Qualitative Data Analysis

Andreas Kaufmann and Dirk Riehle, Univ. Erlangen

NYT B03

Fundamentals

Overview 1 / 2

- What is qualitative data?
  - Any information that can not be reduced to a numerical representation
    - For example: Data related to concepts, opinions, behavior and social activities
- Methods of data collection
  - Interviews (in-depth / semi-structured / structured)
  - Focus groups
  - Observation
  - Immersion
  - Literature search

Overview 2 / 2

- What is qualitative data analysis?
  - A form of inquiry to answer questions about what, why and how people think, act or experience a phenomenon
  - Interpretative analysis
- Where are the analysis results used?
  - Theory building research, very heavily in sociology and market research
- How is the data analyzed?
  - Coding (more on that later)

Qualitative vs. Quantitative Research

	Qualitative	Quantitative
Purpose	Gaining insight into the phenomenon of interest	Measuring magnitude (How common is the phenomenon? Are there shared statistical significant characteristics among the participants?)
Format	Open discussions, interviews	Structured categories of data
Data	Explanatory data from small sample	Data from representative sample
Analysis	Hypothesis generating: form of outcome is not pre-defined a-priori	Hypothesis testing
Process	Iterative	Linear
Result	Common concepts and ideas (the theory); individual responses	Numerical aggregation of clustered responses; accepted or rejected hypotheses
Sampling	Theoretical	Statistical

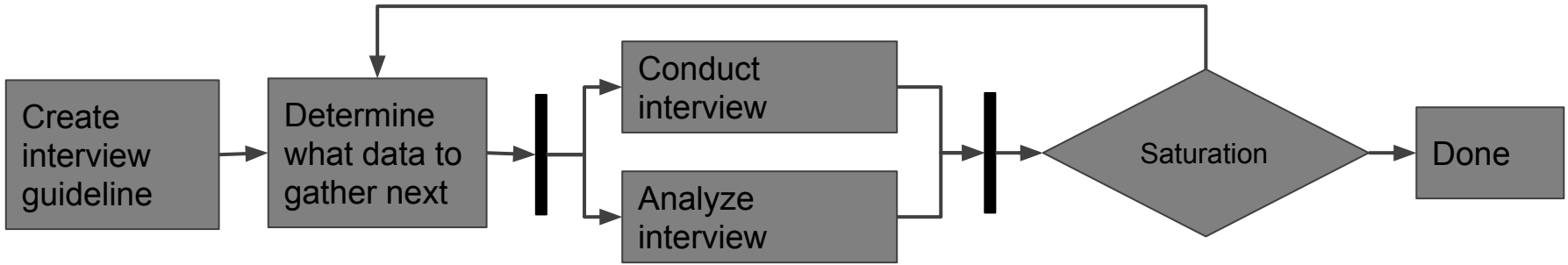
uni1.de/nyt

Research Framework

- Can be performed in the context of a qualitative research framework
  - Grounded Theory (GT), Dimensional Analysis, etc.
- Can be performed independent of a qualitative research framework
  - Technique useful for SLRs, Qualitative Surveys, Case Studies, etc.
  - Some methods, such as thematic analysis, are explicitly independent of any frameworks
- Most appropriate for theory-building research

Basic Principles

- Iterative Process
  - Data gathering and analysis in parallel
- Stopping criterion: Saturation
  - Data should be collected as long as the subsequent analysis yields no changes to the resulting theory any more.



