Literature Mapping Study for Machine Learning Interpretability Techniques Tim Korjakow, Jesse Josua Benjamin,

Christoph Kinkeldey, Claudia Müller-Birn TR-B19-04 2019

### Abstract

Project IKON aims to explore potentials for transferring knowledge generated in research projects at a major Berlin research institution, the Museum für Naturkunde (MfN, natural history museum). Knowledge transfer concerns both the exchange of knowledge among the employees (researchers as well as communicators or management staff), and with the broader public. IKON as a research project coincides with a continuous effort by the research institution to open itself up in terms of its activities and stored knowledge. To understand the specific requirements of IKON, we conducted semi-structured interviews with 5 employees of varying positions (researchers, management staff, employees) at the research institution. Our questions were conceived to understand: (1) what unites individual perspectives of knowledge transfer; (2) by whom and through which means (i.e., actors and infrastructures) is knowledge transfer carried out; (3) where relations between actors and infrastructures in the current state of knowledge transfer need support or intervention. From our results, we infer high-level implications for the design of an application that can support employees of the MfN in (1) collaborating with each other and (2) conceptualize Knowledge Transfer Activities based on semantically related research projects.

*Keywords:* Qualitative Research, Semi-Structured Interviews, Human-Computer Interaction, Thematic Analysis.

### 1 Introduction

Project IKON<sup>1</sup> aims to explore potentials for transferring knowledge generated in research projects at a major Berlin research institution, the Museum für Naturkunde (MfN, natural history museum).<sup>2</sup> Knowledge transfer concerns both (1) the exchange of knowledge among the employees, which we have termed intra-organisational knowledge transfer [?], but (2) more widely extra-organisational with specific publics such as the education sector, citizen science, industry and politics. IKON as a research project coincides with a continuous effort by the research institution to open itself up in terms of its activities and stored knowledge. There are two major artefacts that are expected to be developed in project IKON. First, a Semantic MediaWiki<sup>3</sup> which stores structured data of research projects and knowledge transfer activities. Second, and fed by the Semantic MediaWiki, a data-driven visualization application that explicates potential knowledge transfer from research projects. This technical report concerns the second application,<sup>4</sup> and seeks to illustrate how we conducted qualitative research to

Ihttps://www.mi.fu-berlin.de/en/inf/groups/hcc/research/projects/ ikon/index.html, accessed 09/07/2019.

<sup>&</sup>lt;sup>2</sup>https://www.museumfuernaturkunde.berlin/en, accessed 09/07/2019.

<sup>&</sup>lt;sup>3</sup>https://www.semantic-mediawiki.org/wiki/Semantic\_MediaWiki, accessed 26/07/2019

<sup>&</sup>lt;sup>4</sup>Currently under development at https://github.com/FUB-HCC/IKON-projektor/tree/master,accessed 29/07/2019.

understand how we may visualize the complexities of knowledge transfer at the MfN in an actionable manner for its employees. To understand the specific requirements for the application, we chose to conduct semi-structured interviews with employees of the MfN. In this technical report, we present our study design, analysis, evaluation, discussion and perspectives on future work.

### 2 Study Design

Our study was conceived as a qualitative probe into the understanding of knowledge transfer at the institution from the perspective of employees of various organizational strata.

### 2.1 Development of Semi-Structured Interview Questions

Over a period of four months, we iteratively developed the catalogue of questions that formed the guideline for our semi-structured interview in group discussions. Our final set of questions (Figure 5, p. 12, in German) were conceived to study: (1) what characterizes or unites individual perspectives of knowledge transfer; (2) by whom and through which means (i.e., actors and infrastructures) knowledge transfer is carried out; (3) where relations between actors and infrastructures in the current state of knowledge transfer need support or intervention.

