Hochschule für
Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Master's Thesis

Doing Good with Data

Development of a Maturity Model for Data Literacy in Non-Governmental Organizations

Presented in Fulfillment of the Degree of Master of Science in Business Intelligence and Process Management at the Berlin School of Economics and Law

Helena Sternkopf | 334941 | sternkopf.h@gmail.com

M. Sc. Business Intelligence and Process Management

Hochschule für Wirtschaft und Recht Berlin | Berlin School of Economics and Law

Faculty of Business and Economics

First Supervisor: Prof. Dr. Roland Müller
roland.mueller@hwr-berlin.de

Second Supervisor: Prof. Dr. Frank Habermann
frank.habermann@hwr-berlin.de

Date of Submission: 23.01.2017 Word Count: 18.077

Abstract

This thesis contributes to the growing scholarly interest in exploring and describing data literacy. It is set out to identify and outline data literacy competencies for non-governmental organizations (NGOs) in an initial exploration of representing the topic in a maturity model. The aim is to offer a model that opens the discussion on data literacy and to stimulate increased awareness of what can be understood and expected from it. Based on the development of a preliminary maturity model with the help of a structured review on relevant literature streams, action design research is used to illuminate the question of how data literacy can be described in a maturity model for NGOs throughout three development iteration phases. The study draws on data from data practitioners of the Datenschule as well as a partnering NGO. The Datenschule is an educational program of the Open Knowledge Foundation Germany that provides technical training and consulting to NGOs on data topics. These organizations represent study objects that aspire to raise the awareness on data handling and conversion skills and want to discover new possibilities with the help of data. A thorough, qualitative content analysis of transcripts from five semi-structured expert interviews in the development phase of the artifact and four testing feedbacks yielded a data literacy maturity model that describes eleven competences to asses in organizations. Data has been collected between December 2016 and January 2017. The thesis describes the evolution of the model throughout the different iteration phases and provides a data literacy maturity grid that is complemented by a self-assessment excel tool for a comprehensive overview on different data literacy competencies at different levels. The results reveal that a maturity model for describing data literacy is a helpful tool to raise awareness and educate on the topic, but has difficulties to be sufficiently configurable for different organizational contexts. The author recommends further exploration and modification of the model regarding its application to varying contexts.

Keywords: Data Literacy, Maturity Model, NGO, Action Design Research, Data, Social Change, Artifact Design

Acknowledgement

This thesis does not only represent my work at the keyboard, but is a key milestone of the last three semesters of my graduate studies in Business Intelligence and Process Management at the Berlin School of Economics and Law (BSEL). The presented thesis is the result of many experiences and encounters I have made at the BSEL, and therefore I would like to express my gratitude to all those who gave me the possibility to complete this thesis.

First and foremost, I wish to thank my thesis advisors, Professor Dr. Roland Müller and Professor Dr. Habermann, directors of the BIPM program, who answered any upcoming questions rapidly and thoroughly at any time. They provided many new impulses, promoted creative thinking throughout my graduate studies and supported me from the beginning of my thesis. They gave me the freedom to explore this research topic based on my individual preferences and ideas, while providing academic assistance and moral support along the way.

Furthermore, I would like to express great appreciation to my interview partners, the data they provided and the time they invested for this research project. Especially, Helene Hahn, Moritz Neujeffski and Knut Perseke from the Datenschule. Thank you for investing your commitment and offering your guidance throughout different phases of this research project. Thanks to them, I had the opportunity to build the data literacy maturity model and test it in a real-world organizational setting. They also helped me to come up with the thesis topic and guided me over the past months of development.

A special thank goes to the unconditional support of my friends and family during the ups and downs of my graduate studies. Without their moral assistance throughout the last semesters and before, all past and future accomplishments would not have been possible.

Affidavit / Eidesstattliche Erklärung

I declare that I wrote this thesis independently and on my own. I clearly marked any language or ideas borrowed from other sources that are not my own and documented their sources. The thesis does not contain any work that I have handed in or have had graded as a Prüfungsleistung earlier on.

I am aware that any failure to do so constitutes plagiarism. Plagiarism is the presentation of another person's thoughts or words as if they were my own—even if I summarize, paraphrase, condense, cut, rearrange, or otherwise alter them. I am aware of the consequences and sanctions plagiarism entails. Among others, consequences may include nullification of the thesis, exclusion from the Master program without a degree, and legal consequences for lying under oath. These consequences also apply retrospectively, i.e. if plagiarism is discovered after the thesis has been accepted and graded.

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quelle als Entlehnung kenntlich gemacht. Die Arbeit enthält kein Material, das ich bereits zu einem früheren Zeitpunkt als Prüfungsleistung eingereicht habe.

Mir ist bewusst, dass die ungekennzeichnete Übernahme fremder Texte oder fremder Ideen als Plagiat gilt, selbst wenn diese zusammengefasst, umschrieben, gekürzt, oderanderweitig verändert wurden. Die Konsequenzen eines Plagiats sind mir bekannt. Die möglichen Konsequenzen umfassen, unter anderem, ein Nichtbestehen der Masterarbeit, den Ausschluss von weiteren Prüfungsleistungen im Studiengang, oder zivilrechtliche Konsequenzen, die mit dieser eidesstattlichen Erklärung verbunden sind. Diese Konsequenzen können auch nachträglich zur Anwendung kommen, also nachdem die Arbeit angenommen und korrigiert wurde.

Name:	Helena Sternkopf
Title of Thesis:	Doing Good with Data - Development of a Maturity Model for Data Literacy in Non-Governmental Organizations
Date:	23.01.2017
Signature:	

Table of Contents

List of Abbreviations

ADR	Action Design Research
BARS	Behaviorally Anchored Rating Scale
BPM	Business Process Management
CMM	Capability Maturity Model
DL	Data Literacy
DR	Design Research
MM	Maturity Model
QM	Quality Management
QMMG	Quality Management Maturity Grid

List of Figures

1 Introduction

1.1 Motivation and Relevance of the Topic

"We are drowning in information, while starving for wisdom. The world henceforth will be run by synthesizers, people able to put together the right information at the right time, think critically about it, and make important choices wisely."

E. O. Wilson

Experts of the UN Data Revolution Group (2014: 2) agree with Wilson's introductory quotation, claiming that "data are the lifeblood of decision-making and the raw material for accountability". Datafication (Flyverbom and Koed Madsen, 2015) transferred itself into an essential point that many countries and organizations put on their agenda (Janssen et al., 2012). The publication and use of data is commonly associated with a vast amount of benefits such as "more accountability, public engagement, transparency, new or improved public services, improved policy-making, economic growth, innovation, [and] optimization of administrative processes" (Janssen et al., 2012: 8-9). Within the course of the last years, data literacy (DL), which can generally be described as "the ability to consume for knowledge, produce coherently, and think critically about data" (Gray et al., 2012: 122), turned out to be a meaningful priority to different groups. Educational programs and nonprofit organizations designed programs to familiarize children with coding, but also the emerging importance of data journalism or analytical skills on the modern job market show that we have to equip ourselves with new skills for the data era (Data-Pop Alliance, 2015).

The motive for promoting data literacy seems unequivocal. Still, large parts of our society do not have much clarity about what data literacy tries to explain and lack the imagination of what to expect from it (Data-Pop Alliance, 2015; Wolff et al., 2016). On the other hand, when regarding the use of data in businesses and science, we get the impression that today using and making sense of data is already common practice in these domains. However, if we are shifting our perspective to social change organizations, the landscape appears quite different. Researchers claim that non-governmental organizations (NGO) are still behind data practitioners in the scientific and business environment (Desouza and Smith, 2014). This makes it obvious that there still is a large gap between the potential of data literacy for individuals and organizations, and its actual use in progressing societal change due to missing resources, capability or opportunity (Desouza and Smith, 2014; UN Data Revolution Group, 2014). Due to the lacking scientific attention to data literacy and social change, this research focuses on investigating possible data literacy practices for social change organizations apart from business and science.

As described earlier, society has little understanding of what data literacy describes (Data-Pop Alliance, 2015; Wolff et al., 2016). Additionally, Bhargava and D'Ignazio (2015: 1) argue that there is "a lack of consistent and appropriate approaches for helping novices learn to 'speak data'" and further argue that most tools are designed for users and not learners. Hence, finding a way to picture data handling and conversion skills and translating them for an audience of different backgrounds and competencies appears essential. One possibility to translate, represent and assess capabilities on different levels are maturity models (Becker et al., 2009; Maier et al., 2012; Pöppelbuß and Röglinger, 2011). While maturity grids often share a common composition, their subject matter can vary depending on the specific context, which makes it a capacity illustration tool that can serve different purposes (Maier et al., 2012). Rummler and Brache (1995) mention maturity models as tools for improving systems and roadmaps for steering organizations and thus can help to raise awareness for data literacy on different levels without neglecting the big picture. This aspect appears essential in the context of data literacy, since today data literacy has to transcend the binary considerations, only differentiating between literate and illiterate (Data-Pop Alliance, 2015). Hence, multiple levels of data literacy have to be represented and appreciated in a respective model (Data-Pop Alliance, 2015). Measuring elements in accordance with their maturity is an emerging concept, that equips decision-makers and practitioners with the right information to assess the status of certain domain knowledge in their organization and drive strategic improvements and resolutions (Enkel et al., 2011). It therefore can help to better describe what to understand and expect from working with data through defining basic thresholds. As a consequence, this study combines the challenges and ambiguity of data literacy with the opportunities and advantages of representing competencies in a maturity model.