Sampling

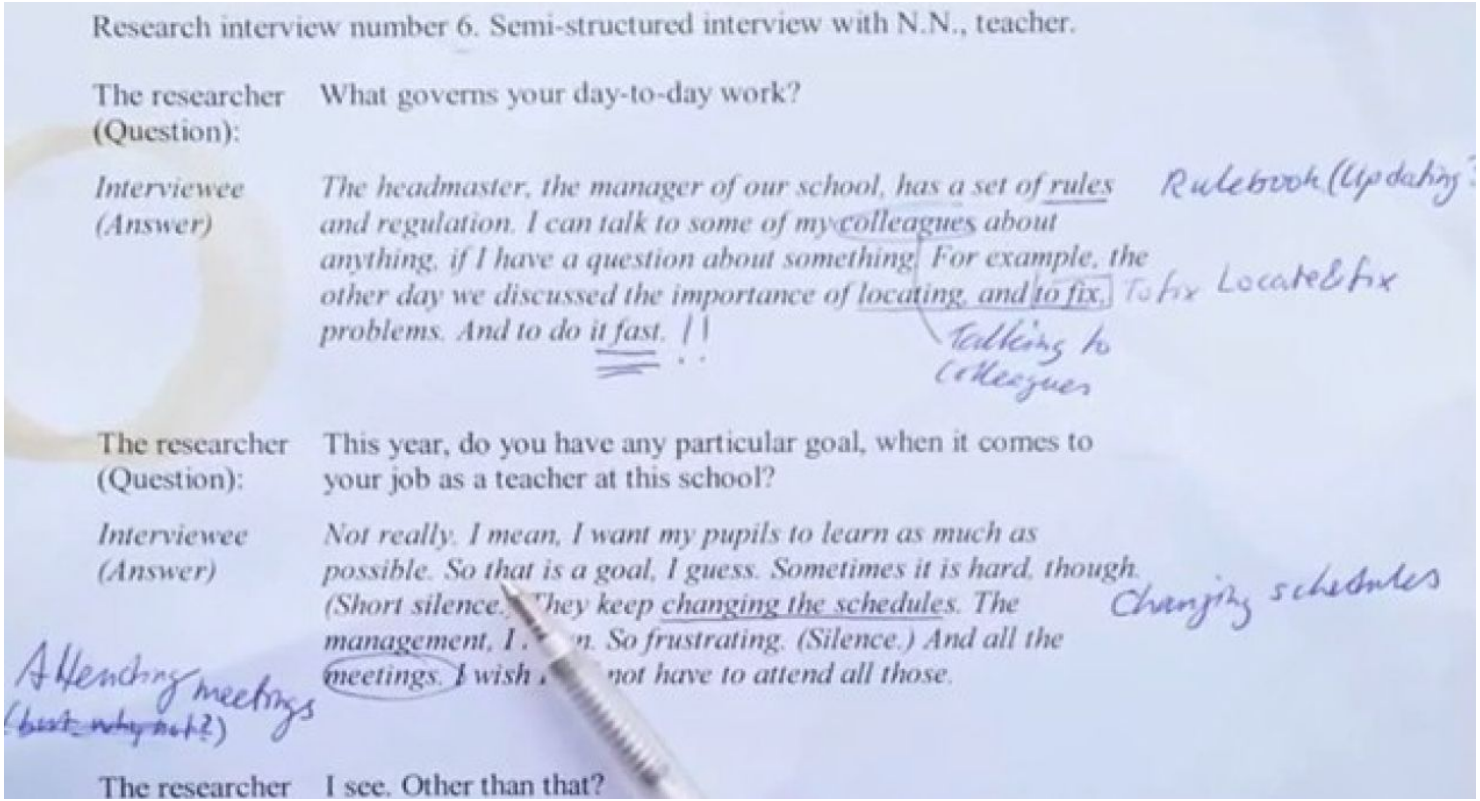
- Theoretical sampling
  - Identify the gaps in the theory and potential conflicts after each iteration and determine appropriate data to gather only for the next iteration
  - By many considered the gold standard of qualitative research
- With defined sampling model
  - Identify the dimensions of relevance through an a-priori literature review
  - Edge-case, Typical-case, Polar-case, etc.
  - Consider stratified random sampling

Coding

Coding

- Coding is a form of data analysis in which theoretical constructs emerge from the data in the form of codes,
- Codes are hierarchically structured
- Segments of data are assigned to the codes representing these constructs
- The codes and their hypothesized relationships are evidenced in the data
- The sum of all codes and their relationships form a theory about the phenomenon under investigation