### 2.2 Finding Participants from the Institution

Recruiting interviewees proved complicated, as the MfN is preparing for a major evaluation in September of the year the study was conducted in, as well as the generally time-restricted everyday work of its employees. We partly relied on our project-partners at the institution to suggest suitable participants, and partly on our own contacts at the institution we had developed during our involvement in project IKON. We contacted potential participants individually, paying special attention to the need for a selection of interviewees that would be representative of the organization as a whole. In the end, we recruited 5 interviewees (3 identifying as female, none as non-binary), specifically; one full-time researcher, two staff employees working in a research project, one employee from middle management who functioned as the head of various research projects, and one employee directly associated with the general directorate.

### 2.3 Conducting the Interviews

We arranged appointments with our participants at the institution, with place and time chosen at their own leisure. The two staff employees as well as the researcher chose to book meeting rooms, whereas the middle-management employee as well as the general directorate associate chose to meet in their offices. At the start of the interview, each interviewee was given (1) a personal introduction about the interviewer, (2) a brief description of the research background of the interview, (3) an overview of the interview procedure and (4) a disclaimer on confidentiality and data protection (Figure 4, p.

11, in German). The interviews lasted approximately 1:15 hours on average. During the interview, we took notes and focused on suitable follow-up question which were interlaced with the pre-determined questions.

### 2.4 Data Preparation

We audio-recorded each interview, and had the files auto-transcribed using an online service.<sup>5</sup> Audio-recording and the fact that its automatic transcription included the temporary storage of data on the service provider's servers, were expressly included in the participant consent form (Figure 19, p. 26, in German). The auto-transcriptions were read and corrected for transcription errors, as well as colloquialisms and specific vocabulary used at the institution (e.g., abbreviations such as MfN for 'Museum für Naturkunde', or FB for 'Forschungsbereich', i.e. research department).

### 3 Data Analysis

In this section, we detail our analytical procedure. To bolster our research, we used Computer-Assisted Qualitative Data Analysis Software, namely atlas.ti.<sup>6</sup> We conducted a theoretical thematic analysis, meaning we focused on a specific area of interest with a set of high-level concepts already defined (e.g., Knowledge Transfer, Institution, Individual) while still allowing for emergent codes [?].

### 3.1 Code Development

Due to our two-year in-depth involvement in project IKON, we had already developed a fair understanding of the general structure of the institution. Additionally, due to the focus of project IKON, we had developed a superficial understanding of knowledge transfer being strongly related to individual actors. Therefore, 'Institution', 'Individual' and 'Knowledge Transfer' served as initial high-level codes. Following theoretical thematic analysis, we then systematically coded the transcripts on the level of sentences [?], focusing on the details of how knowledge transfer actually happens and which aspects of the institution (e.g., infrastructures, values, individuals) are involved in which manner, diversifying our thematic focus accordingly. To guide this procedure, we referred to our notes, in which we focused on the bodily behaviour of the interviewees, such as gestures and tone of voice, which aided us in coding interview statements according to expressed sentiment and significance. Specifically, the emergence of codes such as 'Exchange' and 'Benefit', as well as splitting 'Knowledge Transfer' into two separate coding concepts, proved fundamental to develop a more fine-grained understanding. Our coding procedure resulted in a final list of 12 high-level concepts, namely: Benefit, Constraints, Exchange, Individual, Infrastructure, Institution, Knowledge, Public, Research, Specification, Transfer and Values (see Figures 6 - 18, pp. 13 -25).

<sup>&</sup>lt;sup>5</sup>https://transcribe.wreally.com/, accessed 26/07/2019.

<sup>&</sup>lt;sup>6</sup>https://atlasti.com, accessed 26/07/2019.

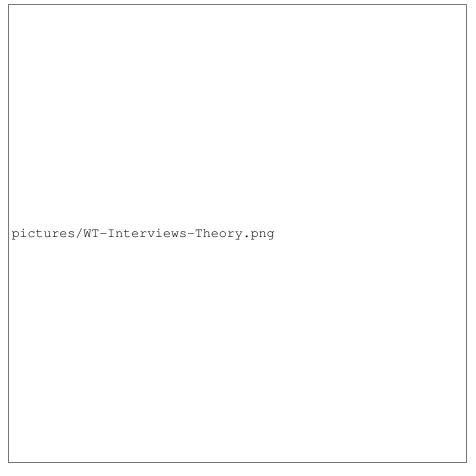


Figure 1: High-level analytical network on the state of Knowledge Transfer at the research institution, interlinked and with selected quotes (in German).