1.2 Aim of the Research

This study aims at describing data literacy in a maturity model for NGOs, in a world where data experts explain that "Data literacy is a 'strange thing' that can only be promoted when specifying what is meant and expected from it." (Data-Pop Alliance, 2015: 25). In this regard, the expected outcome of this research will especially focus on the creation of a clear understanding of the competencies that are needed to engage with data and what can be expected from it. It will help to evaluate data literacy skills on different competence levels, and thereby support to plan and steer future activities. Nevertheless, it will not serve as a unique solution that can evaluate satisfactory or unsatisfactory performance, since this needs to be assessed based on the specific context of the organization under investigation. The data literacy maturity model (DLMM) should rather be understood as an educational endeavor to raise awareness regarding data handling and conversion skills and help to make project work more scalable through a better overview on existing qualifications and gaps. As a consequence, future training programs can be planned for in accordance with the evaluation result, which can save time and costs.

Researchers in this field have not yet sufficiently included effective ways to bridge the gap between the outstanding possibilities of data and the lack of public engagement due to a lack of understanding. They therefore prescribe to direct further studies concentrating on promoting the interests and ability to (re)use data (Hellberg and Hedström, 2015), to which this study aims to contribute. With data literacy describing the abilities regarding the use of data for thinking and arguing for real-world problem solving, data handling skills emerge to become a "life skill, as daily interactions with data become evermore commonplace" (Wolff et al., 2016: 10). This study tries to answer the questions around what competencies do individuals and organizations need to learn from and solve problems with data.

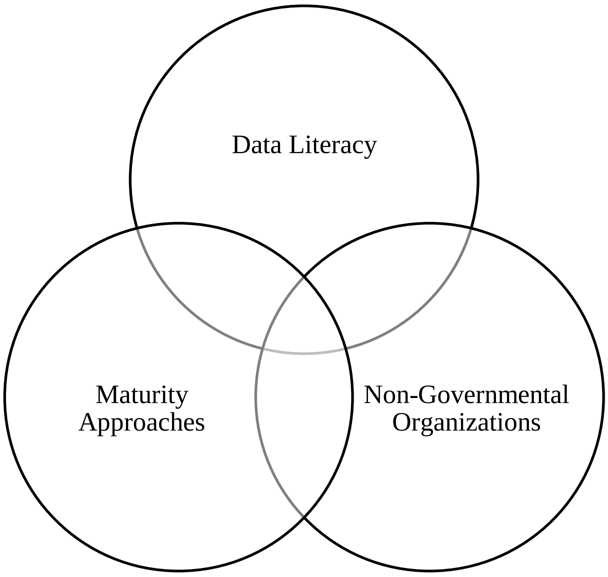
Therefore, the overall research question of this investigation is:

RQ: *How can we describe data literacy in a maturity model for non-governmental organizations?*

1.3 Research Structure

The study investigates three main streams of thoughts with the help of a structured review of existing literature (see Figure 1). First, the status quo of technology and social change organizations, as a central study object, will be presented to better understand their challenges and opportunities. Second, data literacy will be defined through considering definitions from different domains to find a common understanding and appreciate the ambiguity of it. After that, maturity approaches are introduced to better apprehend the mechanisms and underlying rationale of maturity models and their development.

Figure 1 Literature Review Overview



The theoretical foundations will close with a synthesize of the different research streams through the development of a preliminary maturity grid, and a critical reflection on maturity approaches. The preliminary data literacy maturity model will be enriched and iteratively developed with insights from qualitative expert interviews with data practitioners. Later it will be tested and evaluated in a real organizational context using action design research. The findings of the model development will be put into a broader perspective and answers to potential criticism will be provided. The paper closes with a conclusion of the study containing limitations and possible future research directions.

2 Theoretical Foundations

2.1 Technology and Social Change

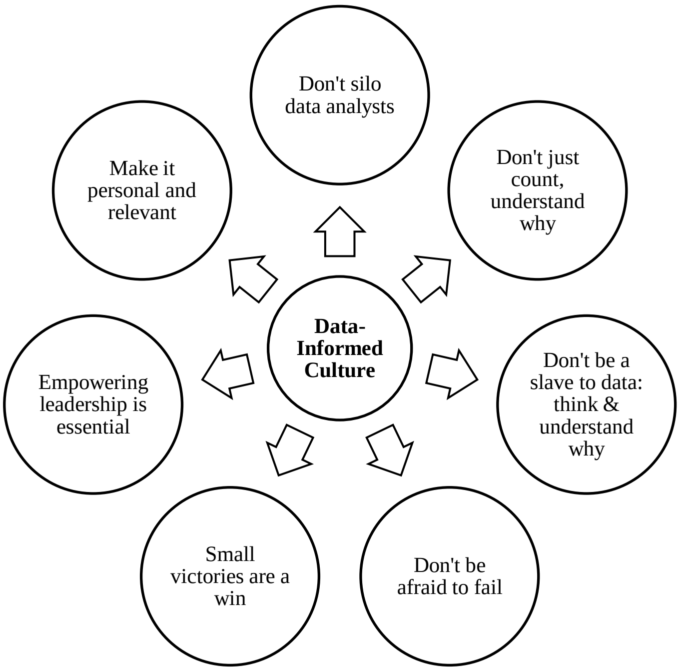
The term non-governmental organization (NGO) was first coined in 1945, because of the need for the UN to distinguish between participation guidelines for intergovernmental agencies and international private organizations (Kütting, 2009). According to the UN, many kinds of private organizations can be denoted as NGOs, as long as they are independent from governmental authorities, do not object governments as a party, are not-for-profit and do not constitute a criminal group (Kütting, 2009). Many different bodies are described as NGOs and there is no universally accepted definition for NGOs (Kütting, 2009). Nevertheless, the features described earlier are usually shared among NGOs that can be active in various areas such as human rights, education, immigration, gender equality and many more social change topics and can be organized at local, national or international level (Kütting, 2009). The broad definition of NGOs as proposed by Kütting (2009) will serve as a base for this study.

Looking at social change organizations and technology-based approaches, it becomes obvious that they are actually not too alien from each other. Certain social problems already integrate data to problem-solving. One example is the use of traffic data to improve traffic in urban areas or weather data to better react to menacing natural catastrophes (Desouza and Smith, 2014). At the same time, if we are looking for solutions to more wicked social challenges, education or homelessness for example, data-driven efforts are still in their infancy. Social problems are more complicated compared to their technical counterparts because of the different stakeholders involved (Desouza and Smith, 2014). Desouza and Smith (2014) state that tackling these wicked problems with the help of data are on the agenda of various government and nonprofit organizations. However, there is only “limited cooperation and data sharing among them” (Desouza and Smith, 2014: 1). On the one hand, this is due to inadequate resources regarding information technology when compared to sciences or business, which might have huge access to financial, customer and other data. On the other hand, the infrastructural obstacles are complemented by lacking data quality and missing information, which can only be solved through data-sharing agreements and data collaboration among stakeholders (Desouza and Smith, 2014).

Despite the many challenges that still exist when it comes to promoting the benefits of data for social change, it is obvious that NGOs and other organizations will continue to invest in exploring this topic in the future (Desouza and Smith, 2014). And already today, there are success stories. One successful example of using data for the social good is how Polaris and two other international anti-human trafficking organizations were able to aggregate their hotline data to transfer it to an international hotline for fighting human trafficking (Desouza and Smith, 2014). Moreover, datafication can also stimulate new forms of philanthropy (Flyverbom and Koed Madsen, 2015). In this context, the term data philanthropy (Pawelke and Tatevossian, 2016) has been coined, which depicts the moment of charitable giving when corporations and governments donate data, which "could help to track diseases, avert economic crises, relieve traffic congestion, and aid development" (Pawelke and Tatevossian, 2016: 1). This example demonstrates that data is already understood as an asset that needs to be shared and used.

In this momentous instant with a lot of ambiguity, but also freedom for creative participation, especially nonprofit organizations can use the change to drive their activities and become data-informed. A data-informed culture describes a state of "agile, responsive, and intelligent [organizations] that are better able to succeed in a rapidly changing environment" (Kanter et al., 2012: 30). One example of how to move towards a successful integration of data activities in an organizational culture is DoSomething.org, which is an initiative that promotes volunteer services. DoSomething.org showcased which aspects to consider when transferring into a data-informed organization. Figure 2 describes key elements of this process, that other organizations should consider when engaging with data in an organizational context.

Figure 2 Elements of a Data-Informed Culture



Based on Kanter et al. (2012: 32-34)

Having internalized basic principles to create a data-informed culture, eminently nonprofit organizations can use data in their sociopolitical work as a complementing information source to advance projects and offer an increased social inclusion as well as to improve internal processes and critical decision-making (Datenschule, 2016). However, data experts still agree that only a few nonprofit organizations have discovered the value of data for their activities (Desouza and Smith, 2014). In this regard, it is essential to raise awareness and educate on where to find data, how to access, clean, analyze and visualize it with digital tools and how to integrate it into the project portfolio (Datenschule, 2016).

2.2 Data Literacy and its Dimensions

Key principles of data literacy have their roots in diverse areas such as mathematics, data mining, statistics, graphic design, and information visualization (Calzada Prado and Marzal, 2013; Fry, 2004). The term data literacy has gradually developed to a widespread buzzword of the recent discourse on the data revolution, which depicts the application and indications of data as a social peculiarity (Data-Pop Alliance, 2015; UN Data Revolution Group, 2014). Nevertheless, even with this attention, the meaning of data literacy as well as what can be expected from it remains ambiguous (Bhargava and D'Ignazio, 2015; Calzada Prado and Marzal, 2013; Data-Pop Alliance, 2015; Gray et al., 2012).