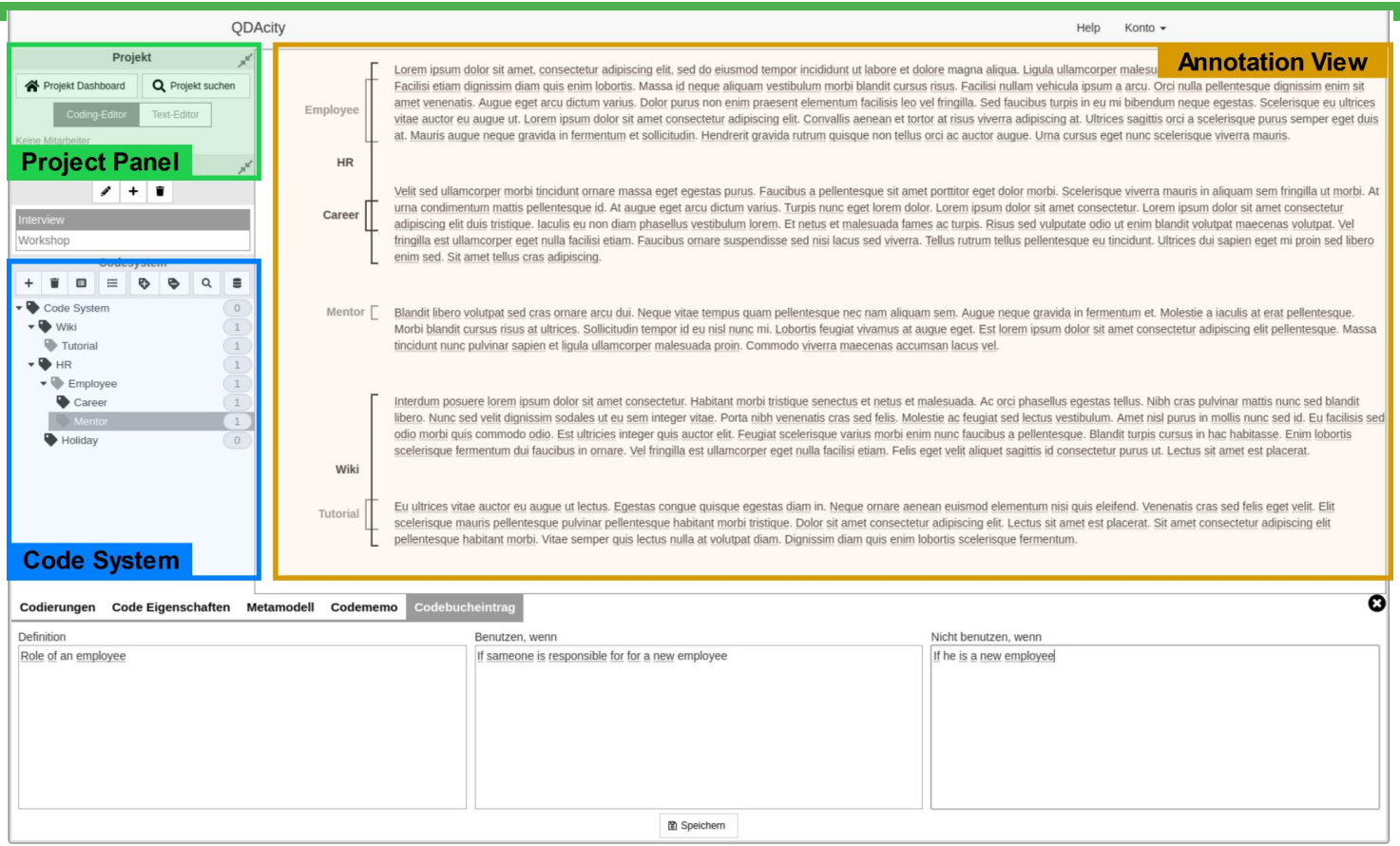
Coding, then



Research interview number 6. Semi-structured interview with N.N., teacher.

The researcher (Question):	What governs your day-to-day work?
Interviewee (Answer)	*The headmaster, the manager of our school, has a set of rules and regulation. I can talk to some of my colleagues about anything, if I have a question about something. For example, the other day we discussed the importance of locating, and to fix, problems. And to do it fast.*
The researcher (Question):	This year, do you have any particular goal, when it comes to your job as a teacher at this school?
Interviewee (Answer)	*Not really. I mean, I want my pupils to learn as much as possible. So that is a goal, I guess. Sometimes it is hard, though. (Short silence.) They keep changing the schedules. The management, I [illegible]n. So frustrating. (Silence.) And all the meetings. I wish I [illegible] not have to attend all those.*
The researcher	I see. Other than that?