### 3.2 Thematic Analysis

Alongside the development of our coding procedure, we began formulating relationships among the aforementioned high-level concepts (see Figures 6 - 18, pp. 13 - 25). These were inferred from links between coded sentences on the level of individual expression, as well as those that correlated across transcripts. Specifically, the developed relationships of  $Transfer \rightarrow needs \rightarrow Infrastructure$ ,  $Individual \rightarrow needs \rightarrow Exchange$ , and  $Infrastructure \rightarrow supports \rightarrow Exchange$  illustrated the fundamental role played by infrastructures at the MfN in conceptualizing and conducting knowledge transfer. Based on these semantic links, we sought to develop analytical networks, of which we considered two. The first (Figure 1) was conceived to provide a model for understanding the state of knowledge transfer at the institution as expressed by our interviewees. This model then served as a lens to develop the second (Figure 2), which highlights

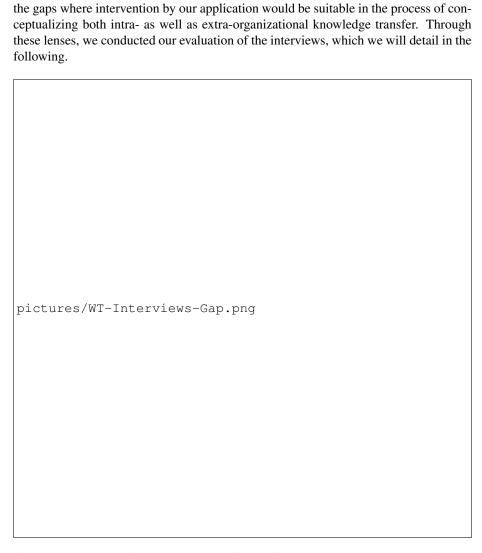


Figure 2: Condensed focus on the specific gaps for intervention that emerged in thematic analysis.

## 4 Results

In this section, we present the major insights we have gathered from our analysis procedure detailed above.

### 4.1 Differential and Localized Knowledge

At the research institution, knowledge itself necessitates a diversified understanding: it may concern knowledge about a specific artefact from the collection, knowledge about how to maintain a collection, or knowledge about how to translate collection material into an exhibition. Interviewees uniformly saw the individual researcher or employee as 'carriers of knowledge'; as expertise across subjects is highly specified. Crucially, exchange rarely happens below or even between the organizational structure of working groups; which typically comprise between 20 or 30 individuals. This suggests that there are significant gaps in knowledge exchange when considering a total staff of at least 300,<sup>7</sup> not to mention visiting scholars and students. An additional challenge to this localized expertise was seen in the form of the fixed-terms contracts of the research funding systems: once individuals leave the institution, their knowledge leaves with them. Documentation of individual research usually does not extend beyond scientific papers; with documentation of, for example, laboratory procedures being hand-written and rarely digitized. Another reason for a lack of knowledge transfer is the actual physical architecture of the MfN, a late 19th-century building which has been partly demolished and refurbished numerous times.<sup>8</sup> A consequence of this history is the absence of social spaces (e.g., cafeterias, canteens or lounges) where employees may congregate and communicate.

