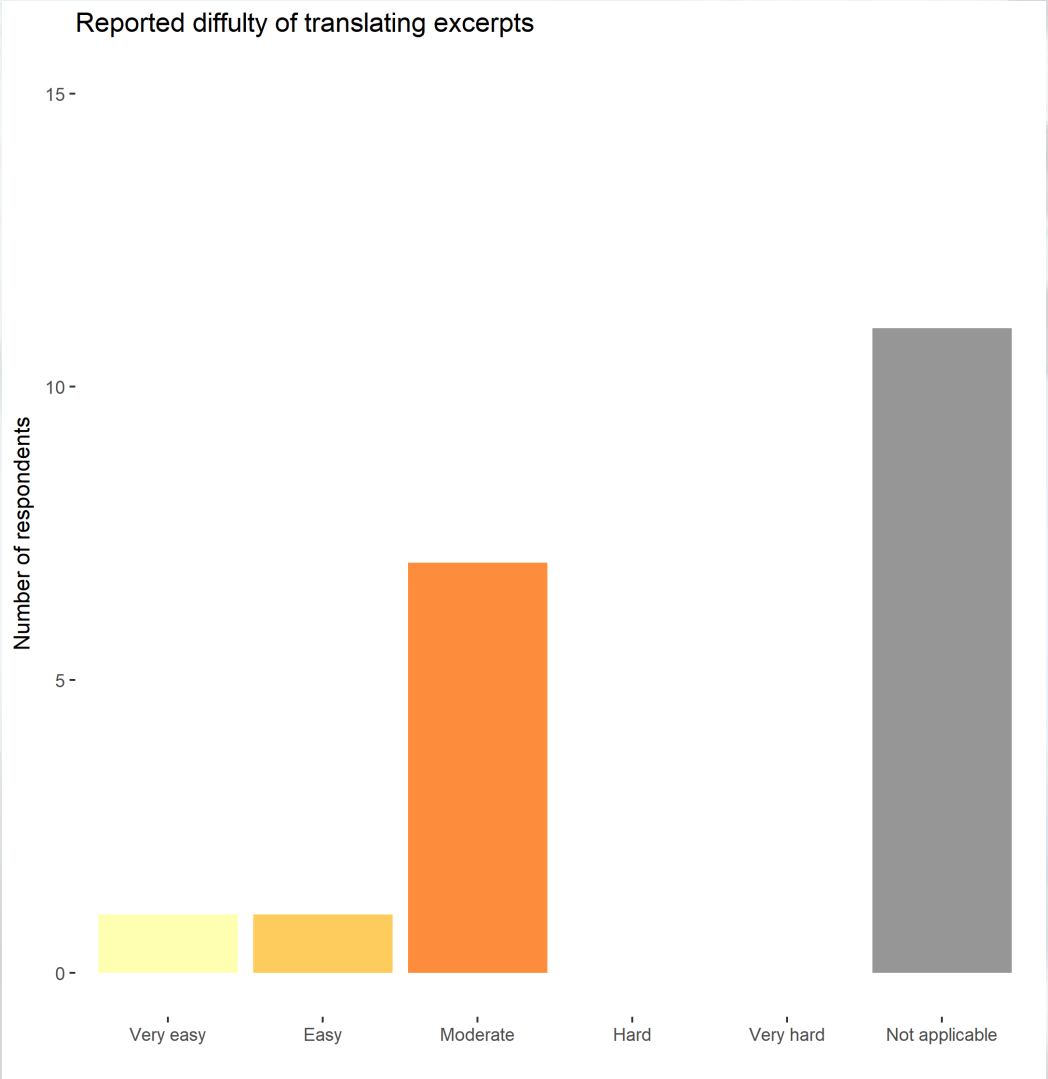
QDR

Translating excerpts?

- How hard/easy was it to translate excerpts for the annotations?
- Only relevant for non-English sources, and optional.