Coding, now



uni1.de/nyt

Coding

- Can be performed in three distinct steps
  - Open coding
  - Axial coding
  - Selective Coding
- Other types of process-phases exist (compare thematic analysis)

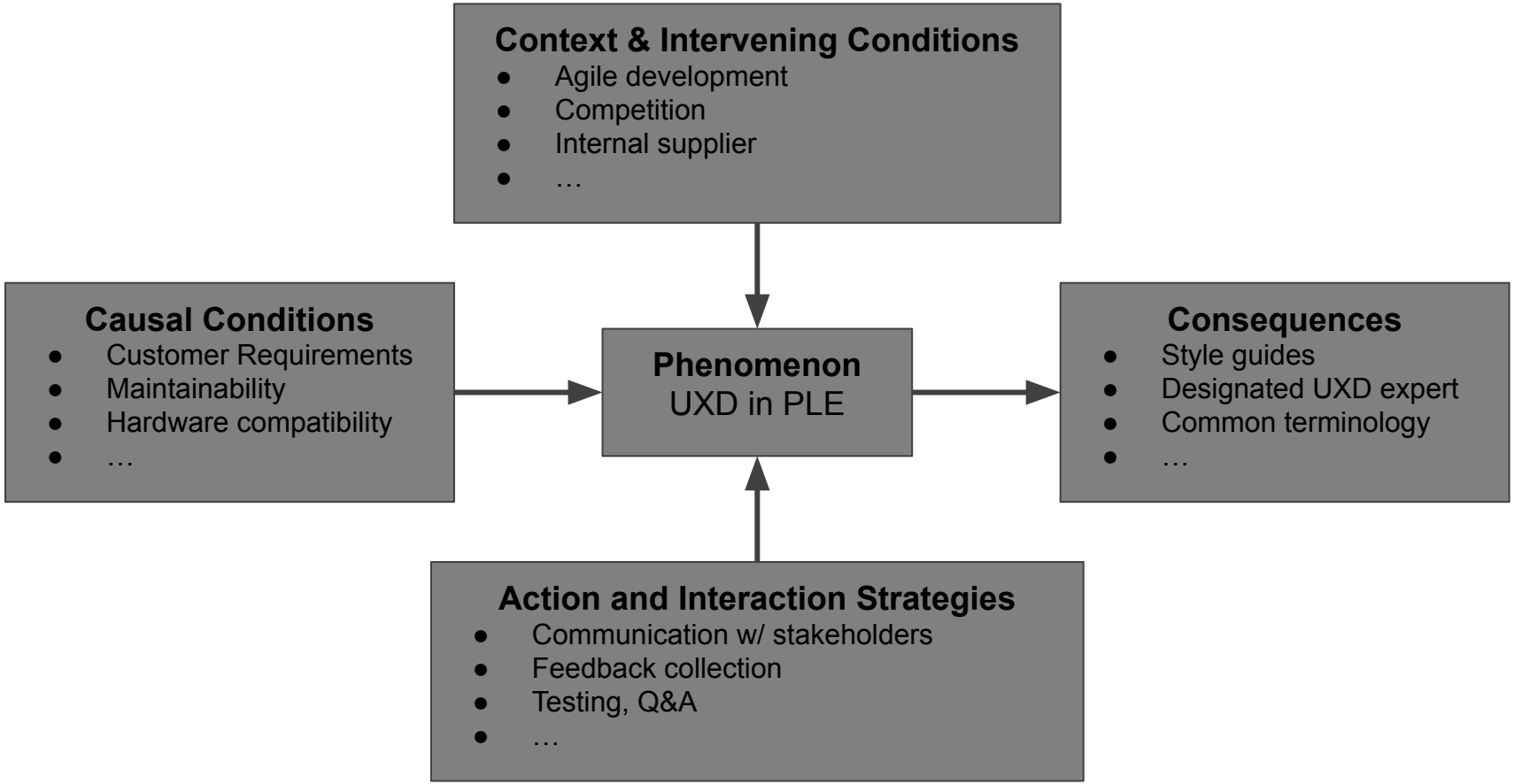
Open Coding

- Identify key concepts
  - Identify similar terms, synonyms
  - Determine concept codes
  - May be terms, actions, relationships
  - Codify using symbol, term, or desc.
- Annotate text with concept codes
  - Annotate using concept codes
  - Use codes consistently
- Memo writing!