### 4.2 Knowledge Transfer as a Bridging Practice

Knowledge transfer was uniformly seen as transposing knowledge (in all its diversified understandings) to audiences that do not share the same focus; i.e., not transferring knowledge from mineralogist to mineralogist, but the mineralogist's expertise to cultural historians studying stone ornaments. Furthermore, we have found that knowledge transfer is tied to the perceived values of the research institution: openness, sustainability, interdisciplinary research. The dominant issue, however, was expressed to lie in being recognized as a research institution instead of 'only' as a museum. This is seen as two-fold: individual employees benefit from the reputation of the museum, which enables them to seek venues for knowledge transfer, yet also constrains them as the institution is commonly perceived as a museum. Knowledge transfer, as an emerging topic of discussion at the institution, is sometimes seen in this light: a uni-directional 'sale' of knowledge to an audience, without reciprocity for the institution. However, interviewees agreed that employees generally view and practice knowledge transfer positively without necessarily using the term, some for many years. Knowledge transfer, then, does not necessarily constitute a new concept at the museum: it is merely a term covering the established practices of employees interested in bridging (1) the perceived absence in public perception of the research activities of the institution, (2) the gap between scientific work and non-peer publics. We can see, therefore, that knowledge transfer is an established practice, though as a defining term it is not without controversy.

<sup>&</sup>lt;sup>7</sup>https://www.museumfuernaturkunde.berlin/en/team, accessed 05/22/2019.

<sup>8</sup>https://www.museumfuernaturkunde.berlin/en/building, accessed 05/22/2019.

### 4.3 Knowledge Transfer Activities as Individual Challenge

Concerning the actual activities that constitute knowledge transfer, interviewees agreed that there is general support at the directorate level, and that the institution as well as the individual benefit from the activities both publicly (in terms of reputation) and scientifically (when there is reciprocal transfer between research communities). Additionally, knowledge transfer was seen as a predominantly individual activity, relying on the drive of each employee to engage in any form of communication beyond scientific publication. However, there are numerous challenges to the specifics of conducting knowledge transfer activities (KTAs). There is, on a systemic level, very little time for researchers, as they are only systematically evaluated in terms of research papers. A lack of appreciation for the benefits of KTAs discourages employees from investing time in them. KTAs are seen as very time consuming, because they necessitate (1) understanding the relevance of a topic for a specific audience and vice versa, (2) identifying the appropriate format (e.g. tours, presentations, lectures, workshops) for transferring knowledge, (3) preparing (mostly scientific) material for non-expert audiences and (4) communicating with stakeholders about the activity. Common challenges lie in: (i) employees not knowing which types of audiences are suitable in the first place, (ii) a lack of experience in different types of formats for KTAs, (iii) material not being readily available for presentation, and (iv) employees not knowing what types of support on these issues are available at the institution, and who to ask about support in the first place. From our analysis, it is clear that there is a significant lack of exchange opportunities in general, and specifically a lack in exchange opportunities for the specifics (e.g., who has the expertise? What can be done? What has been done?) of KTAs.

### 4.4 Research Infrastructures as Mediators

The institutional infrastructure (i.e. laboratories, appliances and collections as well as digital infrastructures) was expressed as particularly well-suited for knowledge transfer; both as being representative of the knowledge generated at the institution, as well as the means for the actual practice of transferring knowledge. A collection, for instance, can serve as the means for situated transfer (e.g., in a tour or exhibition), as well as metaphorical instances of research knowledge (e.g., a fossil as used as a signifier for a research project). In a more narrow sense, the infrastructures that support data collection, storage and presentation were expressed to play a critical role in mediating the possibilities for knowledge transfer; such as the availability of structured data and annotated media files (e.g., images or 3d scans) for online access.

### 5 Discussion

From our analysis, we surmise that there is a lack in exchange opportunities between employees across the institutional structures: the exchange of knowledge about KTAs; concerning specifically (1) what formats of existing KTAs have made use of (2) research and institutional infrastructures in order to communicate to (3) specific audiences.

### 5.1 Implications

Consequently, we see various high level implications for our data-driven visualisation application. First, it should correlate KTAs according to their types of audiences and formats. Second, to support KTAs, our application should then provide a possibility to see what research topics have been used for what type of KTA. Third, it should allow employees to filter research through thematic rather than organizational lenses, to see how similar research has approached the challenges of knowledge transfer discussed above. Fourth, conducting KTAs require an assemblage of knowledge as discussed above. Our 'entities' in this assemblage are: infrastructures, research projects arranged by thematic focus, and existing KTAs. Our application should be able to visualize relationships between those entities. Lastly, it should offer means to easily share all of these aspects; both individually (with any member of the research institution) as well as directed—such as to teams with expertise in KTAs (e.g., the exhibition working group, the press office).