To overcome this ambiguity when describing DL, we first start by looking at the two words separately. Data can be understood as nearly anything: text, images, videos, audio, numbers, etc. It can be described as an object, variable, or piece of information that is identifiable and has the capability to be collected and stored (Data-Pop Alliance, 2015). Generally, data can be distinguished between structured and unstructured. Felden (2006) defines structured data as objects that are machine-readable and unstructured data as items that are not easily machine-readable.

Literacy has proven to be a complex and dynamic concept, which is still interpreted in various ways (UNESCO, 2004). Researchers from UNESCO (2004: 13) define literacy as "the ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts. Literacy involves a continuum of learning in enabling individuals to achieve their goals, to develop their knowledge and potential, and to participate fully in their community and wider society.".

Combining the two terms and formulating a definition of data literacy, leaves us with a description that could read as follows:

> *Data literacy is a continuous learning journey that creates the ability to identify, understand, interpret, create, communicate and compute pieces of information (data), to develop knowledge and the ability to participate fully in our society.*

This in fact, is not too distant from the definitions we find in literature (see Figure 3). Researchers who have investigated the field of data literacy, such as Calzada Prado and Marzal (2013) argue that "the identification of the competencies needed to be data literate is a matter presently under study by a fairly large community of researchers […] in different domains", highlighting that these investigations are all subject to the research context and purpose which makes the topic all the more ambiguous. All data literacy definitions that are reviewed in the following are based on specific contexts, which illustrates the importance of data literacy in many areas, but also delimits to find one common definition (see Figure 4). The reviewed literature has been analyzed in accordance with the concept-centric literature review as described by Webster and Watson (2002).

Vahey et al. (2006), Mandinach and Gummer (2013) and Deahl (2014) suggest definitions in the context of data literacy at schools.

Vahey et al. (2006: 1) propose that "data literacy includes the ability to formulate and answer questions using data as part of evidence-based thinking; use appropriate data, tools, and representations to support this thinking; interpret information from data; develop and evaluate data-based inferences and explanations; and use data to solve real problems and communicate their solutions."

Mandinach and Gummer (2013) describe data literacy regarding how teachers use student data to improve their teaching. According to them (Mandinach and Gummer, 2013: 2 , data literacy is "the ability to understand and use data effectively to inform decisions. It is composed of a specific skill set and knowledge base that enables […] to transform data into information and ultimately into actionable knowledge. These skills include knowing how to identify, collect, organise, analyse, summarise and prioritise data. They also include how to develop hypotheses, identify problems, interpret the data, and determine, plan, implement, and monitor courses of action."

In Deahl's view (2014: 41), data literacy is "the ability to understand, find, collect, interpret, visualize, and support arguments using quantitative and qualitative data."

Investigating data literacy especially for librarians, Calzada Prado and Marzal (2013: 130-131) describe data literacy competencies as the ability to: understand data, find and/or obtain data, read, interpret and evaluate data, manage data, use data.

Bhargava and D'Ignazio (2015: 1) conducted research on evaluating and designing data literacy learning tools and create a leaner picture of data literacy, defining it as "the ability to read, work with, analyze and argue with data".

Apart from this, a recent study of the School of Data (Slater, 2016) builds on the data literacy definition as proposed by Bhargava and D'Ignazio (2015) and extents it based on the insights from their research, with special focus on data journalists. During this study, they identified the following key properties of data literacy: know how to find data, apply critical thinking skills to data, ask questions to data and find answers, find specific outputs (stories or visualizations) in data, use it to advance one's goals, feel comfortable around data and working with it, do statistical analysis with data.

Wolff et al. (2016: 23) interpret data literacy as "the ability to ask and answer real-world questions from large and small data sets through an inquiry process, with consideration of ethical use of data. It is based on core practical and creative skills, with the ability to extend knowledge of specialist data handling skills according to goals. These include the abilities to select, clean, analyse, visualise, critique and interpret data, as well as to communicate stories from data.".

Figure 3 Defining Data Literacy

Author(s)	Focus	Data Literacy Properties	Explanation
Vahey, P., Yarnall, L., Patton, C., Zalles, D., & Swan, K. (2006). Mathematizing middle school: Results from a cross-disciplinary study of data literacy. In Annual Meeting of the American Educational Research Association, San Francisco, CA.	School Data Literacy	Formulate and answer questions Use appropriate data, tools, and representations Interpret Develop, evaluate, solve real problems Communicate	NA
Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. Educational Researcher, 42(1), 30-37.	School Data Literacy	Identify Collect Organize Analyze Summarize and prioritize Develop hypotheses Identify problems Interpret Determine, plan, implement, and monitor courses of action	NA
Calzada Prado, J., and Marzal, M.Á. (2013). Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. Libri 63, 123–134.	Librarian Data Literacy	Understand	know what is meant by data and be aware of the various possible types of data, be aware of the role of data in society, how and by whom they are generated, and their possible application
		Find and/or obtain	be aware of the possible data sources, be able to evaluate them and select the ones most relevant to a given problem, be able to detect when a given problem cannot be solved with existing data and undertake research to obtain new data
		Read, interpret and evaluate	be aware of the various forms in which data can be presented (written, numerical or graphic) and how to interpret them, be able to evaluate data critically
		Manage	be aware of the need to save the data selected or generated and of descriptive or other data associated therewith, for due identification, management and subsequent reuse
		Use	be able to prepare data for analysis, analyze them with necessary tools, be able to synthesize and represent the results suited to their purpose and the audience, make ethical use of data, interpreting results transparently and honestly
Deahl, E. (2014). Better the Data You Know: Developing Youth Data Literacy in Schools and Informal Learning Environments. M.S. Thesis, Massachusetts Institute of Technology	School Data Literacy	Understand	Requires an understanding of what data is, what types of data exist, and how data is generated. Have knowledge of the role and impact of data in society across different contexts and the ethical implications of using data
		Find	Requires the skills necessary to ask questions that can be researched using data, find relevant data sources, compare and evaluate sources, and check for bias and inaccuracy

		Collect	Encompasses the ability to collect qualitative and quantitative data, including methods such as conducting interviews, creating surveys, making observations, and taking measurements
		Interpret	Requires the ability to prepare data for analysis, critically analyze data in a range of formats, and develop inferences
		Visualize	Includes the ability to communicate data using a range of visual representation methods such as tables, graphs, and maps
		Support arguments	Includes the ability to use data as evidence to support arguments and tell stories, taking into consideration the larger cultural, social, and political context. Ability to use data to solve problems and communicate their solutions
Bhargava, R., and D'Ignazio, C. (2015). Designing Tools and Activities for Data Literacy Learners. MIT Center for Civic Media, Emerson Engagement Game Lab.	Data Literacy Tool Design	Read	understand what data is, and what aspects of the world it represents
		Work with	create, acquire, clean, and manage data
		Analyzing	filter, sort, aggregate, compare, and perform analytic operations on data
		Argue	use data to support a larger narrative to communicate messages
Slater, D. (2016). Research Results Part 1: Defining Data Literacy. School of Data.	Data Journalism	Know how to find data	track down sources of existing data, know how to collect data if it does not exist yet
		Apply critical thinking skills to data	ability to do data quality assessment, contextualizing specific information to other aspects
		Ask questions to the data and find answers	ability to ask questions to data and ultimately find answers as one of the goals of data literacy trainings
		Find specific outputs (stories or visualizations) in data	recurred among participants from the field of data journalism, importance of finding stories and other journalistic outputs
		Use it to advance one's goals	link between data and action was evident
		Feel comfortable around data and working with it	promoting comfort around data (and bringing down the psychological barriers that exist between people and data)
		Do statistical analysis with data	ability to work with basic statistics
Wolff, A., Gooch, D., Cavero Montaner, J. J., Rashid, U. and Kortuem, G. (2016) 'Creating an understanding of data literacy for a data-driven society', The Journal of Community Informatics, pp. 9–26.	Community Data Literacy	Ask and answer questions Ethical use of data Select Clean Analyze Visualize Critique and interpret Communicate stories	NA

Despite the obvious overlaps like finding, analyzing and communicating data, it becomes obvious that due to their different foci the definitions differ slightly from each other and vary in their preciseness.

Apart from the data literacy properties, which already provide an overview of the abilities of data literate individuals, practitioners and learners should transcend the dichotomy of data literacy into literate and illiterate and appreciate the many-sidedness of it (Data-Pop Alliance, 2015). When describing data literacy, multiple ways to consider it in accordance to individual contexts and capacities must be acknowledged (Data-Pop Alliance, 2015). This will be reflected later on and examined further in the development of the dimensions of the maturity model to find a suitable definition for NGOs.

2.3 Maturity Approaches and Design Principles

To establish a reasonable maturity model for data literacy, basic principles of maturity approaches need to be understood. Traditionally, measuring capabilities is complex, but can help organizations “achieving the company’s objectives” (Chiesa et al., 2008) since it can support various areas in an organization including decision-making, motivating employees, stimulating learning as well as improving communication and strategic planning (Loch and Tapper, 2002). One approach of assessing organizational capabilities and stimulating improvement is by means of maturity models (Pöppelbuß and Röglinger, 2011). The basic concept of all models is based on the fact that things change over time and that these changes can be predicted and regulated to a certain extent (Ross, 2003). Maturity, broadly spoken, depicts that someone or something has arrived at a state of plentitude and readiness (Cambridge Dictionary, 2016). Based on the presumption of anticipated patterns, maturity models illustrate how organizational capabilities evolve in a stage-by-stage manner along a plausible maturation path (Enkel et al., 2011; Gottschalk, 2009; Vallerand et al., 2015). Essentially, maturity models briefly describe prevalent behavior of an entity at different levels of maturity, for several elements of the area under study. This study concentrates on the process of how to define data literacy and its path to maturity, through considering different levels and dimensions / competencies. Knowing which elements to alter can thus help organizations to move to their desired level in the respective area (Maier et al., 2012). Nevertheless, the top level does not necessarily need to be the most desirable one to reach, but has to be evaluated in close consideration of specific organizational contexts and individual objectives (Maier et al., 2012; Ross, 2003).