Axial Coding

- Identify relationships between codes
  - Equivalence class of atomic concepts
- Category hierarchy follows
  - Should form a single-rooted hierarchy
  - Mutually exclusive, completely exhaustive
- Other types of relationships can be documented in memos

Axial Coding: Coding Paradigm



Selective Coding

- Giving categories context and making sense of them
- Writing memos, code book
- A theory (from data) is the root category of the category hierarchy
  - “Core category”
- At any given level of the hierarchy, term categories link by relationship categories
- At this stage, no need for validation, but theory should “make sense”

Rigor in Qualitative Data Analysis

Dimensions of Rigor

- Qualitative theory building research is often assessed in the naturalistic research paradigm
- Guba & Lincoln: Trustworthiness
  - Credibility
  - Transferability
  - Dependability
  - Confirmability

Dimensions of Rigor

- Showing rigor of execution
- Explicit evaluation
- You need both!

Peer debriefing

- Design and process is laid out to a researcher from outside the project
- The external researcher questions and critiques the approach to find possible flaws
- Session is documented with a peer debriefing protocol
- Should be done at multiple stages to validate the
  - Research design and fit for the research question
  - Recruiting and sampling
  - Data gathering
  - Data analysis

Intercoder Agreement

- Intercoder agreement is the extent to which two or more individuals agree on the coding of qualitative data
- Interrater reliability can be used to control the researcher's bias towards the studied phenomenon
- May be applied to different units of coding (e.g. sentence or paragraph level)
- Most common metrics for measuring: Kappa Statistic, Krippendorf's Alpha

Code Book

- A code book describes the code system
- It can include
  - A definition for each code
  - Instructions for when to use the code
  - Possible disambiguations towards other (similar) codes
  - An exemplary piece of data to be coded with each code

Measuring Saturation

- Track changes within your project
  - New codes
  - Changes to definitions
  - etc.
- Plot those changes against amount of new data gathered
- The code system should reach a steady state where there are little to no changes when new data is added

Member Checking

- Feedback from study participants / stakeholders
- Increases credibility and validity of the study
- During data gathering (interviewing)
  - Active / reflective listening
- After analysis
  - Mirror findings back to stakeholders / experts to confirm correct understanding of the phenomenon

Triangulation

- Data triangulation
  - The code system can ensure consistency between different types and sources of data
- Investigator triangulation
  - Intercoder agreement
  - Peer debriefing
- Theory triangulation
  - Peer debriefing
- Methodological triangulation
  - Different types of data gathering
  - Mixed methods (in addition to the qualitative part)
- Environmental triangulation
  - Awareness of the impact of different settings (in the field vs isolated)
- See Guion (2002) for a very short but good overview of the definitions

Coding Exercise

Overview

- Qualitatively code a group of documents following a provided codebook
  - Download codebook from StudOn
- Each student will be given 6 documents
  - Documents may not be shared with anyone
- Four exercises
  - QDA Exercise 1 – code documents “Rick” and “Morty”
  - QDA Exercise 2 – code documents “Margery” and “Homer”
  - QDA Exercise 3 – code documents “Kyle” and “Stan”
  - QDA Exercise 4 – 1-2 page conclusion to the research question of your choice based on your coding
- QDA Exercise 1-3 are performed on QDAcity. Submit your work on the “Exercise Dashboard”
- QDA Exercise 4 requires submission on StudOn
- For explanation how to join the course on qdacity see here

Learning Process

- Read the codebook very carefully
- Use QDAcity for coding
- Work individually
- We will evaluate your weekly work on Mondays after class
- Recode the already coded document in each new week to make your coding consistent
- Post your questions on the StudOn form for NYT (be specific about what you don't understand or find unclear)

Process on QDAcity

- In QDAcity the QDA exercise is just one exercise which you continue to work on
  - However, the deadlines for the specific documents to be coded in Exercise 1, 2 and 3 are still relevant
- After each week we will create an intercoder agreement report
  - This will show up on your exercise dashboard, where you also see the statistics of your project