### 5.2 Limitations

A clear limitation lies in the small sample of interviews we conducted. With five interviews, the social, technical and institutional complexities of knowledge transfer at the MfN cannot be exhaustively described. However, due to the diverse organisational positions of interviewees as well as our familiarity with project IKON, we have sought to alleviate this constrained view. Furthermore, we did not conduct redundant coding of our data, thereby not fulfilling strict coder inter-reliability criteria. It is primarily for these two limitations that we have framed this work as a qualitative probe rather than a grounded theory. However, due to our iterative development of research questions and discussions during the evaluation procedure, we are confident that this qualitative probe into the mentioned complexities will prove fruitful for future research, as we discuss in our concluding section.

# 6 Conclusion and Perspectives

Our interviews have led us to a significant premise for the future development of the application and our research. The fundamental element of knowledge transfer lies in the research projects that are conducted by individual researchers. However, instead of reifying the existing organizational structures, and thereby possibly perpetuating a separation rather than integration of knowledge, our new premise is to show thematic similarities in the actual research activities. To this end, rather than visualizing research projects by departments, we are developing a Natural Language Processing (NLP) approach that assesses projects on the semantic similarity of their scientific abstracts. However, this approach necessitates the deployment of Machine Learning (ML) algorithms. Thereby, due to the inherent difficulty of interpreting ML outputs [?], we are confronted with a new research area: making ML outputs interpretable for nontechnical experts. This also connects to wider socio-political issues such as biases that become encoded in technology [?, ?]. Though we have critically reflected on this issue

by interviewing the developer of our NLP pipeline [?], we consider the interpretation of ML outputs by the MfN researchers and employees an especially promising focus of research. In this light, we have begun developing techniques to extract information from the NLP pipeline in order to foster informed usage, e.g. visualizing the uncertainty of technological decision-making [?]. In light of these efforts and the insights gathered from the interviews, we are developing a participatory design workshop, in which we aim to engage MfN employees in the selection of such techniques based on their actual need. This, again, is a particularly timely and challenging research avenue, as the principles of participatory design as an exploratory approach are challenged by the emergent dynamics [?] of ML-driven systems. With the use case of project IKON and the qualitative understanding we have developed, therefore, we see highly relevant research perspectives emerging for future work.

### Acknowledgements

This work is supported by the German Federal Ministry of Education and Research, (BMBF), grant 03IO1633 ("IKON – Wissenstransferkonzept für Forschungsinhalte, -methoden und -kompetenzen in Forschungsmuseen").

# 7 Appendix

This section presents the pages of the semi-structured interview guideline, as well as the code-book and high-level concept networks that have been generated from the qualitative interviews.

# pictures/IKON-MfN\_Leitfaden\_V3-1.pdf

Figure 3: Cover page of the semi-structured interview guideline.

### 7.2 Code Book

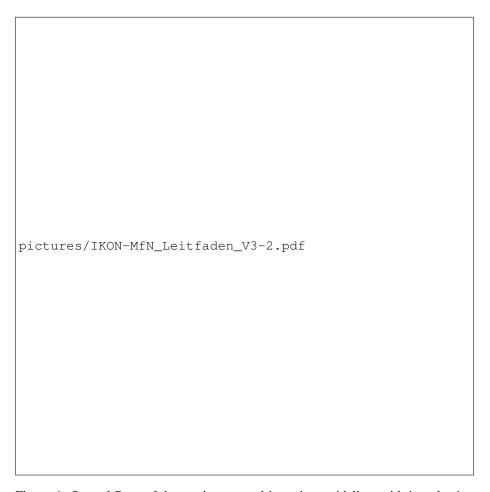


Figure 4: Second Page of the semi-structured interview guideline, with introduction and disclaimer.