That being said, there are also different purposes of maturity models that can be distinguished between descriptive, prescriptive and comparative (see Figure 4) (Becker et al., 2009; Maier et al., 2012). For the purpose of this study which is restricted by the given time limit, a descriptive maturity model is developed.

Figure 4 Comparing Descriptive, Prescriptive and Comparative MM

Descriptive MM	**Prescriptive MM**	**Comparative MM**
AS-IS assessment	Indicates to identify desirable maturity levels and provides guidelines on improvement measures	Allows for internal or external benchmarking
Current capabilities of entity are assessed to given criteria	A specific and detailed course of action is suggested	Given sufficient historical data from many assessment participants, maturity levels of similar organizations can be compared
MM used as diagnostic tool maturity levels and status can be reported to internal & external stakeholders		

Based on Becker et al. (2009) Maier et al. (2012)

The design of the data literacy maturity model will follow a combination of two procedure models that have been developed by other researchers (Becker et al., 2009; Maier et al., 2012) to cover all necessary aspects of the maturity model design and guarantee a clear development structure. Generally, the development of maturity frameworks can be considered as part of design research (Becker et al., 2009). Becker et al. (2009: 214) describe that "maturity models can be understood as artifacts which serve to solve the problems of determining a company's status quo of its capabilities and deriving measures for improvement therefrom.". This is in close conformance with the principle of design science, which aims at improving problem-solving capacities through the creation of artifacts, such as models or methods (Hevner et al., 2004).

As for the design of the data literacy maturity model, development procedures as suggested by Becker et al. (2009) and Maier et al. (2012) (see Appendix A and B) will be considered. Becker et al. (2009) extract their requirements (see Figure 6) and an eight-phase procedure model (see Figure 5) from Hevner's (2004) design science guidelines.

Figure 5 Requirements for the Development of Maturity Models

Requirement	Description
1 Comparison with existing maturity models	The need for the development of a new maturity model must be substantiated by a comparison with existing models. The new model may also just be an improvement of an already existing one.
2 Iterative procedure	Maturity models must be developed iteratively.
3 Evaluation	All principles and premises for the development of a maturity model, as well as usefulness, quality and effectiveness of the artifact, must be evaluated iteratively
4 Multi-methodological procedure	The development of maturity models employs a variety of research methods, the use of which needs to be well-founded and finely attuned.
5 Identification of problem relevance	The relevance of the problem solution proposed by the projected maturity model for researchers and/or practitioners must be demonstrated.
6 Problem definition	The prospective application domain of the maturity model, as well as the conditions for its application and the intended benefits, must be determined prior to design.
7 Targeted presentation of results	The presentation of the maturity model must be targeted regarding the conditions of its application and the needs of its users.
8 Scientific documentation	The design process of the maturity model needs to be documented in detail, considering each step of the process, the parties involved, the applied methods, and the results.

Based on Becker et al. (2009:214)

Not only because of the considerable attention of Hevner’s design principles (2004), but also because of its applied use in established companies such as Deloitte, Becker’s design procedure (2009) will be consulted. Maier et al. (2012: 141-144) propose a four-phase procedure, aimed at presenting a reference point and guidance for developing maturity grids based on a review of 24 existing maturity grids. Parts of Maier’s procedure model (2012) will be integrated into the guidelines as suggested by Becker (2009), especially in phase four – the iterative design of the maturity model since the descriptions of Maier complement the step in a more precise way.

Figure 6 Phases of the Maturity Model Development

Development Phase	Description	Related Requirements
1 Problem definition	For this purpose, both the targeted domain vs. partial discipline and the target group (e. g. intra-corporate vs. external) of the maturity model need to be determined. The problem relevance, i.e. the actual demand for the maturity model, must be clearly demonstrated.	6 5
2 Comparison of existing maturity models	A comparison of existing maturity models can be found in all reviewed examples. Shortcomings or lack of transferability often motivate improvements of older models.	1
3 Determination of development strategy	The most important basic strategies are: the completely new model design, or the enhancement of an existing model; the combination of several models into a new one; and the transfer of structures or contents from existing models to new application domains.	8
4 Iterative design of the maturity model	Sub-steps of this phase: selecting the design level, selecting the approach, designing the model section, and testing the results will be iterated. In the next step, the selected model section needs to be designed in accordance with the chosen procedure. Maier et al. 2009 suggest the following steps: a. Select process areas / dimensions b. Select maturity levels c. Formulate cell text d. Define administrative mechanisms The result then must be tested for comprehensiveness, consistency, and problem adequacy.	2 4 3
5 Conception of transfer and evaluation	Different forms of result transfer for the academic and the user communities need to be determined. Besides the widespread publication of document-based check lists, and manuals, software-tool supported accessibility of the maturity model (e. g. via internet) offers another alternative. Possibilities for the evaluation of the problem solution proposed by the maturity model should be incorporated into the transfer design (possibility for feedback as early as the design stage, e. g. questionnaires, software tools, workshops).	4
6 Implementation of transfer media	Make the maturity model accessible in the planned fashion for all previously defined user groups. In the reviewed projects, voluminous reports prevail. Self-assessment questionnaires are sometimes available, but are, for commercial reasons, often not made generally accessible.	7
7 Evaluation	Should establish whether the maturity model provides the projected benefits and an improved solution for the defined problem. The defined goals are to be compared with real-life observations. For this purpose, case studies, or online evaluations may be made accessible. The outcome of the evaluation may cause a reiteration of the design process. It is also possible that the maturity model may be retained unchanged, while the conception of transfer and evaluation may need to be modified.	3 2
8 Rejection of maturity model	Negative results may lead to a rejection of the model, in which case the model should be taken off the market.	

Based on Becker et al. (2009: 217-219), Maier et al. (2012: 149-155)

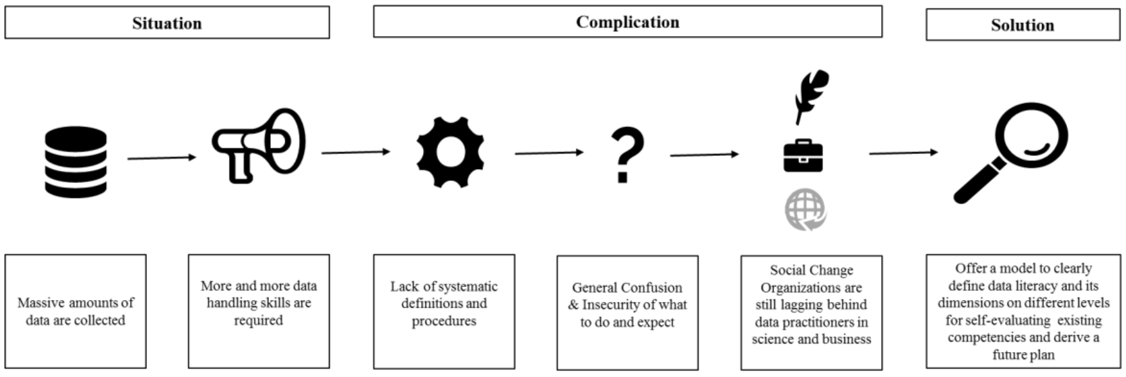
3 Synthesis: Development of a Preliminary Data Literacy Maturity Model

As mentioned earlier, a combination of both procedure models as proposed by Becker et al. (2009) as well as Maier et al. (2012) will be employed for the development of the preliminary data literacy maturity model, in which Maier's development phase will be integrated into Becker's iterative design process. The preliminary model will afterwards be enriched and refined using action design research (Sein et al., 2011).

3.1 Problem Definition

We live in times in which massive amounts of data are collected every day. Data streams emerge from various new sources, such as mobile phones, credit cards, televisions, city infrastructure, sensor-equipped buses, trains or buildings, which makes the data flow so fast that within the two years prior to 2014 alone a zettabyte of data was accumulated. This belittles any prior record of human civilization (Shaw, 2014). However, it is not only the revolutionary amount of data that is collected, but also the fact that through the growing sophistication of statistical and computational methods that data can be used for innovative and powerful application domains (King, 2014). It therefore becomes obvious that more and more data handling skills are required (Data-Pop Alliance, 2015). At the same time, there is a lack of systematic definitions and procedures regarding data literacy (Süssenguth, 2015). The fact that data literacy is no binary phenomenon that can be distinguished between literate and illiterate causes general confusion and insecurities of what to do and expect from data (Data-Pop Alliance, 2015).

Figure 7 Problem Definition - Situation Analysis



Hence, frameworks that try to describe data literacy need to understand, appreciate and represent the multiplicity of it. One possible way to represent this, is by means of a maturity model. While data practitioners from sciences and business are more and more progressing in the world of data, social change organizations are still lagging behind when it comes to integrating data into their activities (Desouza and Smith, 2014). To close this gap, the data literacy maturity model will help to define and evaluate data handling and conversion capabilities, identify strengths and weaknesses and offer a guidance for future planning. Hence, the model is aimed at raising awareness on the topic of data literacy and provide a clear definition of competencies at different levels regarding several dimensions (see Figure 7).

Stakeholders of the model are nonprofit organizations that want to evaluate their data literacy capabilities or plan future actions in this field. A major stakeholder will be the Datenschule that is planning to use the model in their project work. The Datenschule (German for School of Data) is an educational program closely related to the topic of data literacy. Under the slogan “Do Good with Data”, the offer is especially targeted to non-profit organizations and conveys practical knowledge to support NGOs in realizing their societal goals. The project is based on a combination of workshops, strategy consulting and technical training – individually tailored to the specific requirements and practices of non-governmental organizations (Datenschule, 2016).