Peer Feedback on QDAcity

- After the report is generated you have the **option** of giving other course participants feedback on their coding
- For you this might have the benefit of seeing how other people work and potentially be inspired or see things that you overlooked and might do differently.
- If you do request other participants' submission for peer feedback, do make an effort to give constructive feedback
    - Giving feedback is not required, but if you do do it, do it so it helps out your fellow students

Final Artifact

- For Method Exercise 4, submit your 1-2 page conclusion to StudOn
  - Submit as a PDF file
  - State which research question you are answering
    - You may chose the specific wording of the research question your conclusion should answer (using the data from the project)
    - If you don’t want to come up with your own specific question, you may write about answering the general research question of “What are the most important challenges for remote work, and how are they addressed in industry?”
  - Maximum of 2 pages

Grading

- Grading of coding (ME 1 - ME 3) based on intercoder agreement
  - Your work is compared to original coding
  - We examine accuracy – both precision and recall – so indiscriminate coding is not advised
  - Differences of a few words or punctuation are not considered
  - Grading is automated

- Sum of coding exercises calculated as follows:
  - ME 1 – 15%
  - ME 2 – 15%
  - ME 3 – 40%
  - ME 4 – 30%

Code Hierarchy

- Do not apply category codes (eg. Codes with other codes below it); only use codes with no codes beneath them in the hierarchy.
- In the following example, code A should not be used:
- Category A
  - Concept B
  - Concept D
  - Concept E

Question Prompts

- Do not code the interviewer's question, except when it is needed to supply context.
  - In this example, the question provides context and is coded if the answer is coded:
    - Interviewer: Have you ever seen a zebra?
    - Interviewee: Yes, in Tanzania.
  - In this example, the question is not necessary to understand and is not coded even if the answer is:
    - Interviewer: Have you encountered a monkey?
    - Interviewee: A monkey stole my glasses in Shimla.

Coding Granularity

- Coding can be done at the multi-paragraph, paragraph, sentence or phrase level.
  - These examples are all valid codings for an imaginary code 'cat' defined as all references to cats:
    - I like cats, but I also like dogs. My neighbors got a new cat last month, but it likes to sleep at my house.
    - Cats are the best when they come to to snuggle you when you're feeling down. I know people say cats are aloof, but mine always knew when I was sad and would come to cheer me up.

      I do find them annoying at times, for instance when they dig up the garden or yowl at night.
  - Note that examples in the codebook do not necessarily contain all adjacent text employing the code, only sufficient text for understanding the concept.

Positive and Negative

- Both positive and negative examples of a concept should be coded. (unless otherwise specified)
  - These are both examples of the code 'helpfulness':
    - You only need to ask him and he's happy to volunteer – making phone calls, sending out postcards – anything the organization needs, really.
    - They are not good for the group. Even when you ask them explicitly, they won't help you.
  - Positive or negative opinions/practices are also both coded:
    - Organization does or doesn't engage in a practice
    - Speaker does or doesn't approve of something

Descriptions and Examples

- Both descriptions and examples should be coded, unless the codebook advises otherwise.
  - In this example, 'style guide' is described:
    - Use this code for description of UXD style guide, its structure, related practices, benefits and use. Style guide mainly focuses on usability implementation, but can also have UXD design components.
  - In this example, 'style guide' is demonstrated by example:
    - We also define the style guide. If we are not able to implement everything with base components, then we give out the style guide [saying] in which way the client departments have to develop.

Research Questions

- Our research question: What are the most important challenges for remote work, and how are they addressed in industry?
- For the conclusion you can come up with your own question.

Thank you! Any questions?

dirk.riehle@fau.de – https://oss.cs.fau.de

dirk@riehle.org – https://dirkriehle.com – @dirkriehle

Legal Notices

License

- Licensed under the CC BY 4.0 International license

Copyright

- © 2012-2023 Dirk Riehle, some rights reserved
- © 2015-2023 Univ. Erlangen, some rights reserved

Co-authors

- 2015-2023 Andreas Kaufmann

Different Methods

	Corbin Straus	Thematic Analyss
	Open coding	(Open) coding
	selective coding	searching for themes
	axial coding	
	core categories	themes

Different Forms of Qualitative Data Analysis

Card sorting

Thematic analysis

Theoretical coding