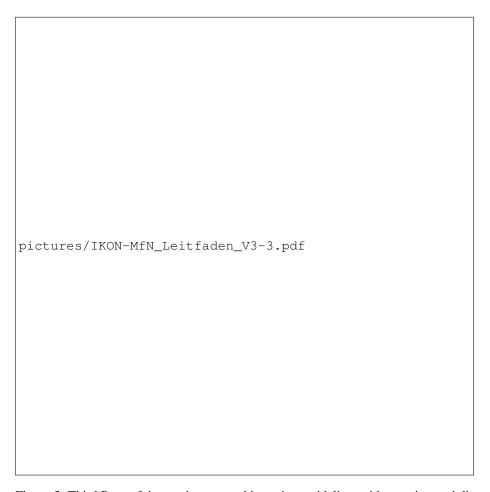


Figure 5: Third Page of the semi-structured interview guideline, with questions subdivided into themes.

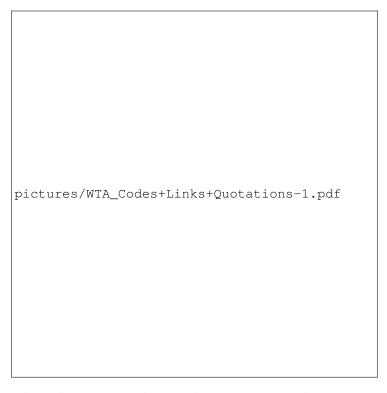


Figure 6: Code-Book with quotations and code-code-links, page 1.

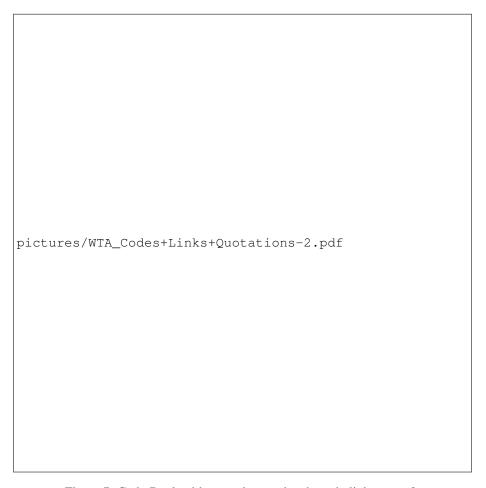


Figure 7: Code-Book with quotations and code-code-links, page 2.

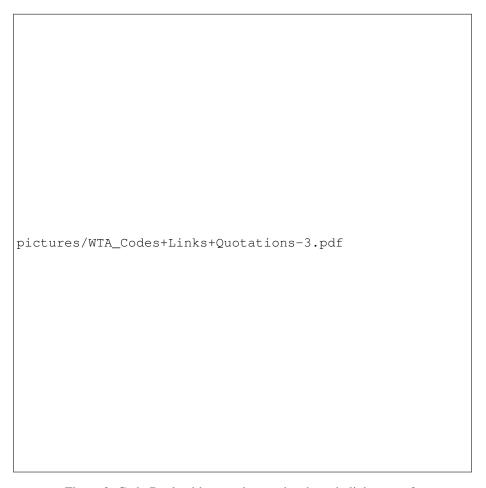


Figure 8: Code-Book with quotations and code-code-links, page 3.

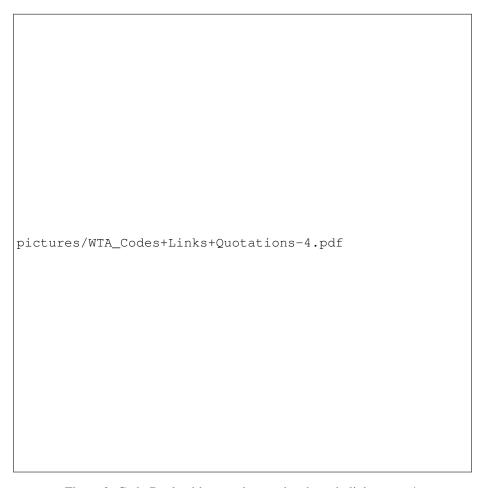


Figure 9: Code-Book with quotations and code-code-links, page 4.