QDR

Difficulty translating excerpts



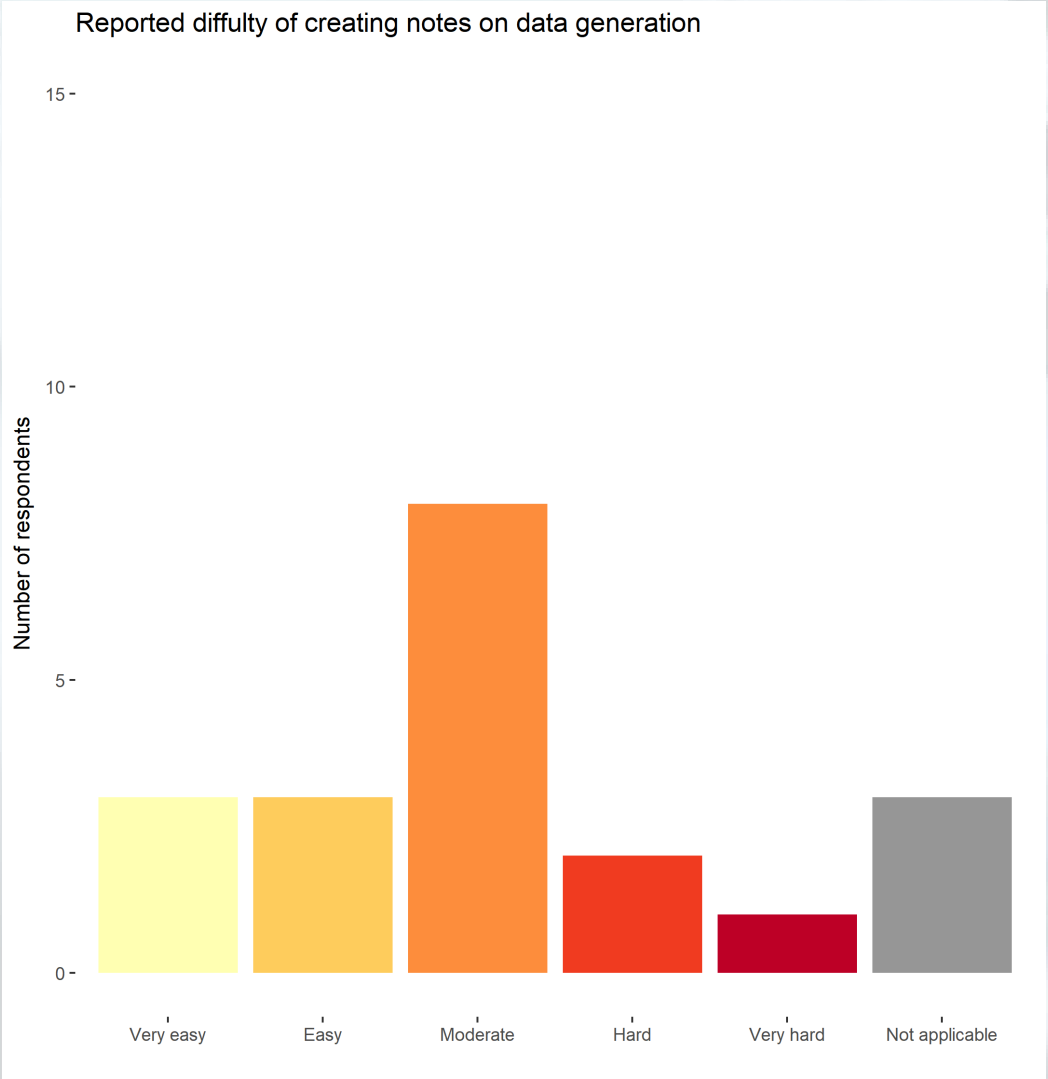
QDR

Creating analytic notes about data generation?

- How hard/easy was it to create notes about how data sources were generated?