3.2 Comparison of Existing Maturity Models

Numerous maturity models have been derived from the generally acknowledged and recognized Quality Management Maturity Grid (QMMG) (Crosby, 1980) and the Capability Maturity Model (CMM), which has been developed for software development (Paulk et al., 1993). The origins of maturity thinking stem from quality management (QM) and Crosby’s pioneering book “Quality is free” (Crosby, 1980) in which he introduces the quality management maturity grid. The QMMG advocates a simple tool that evaluates the quality management spectrum of a company in a few phrases. It represents quality management in five stages and six dimensions (see Appendix C). According to Crosby (1980), companies evolve through five phases of maturity: uncertainty, awakening, enlightenment, wisdom and certainty. It describes six dimensions of QM: management understanding and attitude, quality organization status, problem handling, cost of quality as % of sales, quality improvement actions and summary of company quality procedure, in accordance to the five stages (Crosby, 1980). Based on the QMMG, companies can measure their quality management performance based on a score that is obtained by adding up the individual scores for every measurement category (Crosby, 1980). Stage one represents a score of one, Stage two gives a score of 2, Stage 3 depicts 3, and so on. Hence, companies can obtain a minimum score of 6 (equals all categories at the uncertainty level) and a maximum score of 30 (all categories are at certainty) (Crosby, 1980). The QMMG appreciates that organizations can have stronger capabilities in some dimensions and a weaker performance in others. Consequently, strengths and weaknesses can be easily identified which helps to plan future actions strategically (Crosby, 1980).

Many maturity models have been derived from Crosby's work, such as one of the best known spin-offs: the Capability Maturity Model (CMM) for software (Paulk et al., 1993). It is a model used to describe software process maturity and to which magnitude the specific process is precisely "defined, managed, measured, controlled and effective" (Paulk et al., 1993: 21). According to Paulk et al. (1993) increasing maturity with respect to a certain element can be understood as the formulation of processes through policies and organizational structures. Still, the CMM follows a different approach compared to the QMMG. It describes key process areas which all need to be completed to move on to the subsequent maturity level. The basic idea is to move from one maturity level to another through the completion of certain predefined standards. This is defined as a staged approach and leads to the precise classification into one maturity level which ranges from one to five. In Crosby's model, capabilities in the different dimensions can vary in their level of maturity and their interdependence is not that strict. This aspect proves that there are two possibilities to evaluate the maturity, based on a dimension holistic approach and one based on the individual dimensions. The levels of the CMM are classified as initial, repeatable, defined, managed and optimizing (see Appendix C) (Paulk et al., 1993). General descriptions are provided for each level and include common features (Paulk et al., 1993). The common features are: commitment to perform, ability to perform, activities performed, measurement & analysis and verifying implementation (Paulk et al., 1993). The common features particularize the key practices that can accomplish the aims of the key process area (dimension) (Paulk et al., 1993).

Practitioners tend to circumvent the complexity of CMM-based maturity models, by designing maturity models based on the Crosby grid approach, where performance of a number of key activities is described at different levels (Chiesa et al., 2008; McGrath and McGrath, 1996). Further, the grid approach as presented by Crosby, follows a more organization-neutral approach and thus can be applied to a greater variety of organizations. Although, maturity measurement grids tend to be less complex, they are still effective diagnostic and improvement tools (Maier et al., 2012) and will therefore create the underlying rationale of the data literacy maturity model. After a short introduction to the foundations of maturity models, further models will be reviewed and integrated in subsequent steps.

3.3 Determination of the Development Strategy

Becker et al. (2009: 218) suggest four basic development strategies for the design of a maturity model: “completely new model design, enhancement of an existing model, the combination of several models into a new one, and the transfer of structures or contents from existing models to new application domains”. This study combines different structures and contents from existing models to the new application domain of data literacy, which are illustrated in figure 8. Certain aspects will be put together from existing models to guarantee scientific validity and enrich existing standards through the collection of new data in the empirical part of this study.

Figure 8 Maturity Model Development Strategy - Combination of different Structures

Maturity Model Structure / Content	Based on
Level Names	Chall (1983), Crosby (1980), Hillson (1997)
Level Descriptions	Argyris and Schön (1978), Hillson (1997), Debnath et al. (2015), CobiT (2007)
Dimensions / Competences	Slater (2016)
Dimension Descriptions / Cell Text	Slater (2016)

3.4 Iterative Design of the Maturity Model

The iterative design will be continued in the empirical part of this study through expert interviews and the testing of the solution. This first part concentrates on defining different components of the maturity model through the insights from literature. Although various maturity models in different domains have been developed, Fraser et al. (2002: 4) identify the following shared properties of maturity models that need to be described:

- Several levels (usually 3-6): the top level describes the most mature state where the dimension subjects are systematically managed by the strive for improvement
- Descriptor for each level: e.g. initial / repeatable / defined / managed / optimizing
- Generic description or summary of the characteristics of each level as a whole
- Several dimensions or process areas: define related aspects / activities that describe the phenomenon to be evaluated
- Several elements or activities for each process area: summarize the states that must exist for the process area to be implemented in an effective way
- Description of each activity as it might be performed at each maturity level: describes the elements of the process area often regarding commitment to perform, ability to perform and the activities performed

The development of the model will be done in accordance with phase two of the procedure model as proposed by Maier et al. (2012). Consequently, the following steps need to be completed: a. select process areas, b. select maturity levels, c. formulate cell text, d. define administration mechanism (Maier et al., 2012).

a. Select Process Areas / Dimensions

For the purpose of this study, the definition by Slater (2016) will be adopted since it is one of the most recent studies on data literacy and based on previous studies (Bhargava and D’Ignazio, 2015). Also, it has been developed at the School of Data which is closely connected to the study object of this research and thus represents an adequate definition as a base for further insights. Adapted from the definition by Slater (2016) and the School of Data, the definition of data literacy as presented in figure 9 is used to describe the key dimensions in the preliminary DLMM:

Figure 9 Data Literacy Competences – Dimensions

Dimension	Competence	Description
1	**Data Sources** Know how to find data	track down sources of existing data, know how to collect data if it does not exist yet
2	**Data Thinking** Apply critical thinking skills to data	ability to do data quality assessment, contextualizing specific information to other aspects
3	**Data Inquiry** Ask questions to the data and find answers	ability to ask questions to data and ultimately find answers as one of the goals of data literacy trainings
4	**Data Outputs** Find specific outputs (stories or visualizations) in data	importance of finding stories and other journalistic outputs
5	**Data Objectives** Use it to advance one's goals	link between data and action was evident
6	**Data Culture** Feel comfortable around data and working with it	promoting comfort around data (and bringing down the psychological barriers that exist between people and data)
7	**Data Analysis** Do statistical analysis with data	ability to work with basic statistics

Based on Slater (2016)

b. Select Maturity Levels

The number of levels is to some extent arbitrary (Fraser et al., 2002). The original Crosby Quality Maturity Grid (Crosby, 1980) used five, as well as the software CMM (Paulk et al., 1993), the Risk Maturity Model (Hillson, 1997) describes four as well as the Innovation Audit by Chiesa et al. (2008). Nevertheless, based on the comparison of 24 maturity models by Maier et al. (2012: 141-144), we see that the levels are usually between four and five, which will serve as an orientation for this study.

According to Chall (1983) there are five levels of reading literacy that could represent the different competency levels of data literacy and are described as Acclimation, Early Competencies, Middle Competencies, Late Competencies and Proficiency. Referring again to Crosby's QMMG (1980), the levels can potentially also be described as Uncertainty, Awakening, Enlightenment, Wisdom and Certainty. Last but not least, Hillson (1997) proposes a four-level description for risk maturity that includes Naïve, Novice, Normalized and Natural as can be seen in Figure 10. Due to the growing complexity of maturity models with an increasing number of levels, the data literacy model will be limited to four levels.

Figure 10 Exemplary Level Descriptions

	Chall (1983)	Crosby (1980)	Hillson (1997)
Level 1	Acclimation	Uncertainty	Naive
Level 2	Early Competencies	Awakening	Novice
Level 3	~~Middle Competencies~~	Enlightenment	Normalized
Level 4	Late Competencies	~~Wisdom~~	Natural
Level 5	Proficiency	Certainty	

Based on Chall (1983), Crosby (1980) and Hillson (1997)

c. Formulate Cell Text

One of the major challenges when designing maturity models is the gradual description of the different levels in correspondence to the defined dimensions. In this step, levels of maturity are assigned to key aspects of performance or key activities. Thereby a series of cells that explain the gradual performance differences is created. With maturation, as primary subject matter, maturity models are required to define central constructs related to maturity and maturation (Becker et al., 2010).

To get more clarity in this regard, it is advisable to examine the difference between immaturity and maturity. For this, a theory developed by recognized Harvard scholar Argyris (1970) on an individual's maturation path within an organization is suggested in Figure 11 to better understand, which logic of maturation this study is based on.