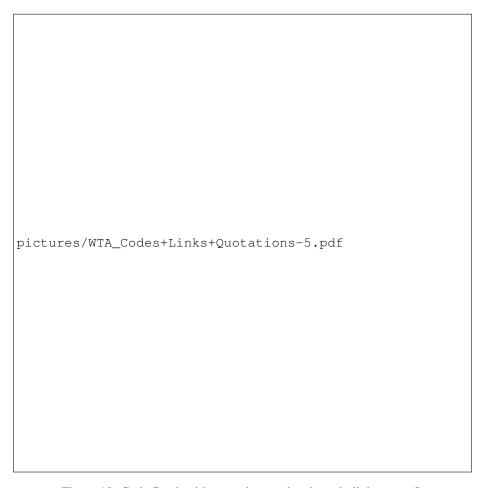


Figure 10: Code-Book with quotations and code-code-links, page 5.

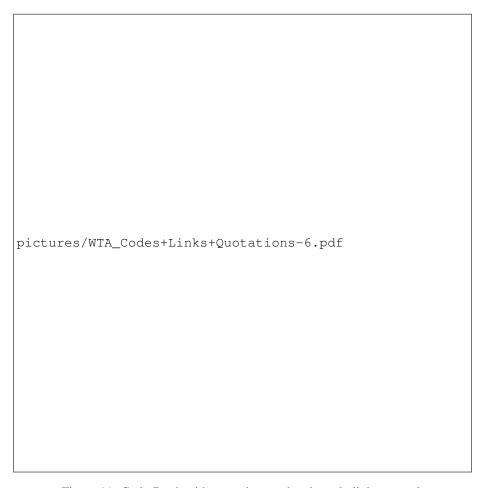


Figure 11: Code-Book with quotations and code-code-links, page 6.

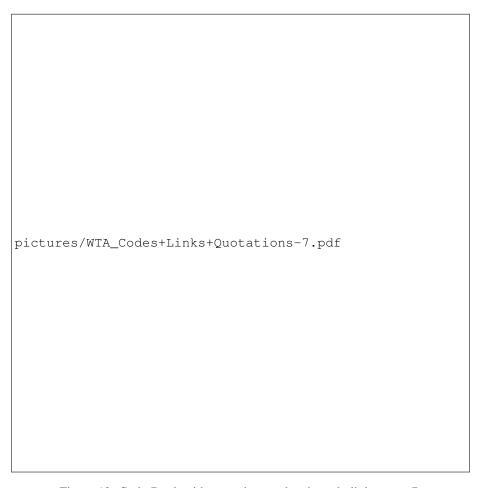


Figure 12: Code-Book with quotations and code-code-links, page 7.

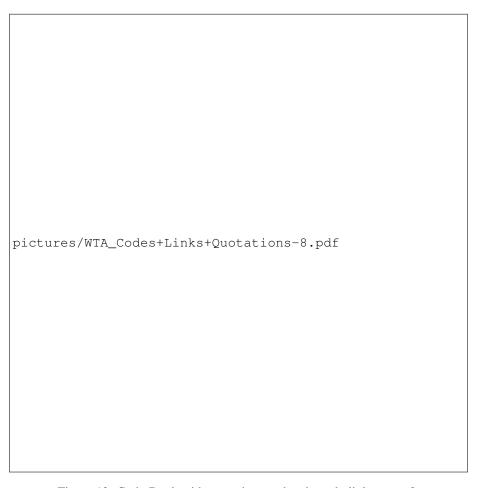


Figure 13: Code-Book with quotations and code-code-links, page 8.