QDR

Difficulty writing notes on data generation



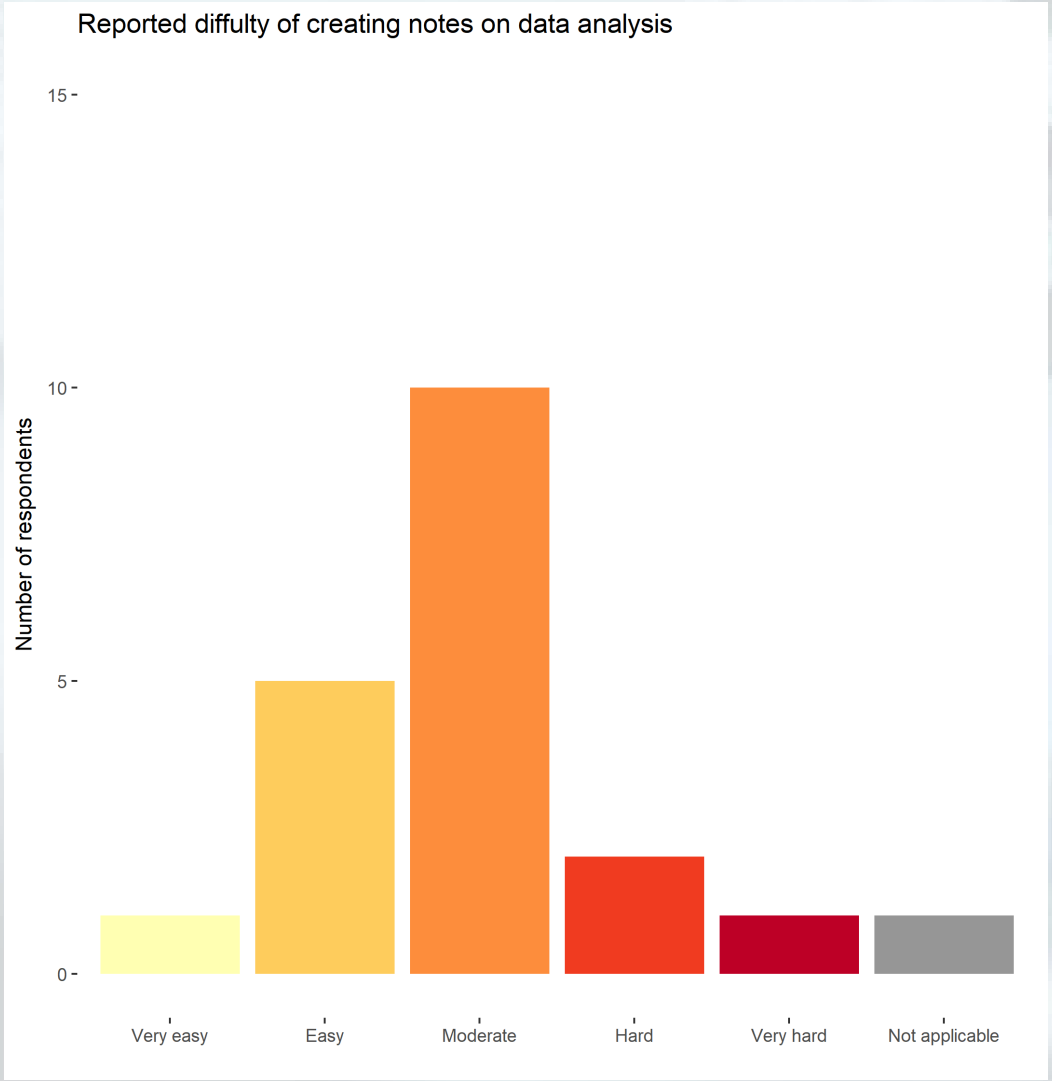
QDR

Creating analytic notes about data analysis?

- How hard/easy was it to create notes about how data sources were analyzed?

QDR

Difficulty writing notes on data analysis



QDR

Distribution of answers across articles

	A	N	Q2	J	K	H2	G	B	O	P	F	Q1	L1	L2	C	E	H1	D	I	M
1. Finding data sources	2	2	3	3	1	3	2	4	3	3	2	2	2	2	2	3	2	3	2	3
2. Origin of data sources	4	1	2	1	1	2	3	2	2	1	1	2	2	1	1	2	2	2	1	2
3. Source rights	2	3	na	1	na	5	4	2	na	1	2	na	na	1	na	2	3	2	na	1
4. Human participants af	5	na	na	3	na	1	na	2	4	2	na	na	2	1	na	na	na	3	1	3
5. Preparing data	5	1	5	1	na	1	3	3	2	3	2	na	2	3	na	3	3	na	na	2
6. Choosing excerpts	3	5	3	5	5	3	3	3	3	4	3	3	2	3	2	3	3	2	1	3
7. Translating excerpts	1	na	3	2	na	3	na	3	3	3	na	3	na	na	na	na	3	na	na	na
8. Data Notes	4	5	4	1	3	1	na	3	3	na	1	3	3		3	3	2	2	2	3
9. Analysis Notes	4	5	3	2	3	1	4	3	3	3	3	na	2	2	3	3	2	3	2	3

QDR

After the workshop

- ATI projects from February workshop on articles published by Cambridge University Press are live.
- Second group of ATI projects from articles from non-CUP journals will launch soon (waiting on some minor teach tweaks from Hypothesis).
- ATI Challenge launched, meritocratic selection of ATI projects for second (November) workshop. Difference from first workshop, annotate contemporaneously and not retrospectively.

QDR