Figure 11 Immaturity / Maturity Continuum

Immaturity / Maturity Continuum		
Immaturity		**Maturity**
Passive	-	Active
Dependence	-	Independence
Behave in few ways	-	Capable of behaving in many ways
Erratic shallow interests	-	Deep and strong interests
Short term perspective	-	Long term perspective
Subordinate position	-	Equal or subordinate position
Lack of awareness of self	-	Awareness and control of self

Based on Argyris (1970)

Based on this, an adaptation of the three learning types in organizations as described by Argyris and Schön (1978) is used in order to differentiate between maturity levels and to facilitate the formulation of the cell text. A preliminary level is added to describe no / very limited engagement with the topic so far, which is missing in Argyris' and Schön's considerations. Following this approach, the different levels can be differentiated as follows:

> *Level 1:* Describes situations, where the problem is not yet addressed and there is no engagement with the related activity. The connection between the factor and the overall topic of the MM is not reflected on.
>
> *Level 2: C*orresponds with the first type of learning suggested by Argyris and Schön and points out that some action is changed to correct a certain behavior. Other than that, tasks are carried out as usual.
>
> *Level 3:* Is based on Argyris' and Schön's second type of learning: people at this stage modify their actions, as well as think critically about existing norms, procedures, policies, and objectives that govern their actions. This means that the general situation is taken into consideration.
>
> *Level 4:* It corresponds to the third type of learning and signifies a stage, where an organization is aware of the influence of a given factor and continuously checks whether the way things are handled or set up is still appropriate for the given situation. At this stage, learning does not only happen when a mistake needs to be corrected: participants have the general mindset of continuously adjusting and improving the situation.

Taking a look at another general description, based on research by Hillson (1997: 37-38), who developed a risk maturity model, the general descriptions of the levels have been taken as an orientation for the data literacy maturity model in order to illustrate the overall rationale that characterize the different levels and also reflects Argyris' and Schön's (1978) considerations.

> *Level 1:* Organizations on Level 1 are unaware of the need for data literacy skills and have no or very vague understanding on what is required. There is no structured approach to dealing with data-related topics. Processes are reactive with no or little attempt to learn from the past or prepare for future data-driven possibilities.
>
> *Level 2:* Organizations are experimenting with the application of data-related topics, usually through a small number of nominated individuals, but has no structured generic processes in place. Although, they are aware of the

potential benefits of managing and using data, organizations have not yet effectively implemented data-related processes and are not gaining the benefits.

Level 3: Organizations have built data-driven activities into their routine business processes and implement data wherever it makes sense (on most or all projects). Generic procedures and standards on how to handle data are formalized and widespread, benefits are understood at all levels of the organization. Still benefits may not be consistently achieved in all cases.

Level 4: Organizations have established a data-informed culture throughout all levels. Data is actively used to improve business processes and gain competitive advantage. Data is used to manage opportunities as well as potential negative impacts.

Apart from the general level descriptions, the dimensions have to be defined accurately as well (Pöppelbuß and Röglinger, 2011). The data for the descriptions of the dimensions will be mainly collected in the expert interviews. However, for further orientation, we can consider certain elements to follow, as described by CobiT (Control Objectives for Information and Related Technology), which is an internationally recognized IT governance and control tool. To gradually describe the different dimensions on different levels, we can use the following aspects to provide guidance for formulations (IT Governance Institute 2007):

(1) awareness and communication

(2) policies, standards, and procedures

(3) tools and automation

(4) skills and expertise

(5) responsibility and accountability

(6) goal setting and measurement

As a further orientation for the formulation of the cell text, the underlying rationale of behaviorally anchored rating scales (BARS) will be used, which is suggested by Maier et al. (2012). It is a method that is often used in human resource management to obtain a precise description of different performance levels regarding specified tasks (Snell et al., 2016). BARS consist of a set of scales that present a major performance dimension of a particular task and are usually anchored by five or more critical incidences which reflect highly effective to highly ineffective observable behaviors relevant to the dimension under consideration (Snell et al., 2016). The scale values as well as the number of critical incidences anchored on a scale can vary, depending on the development procedure and appropriateness of the situation (Debnath et al., 2015). The major critical incidents for each dimension will be based on selected insights of the literature review and will be more clearly identified through expert interviews later on. For this study, a four-level scale is applied containing the following rationale:

(4) - Excellent performance

(3) - Fully competent performance

(2) - Marginal performance

(1) - Unsatisfactory performance

The preliminary cell text (see Figure 12) is based on the researcher's own formulations in accordance with the definitions of the selected data literacy description by Slater (2016).

Figure 12 Preliminary Data Literacy Maturity Model

BARS Logic (Debnath et al., 2015)	**Unsatisfactory Performance**	**Marginal Performance**	**Fully Competent Performance**	**Excellent Performance**
Level	**Level 1**	**Level 2**	**Level 3**	**Level 4**
Underlying Rationale based on learning types described by Argyris and Schön (1978)	Describes situations, where the problem is not yet addressed and there is no engagement with the related activity. The connection between the factor and the overall topic of the MM is not reflected on	Corresponds with the first type of learning suggested by Argyris and Schön, and points out that some action is changed to correct a certain behavior. Other than that, tasks are carried out as usual.	Is based on Argyris' and Schön's second type of learning: people at this stage, modify their actions, as well as thinking critically about existing norms, procedures, policies, and objectives that govern their actions. This means that the general situation is taken into consideration.	It corresponds to the third type of learning and signifies a stage, where an organization is aware of the influence of a given factor and continuously checks whether the way things are handled or setup is still appropriate for the given situation. At this stage, learning does not only happen when a mistake needs to be corrected: participants have the general mindset of continuously adjusting and improving the situation.

General Description based on Hillson (1997)		Organizations on Level 1 are unaware of the need for data literacy skills and have no or very vague understanding on what is required. There is no structured approach to dealing with data-related topics. Processes are reactive with no or little attempt to learn from the past or prepare for future data-driven possibilities.	Organizations are experimenting with the application of data-related topics, usually through a small number of nominated individuals, but has no structured generic processes in place. Although, they are aware of the potential benefits of managing and using data, organizations have not yet effectively implemented data-related processes and are not gaining the benefits.	Organizations have built data-driven activities into their routine business processes and implement data wherever it makes sense. Generic procedures and standards on how to handle data are formalized and widespread, benefits are understood at all levels of the organization. Still benefits may not be consistently achieved in all cases.	Organizations have established a data-driven culture throughout all levels. Data is actively used to improve business processes and gain competitive advantage. Data is used to manage opportunities as well as potential negative impacts.
Slater (2016)	**Data Thinking** Apply critical thinking skills to data	Critical evaluation of data does not exist. Data quality is not assessed and data evaluation criteria cannot be described.	Critical evaluation of data is vaguely defined. Data quality is not assessed consistently.	Critical evaluation of data is vaguely defined regarding certain aspects, data quality is assessed consistently.	Ability to do data quality assessment independently. Data evaluation criteria regarding authorship, method of obtaining and analyzing data, comparability and quality are precisely defined. The impact of data on science and society is internalized, so are copyright and licenses influencing data reuse. Ability of contextualizing specific information to other aspects exists throughout the organization.
	Data Inquiry Ask questions to the data and find answers	Lacking ability to formulate questions to find meaningful answers in data.	Questions can be asked to data in limited number of situations and answers questions partially.	Questions to data are formulated precisely and target-oriented to find meaningful answers in most of the cases.	Questions are formulated in accordance of multidimensionality of existing data. Answers to informational needs can be consistently found in data.
	Data Outputs Find specific outputs (stories or visualizations) in data	No awareness of the multiplicity of how data can be presented. No stories or visualizations can be drawn from data.	Limited ability to find specific outputs especially complex stories and visualizations. Focus on representing tables and a selection of index figures.	Basic stories and static visualizations can be created from data.	Ability to synthesize, represent and communicate the results of data analysis in ways suited to the nature of the data, their purpose and the audience. High awareness of the various forms in which data can be presented (written, numerical or graphic). Data Storytelling, static and interactive visualizations are commonly used.
	Data Objectives Use it to advance one's goals	No connection between organizational objectives and supporting function of data is established. Interpretation of data for specific needs and objectives does not exist.	Minor connections between organizational objectives limited to a small application area.	Data is understood as an enabler to achieve goals on a majority of topics (e.g. fundraising, campaign creation, process improvement, etc.).	Ability to interpret data in accordance to specific needs and objectives. Data is integrated when formulating objectives and planning for future actions. Link between data and action is evident.

	Data Culture Feel comfortable around data and working with it	Data is perceived as an ambiguous term which causes insecurities.	Data is perceived as an interesting concept and benefits are appreciated. Insecurities exist regarding use cases and what exactly to expect.	Data is not perceived as a source of insecurities, but rather understood as an enabler for progress and support for existing and planned activities.	Psychological barriers of data have been brought down (insecurities, fear, resignation, etc.) and comfort around data is promoted. Dedicated resources for data handling and data conversion exist.
	Data Analysis Do statistical analysis with data	No knowledge about how to prepare and analyze data. No resources are assigned for data handling and conversion.	Ability to work with basic descriptive and inferential statistics. Only a few resources are allocated to data handling.	Ability to work with advanced statistics (e.g. time series, factor analysis, tree-based models, clustering, etc.). Data handling and conversion skills are promoted throughout the organization.	Ability to apply descriptive, predictive and prescriptive statistics when necessary. Ability to prepare data for analysis, analyze it in keeping with the results sought and know how to use the necessary tools (data analysis tools both locally: Excel, R, SPSS, Stata, etc. and online). Ethical use of data, making sure that used methods are deployed and results interpreted transparently and honestly.

d. Define Administration Mechanism

The last step of Maier's development phase (2012) is to define the administration mechanism. This step includes the description of how the different dimensions will be assessed. In this regard, face-to-face interviews, workshops or surveys are possible methods (Maier et al., 2012). For the purpose of this study, the assessment will be executed through a self-evaluation excel tool for individuals that will refer to the maturity grid based on the results presented in a spider web diagram.