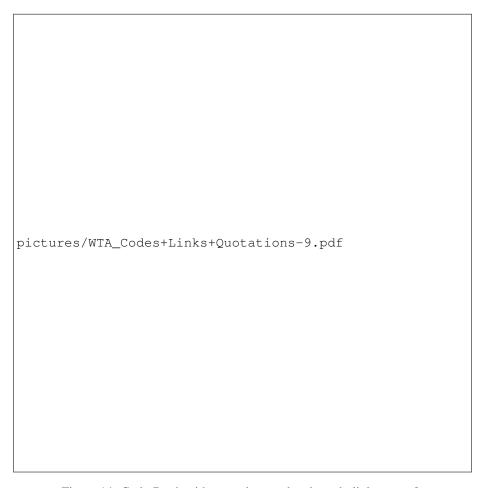


Figure 14: Code-Book with quotations and code-code-links, page 9.

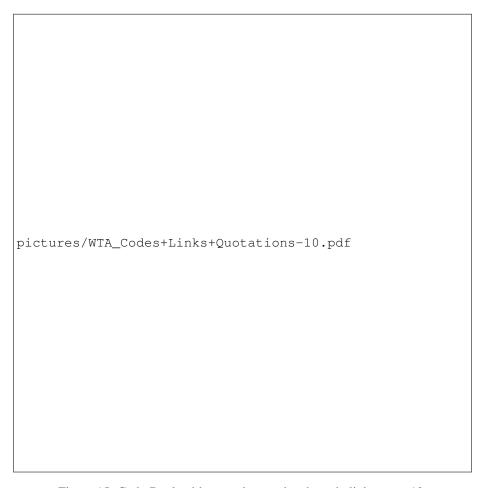


Figure 15: Code-Book with quotations and code-code-links, page 10.



Figure 16: Code-Book with quotations and code-code-links, page 11.

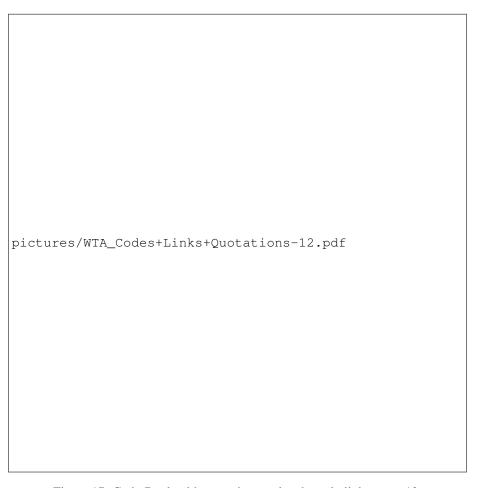


Figure 17: Code-Book with quotations and code-code-links, page 12.

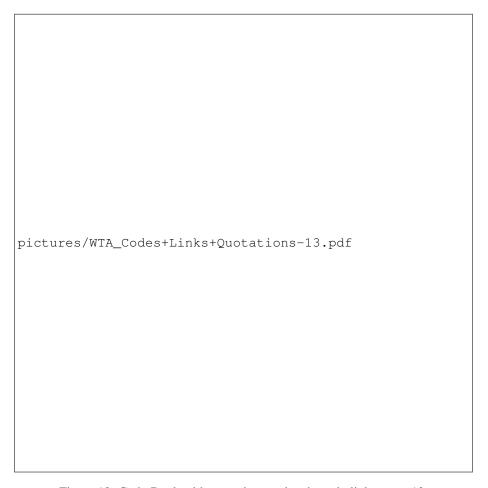


Figure 18: Code-Book with quotations and code-code-links, page 13.

# 7.3 Consent Document

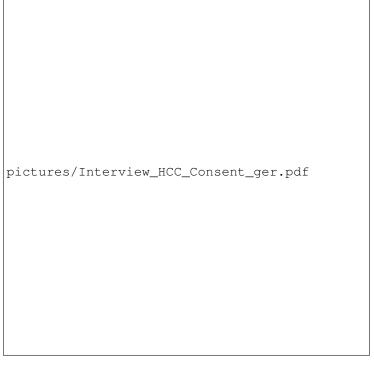


Figure 19: Consent document signed before each interview session.