3.5 Conception of Transfer

Becker et al. (2009) propose to provide different forms of result transfer for the academic and user communities. The results of this study will be made available to the academic community through the publication of the master's thesis. The users will obtain the results of the maturity model development using the maturity grid as well as an excel tool for self-evaluation, that will be developed based on insights from the expert interview in a later step.

3.6 Implementation of Transfer Media

The sixth phase of the development procedure proposed by Becker et al. (2009) describes the implementation of the transfer media. This includes that the maturity model is made accessible for all targeted user groups. As described earlier in point five, this will be guaranteed through the publication of the investigations in the master's thesis, as well as the maturity grid and the excel tool. Nevertheless, the detailed implementation of the model is not within the focus of this study.

3.7 Evaluation

As suggested by Becker (2009), evaluation criteria need to be described. In order to evaluate the model, the following success criteria, based on Pöppelbuß and Röglinger (2011) are predefined: correctness, flexibility, usability, implementability, economic efficiency. The definition of evaluation criteria helps to find out whether the MM provides the presumed benefits and an improvement for the selected topic area (Pöppelbuß and Röglinger, 2011). After the evaluation, the model design process might have to be reiterated if the evaluation was not satisfying. The evaluation of the DLMM will be done through the cooperation with data experts and a NGO, to measure the set criteria as well as the relevance of the model. The evaluations will be incorporated in the model directly. Feedback from the testing with the NGO will be translated into recommendations for a future iteration, due to the limited time of this investigation.

3. 8 Rejection

One last potential step in the procedure model, as proposed by Becker et al. (2009), can be the rejection of the model, due to insufficient performance and relevance of the selected topic area. Whether the model is relevant enough for the planned application will be determined in subsequent rounds of assessment.

4 Critical Reflection on Maturity Models and the Procedure Model

Although, maturity models can provide meaningful insights for change initiatives in organizations, they also have been subject to criticism. A major point for criticism is that maturity models usually describe a step-by-step recipe, which reduces the phenomenon under investigation only to a fraction (de Bruin et al., 2005; McCormack et al., 2009). To circumvent this problem, the data literacy maturity model does not represent an exclusive solution, but rather option values that can be used as a first reference for orientation. This step is in accordance with Mettler and Rohner (2009) that suggest to design MM in a modifiable manner, due to the ever-changing internal and external attributes such as technological advancements, emerging scientific insights or the redefinition of the target audience, which would limit a standardized model's applicability. Apart from that, maturity approaches can ignore the possibility of multiple desirable paths (King and Teo, 1997). For this reason, King and Kraemer (1984), advice that maturity models should not only concentrate on describing a series of levels towards a final state, but also on drivers and barriers that influence evolution and change and appreciate different desirable states depending on specific organizational objectives and conditions, which will be reflected on during data collection. Further criticism is directed towards the growing amount of nearly identical maturity models, as well as the deficient indication of the development process which is often exchanged for a rather non-reflective adaptation of the CMM approach (Becker et al. 2009). Although, there is a growing group of researchers that address the gap of lacking design principles (Becker et al., 2009; Maier et al., 2012; Pöppelbuß and Röglinger, 2011)**,** through the development of the DLMM, more guidance regarding the formulation of cell text and in this regard especially how to effectively determine the different borders between levels would be helpful.

5 Methodology

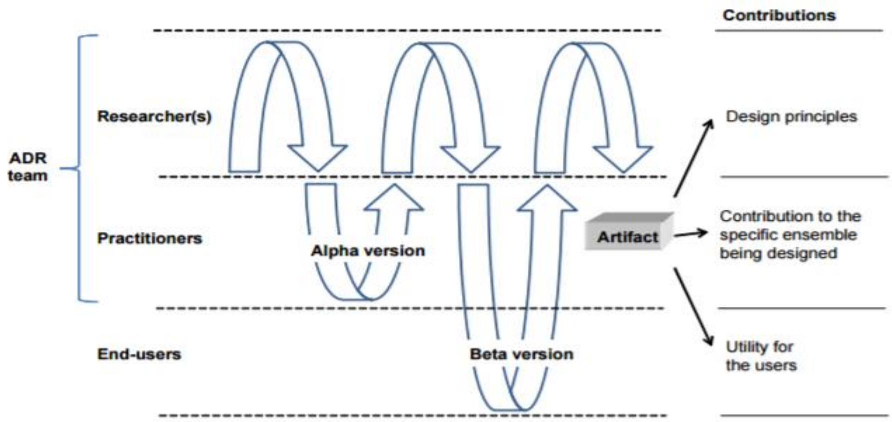
5.1 Preliminary Considerations

The aim of this study is to develop an artifact that helps to describe data literacy in a maturity model. For this reason, Hevner's design research (DR) approach (2004) is used as an underlying rationale of this research. This corresponds to the selected model development procedure as proposed by Becker et al. (2009). According to Hevner (2004) artifacts are a construct, a model, a method, or an instantiation, which can also refer to artifacts such as maturity models. Design research turns out to be an appropriate research method, since it seeks to create new artifacts that represent the intersection between people, organizations and technology (Hevner et al., 2004). Especially, the crucial intersection between these three components for this study, led to the decision, to combine design science with another research method, to further highlight the applied character of the designed artifact. Consequently, we combine design science and action research which is commonly referenced to as action design research (ADR) (Sein et al., 2011). DR is a method predominantly used in information systems research, which usually takes a more technological view and undervalues the influence of the organizational context on the artifact (Sein et al., 2011). Hence, technological accuracy is treated at the expense of application relevance and a shortened recognition of the effects on the interaction between the artifact and the organizational context (Sein et al., 2011: 39). Sein and collaborating researchers (2011) propose action design research as an adaptation of DR to address this issue. As a premise of this study, ADR helps to reflect the interwoven relationship between the artifact design and the shaping influence of the organizational application. The development of the maturity model reflects this through the use of a procedure model (Becker et al., 2009) that incorporates the design research principles as suggested by Hevner et al. (2004) as well as the testing of the artifact with an organization.

5.2 Data Collection

In order to answer the research question of how to design a maturity model for defining different levels of data literacy in NGOs, the collection of data is divided into three iteration phases and is based on an ADR schema by Sein et al. (2011) (see figure 13).

Figure 13 Generic Action Design Research Schema for IT-dominant Artifacts



Based on Sein et al. (2014)

Initially, a preliminary framework for data literacy maturity was developed by the researcher (Preparation Phase). This model incorporates various elements of data literacy that can be derived from relevant literature streams. After that, the preparatory model is reviewed by two data practitioners of the Datenschule in semi-structured expert interviews (interview guidelines are presented in Appendix D), for an initial exploration of a mutual understanding of the model as well as potential adaptations regarding their application context (Iteration Phase 1).

Second, the adapted alpha version (see figure 13) is enriched with the help of data practitioners especially from a more technical point of view to fill the cell text of the maturity grid as well as to identify level metrics to better distinguish between the different levels of the maturity model (see Appendix D). The level distinguishers are used to create the self-evaluation questionnaire (Iteration Phase 2). At this point, the data collection regarding the content and conception of the maturity model is finished. The third iteration phase is dedicated to the development of the beta version (see figure 13) and the model evaluation in an organizational context with the end users. In this phase, the model is reviewed by data practitioners of the Datenschule, and additionally evaluated by representatives of an organization through a testing of the questionnaire and the maturity grid (questionnaire and evaluation table can be seen in Appendix G). The predefined evaluation criteria (see Chapter 2) are: correctness, flexibility, usability, implementability and economic efficiency. Feedback from all phases is directly integrated to the new version of the artifact and brought into the next phase by the researcher. Data has been collected between December 2016 and January 2017.

Since to date, research on data literacy is limited in general and particularly when it comes to evaluating capability maturity, this research focuses on discovering new theory. Therefore exploratory qualitative experts interviews were applied throughout the data collection process, since it can result in the discovery of generalizations and the understanding of phenomena, which have only received limited scientific attention so far (Stebbins, 2001). The description of critical incidents in the interviews (critical incident technique – CIT), is a well-recognized qualitative research approach that offers an approach for collecting and analyzing information (Hughes et al., 2007). The critical incident technique helps to collect contextualized data that represents real-world insights based on experiences and observations of the research participant. This helps to gain insights that are closer to the real application of the model in the future (Hughes et al., 2007). As this study aims at forming a first investigation in developing a data literacy maturity model that can be used in practice, the practicability aspect of the CIT as well as the fact that it provides a knowledge base for future research makes the CIT a valuable research method for this study (Hughes et al., 2007).

5.3 Sampling Frame

The data collection is based on the study object of the Datenschule (German for School of Data), a project of the Open Knowledge Foundation Germany which is based in Berlin. The Datenschule is an educational program closely related to the topic of data handling skills. This educational endeavor covers a broad range from finding data, analyzing and visualizing it, to the development of data-driven campaigns. Under the slogan “Do Good with Data”, the offer is especially targeted at non-profit organizations and conveys practical knowledge to support NGOs in realizing their societal goals, which makes it a relevant study object for this specific research. The project is based on a combination of workshops, strategy consulting and technical training – individually tailored to the specific requirements and practices of NGOs (Datenschule, 2016). Investigating this specific case is an attractive method, since it explores a contemporary phenomenon in a real-world context and thus helps to derive recommendations for data handling practices for NGOs and forms a basis for future research (Yin, 2014).

Figure 14 Sample Overview

Data Collection	Interviewpartner	Organization	Position	Purpose	Date
Expert Group 1	Interviewpartner 1	Datenschule	Project Manager	Model Design	28.11.2016
Expert Group 1	Interviewpartner 2	Datenschule	Workshop Designer	Model Design	28.11.2016
Expert Interview 1	Interviewpartner 3	Data & Society Research Institute, Engine Room	Fellow & Research Lead	Model Design	05.12.2016
Expert Interview 2	Interviewpartner 4	Datenschule	Developer	Model Design	08.12.2016
Expert Interview 3	Interviewpartner 5	HWR Berlin	Data Scientist	Model Design	23.12.2016
Expert Group 2	Interviewpartner 6 (1)	Datenschule	Project Manager	Evaluation	02.01.2017
Expert Group 2	Interviewpartner 7 (2)	Datenschule	Workshop Designer	Evaluation	02.01.2017
Solution Testing	Interviewpartner 8	NGO: Education	Project Collaborator	Evaluation	05.01.2017
Solution Testing	Interviewpartner 9	NGO: Education	Project Lead	Evaluation	05.01.2017

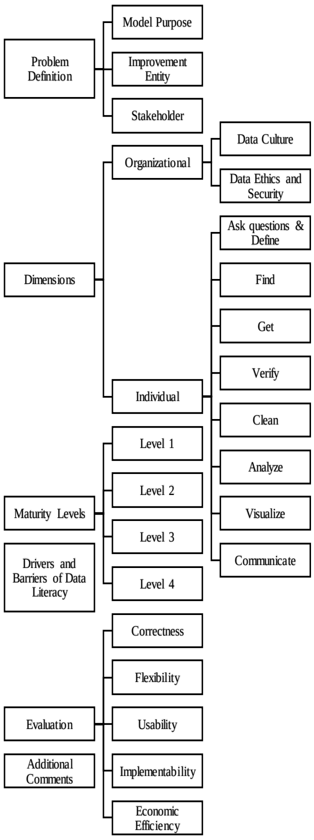
The iterative artifact design process of action design research depicts that the sample strategy is based on purposeful sampling. Consequently, the researcher actively seeks the most effective sample to answer the research question (Marshall, 1996). The selection criteria for the chosen sample are individuals, who are active/interested in data-related activities in non-governmental organizations. More precisely, the research is examining study subjects who have specific experiences (critical case sample) for instance, the data practitioners of the Datenschule as well as the organization to evaluate the model. Additionally, subjects with special expertise on data literacy (key informant sample) such as the interviews with general/technical data literacy experts to enrich the model are investigated. Moreover, action design research makes it necessary to decide for a sample that represents the interconnection between people, organization and technology (Hevner et al., 2004; Sein et al., 2011), which is expressed in the selected sample (see Figure 14). The evaluation of the model is investigated with the help of the same experts as in the first iteration phase to be able to have a clear comparison on the advancements of the model. In a final step, the model is used together with a project partner of the Datenschule to gain insights from a real-world application scenario and to include an organizational perspective in the model development. Especially the recommendations of the representatives of the Datenschule help to contact useful candidates for further investigation (snowball sample), which is essential to the iterative model design (Marshall, 1996).

Since the evaluation of the artifact design includes a subjective, qualitative proportion, the interpretation of the participants has an influence on the quality and content of the data collected. To limit the subjective influence, a variety of sampling perspectives is used. The required sample size to reach data saturation can only be determined as the study progresses and new categories or explanations terminate to stem from the collected data (Marshall, 1996). Due to time constraints of this study, the number of interview partners is limited to nine. This can of course limit the theory developed during this study, but also complies with the purpose of this study to represent a first scientific exploration of the possible design of a data literacy maturity model.

5.4 Data Analysis

After having clarified the research methods and guiding principles of the sample selection, the next step is to define the underlying approaches of the data analysis. This study analyzes the collected data in conformity with Meuser and Nagel (2009) who established a strategy of qualitative content analysis for expert interviews to determine central, common and typical phenomena as well as contradictions of a certain aspect. This procedure helps to minimize the dataset and allows for integrating and comparing the theoretical foundations with the information collected during the interviews. The first step after conducting the interviews is to transcribe the audio files. After that, the data material is screened and codes created deductively in accordance with the insights from the maturity model development of the literature review (Problem Definition, Dimensions, Maturity Levels, Drivers and Barriers of Data Literacy, Evaluation) as well as inductively (Additional Comments) in order not to neglect aspects that were not considered prior to the interviews (see Figure 15).

Figure 15 Code System



After the identification of meaningful codings for each category (code), relevant paragraphs have been paraphrased in accordance with the respective codes / topics. The codes and corresponding codings were documented and recorded with the help of MAXQDA, a software for qualitative data analysis. In accordance with the logic of the CIT the codes as well as the codings represent the clusters of incidents. The next step of the data analysis includes a thematic comparison between the different interviews to identify similarities and contradictions as well as to further refine the formulation of the cell text and the identification of level metrics. The codes determined in the empirical study are compared to the dimensions of the maturity model developed in the literature review and extended or rejected accordingly. On this basis, all relevant components of existing literature and the elements perceived by data practitioners are combined in an adjusted framework. This step helps to formulate generalizations for a data literacy maturity model by setting the insights back into the theoretical context. The analysis of the maturity model evaluation as a second part of the empirical study is based on the testing of the assessment tool and the maturity grid, together with a short feedback form.

6 Analysis and Findings

After the preliminary model synthesis based on relevant literature streams and the data collection, the analysis is divided into three iteration phases in accordance with Sein et al. (2011). The analysis is based on a qualitative content analysis by detecting patterns, repetitions and diction interpretation of the arguments proposed by the interviewees. Most interviews and testings have been executed in German and translated to English afterwards. The final synthesis of the beta-version of the DLMM will be presented in the discussion and is not part of this chapter.

6.1 Iteration Phase 1 – Creating a Common Understanding

The objectives of the first artifact iteration phase was the creation of a common understanding and shaping of the model based on the categories proposed in Becker's maturity model procedure (2009) together with experts from the Datenschule.

Audience and Stakeholders

Based on the assumptions of the preliminary model synthesis in Chapter 2, the practitioners of the Datenschule agreed that the DLMM can be generally helpful for NGOs. They further specified, that these NGOs usually already have a basic understanding of the potential of data (112816_Transcript_Expert1_2, n4) (for referencing see Appendix G). They further explain that this does not necessarily mean that they know how to work with it, but that they already discovered the potential of integrating data to their activities and specific topics to a certain extent (112816_Transcript_Expert1_2, n4, n5). At the same time, it cannot be used for anyone, but only for people that have this basic understanding of data as described before and is rather relevant for organizations that already work digitally and not analogue anymore (112816_Transcript_Expert1_2, n6, n9). This means that they already somehow produce, publish and communicate digital content and want to use data for political and social topics (112816_Transcript_Expert1_2, n7, n8). These NGOs are further interested in working with facts, evidence and numbers to prove something, mostly regarding “fact-checking” and monitoring (112816_Transcript_Expert1_2, n4, n8).

In addition, the experts believe that the model can potentially be transferred to a lot of other application domains and organizations and are not only limited to social change organizations (112816_Transcript_Expert1_2, n1, n2, n3).

Improvement Entity

Having a better understanding of the stakeholders of the DLMM, the improvement entity of the maturity model has to be described. The interviewees state that it is “super important” to create internal data competencies (112816_Transcript_Expert1_2, n12). They therefore strongly believe and repeatedly mentioned that it is crucial to start with one person in an organization, create a strong foundation and spread the knowledge from this point on, from bottom to top on a continuous basis (112816_Transcript_Expert1_2, n13, n14, n15, n16, n17, n18, n19, n23, n24). Additionally, they see the advantage of a facilitated evaluation of data skills and better future competence building on an individual basis (112816_Transcript_Expert1_2, n19). Although the Datenschule works together with teams, their educational endeavors aim at the development of individual competencies, and thus regard the DLMM helpful at this level to derive implications to a broader, organizational context (112816_Transcript_Expert1_2, n22). Interview partner 1 mentioned, that if there is this one person that has some data knowledge, he / she serves as a central touch point for other teams and individuals. Hence, this knowledge can be transferred to different projects and topics within the organization (112816_Transcript_Expert1_2, n15). As a consequence, certain workflows can be established and differentiated data competencies can be built (112816_Transcript_Expert1_2, n15).

Purpose of the Model

After defining the improvement entity, the purpose of the model regarding raising awareness or best practice benchmarking (see Chapter 2, Table 2) was identified.

The experts agreed that using data in their usual activities is a new topic for many NGOs that needs to be tackled from a small starting point (112816_Transcript_Expert1_2, n26, n27). They therefore believe that it makes sense to start evaluating one organization first, define the criteria for the different levels more precisely and elaborate from that point on to other organizations (112816_Transcript_Expert1_2, n26, n26). For them it is more important to start raising awareness on the topic at the beginning, to define a shared understanding and work further on from there (112816_Transcript_Expert1_2, n29, n30, n31). At the same time, they discussed that defining the purpose of the model depends a lot on the country that the organization is operating in, since there are big differences in the advancement of data topics locally (112816_Transcript_Expert1_2, n25). A final thought for the future of the model was to compare organizations that already use data to a major extent with learning organizations and adapt the model to these specific situations to formulate benchmarking standards (112816_Transcript_Expert1_2, n28).

Dimensions

The definition of the model purpose is followed by further development of the model on a content-level. The first step in this regard was to determine whether the dimensions of the preliminary model were meaningful to the data practitioners.

The preliminary maturity grid was provided to the data practitioners to evaluate the appropriateness of the descriptions and elaborate on the insights from literature. Overall, the interviewees agreed that "everything" that is listed in the cell descriptions fits very well to the respective level and dimension (112816_Transcript_Expert1_2, n32, n126). They also appreciated that the formulations were rather general in order to be able to define the skills / metrics more precisely in a later step to use the model in a variety of organizations (112816_Transcript_Expert1_2, n32, n126).

Nevertheless, they mentioned that the order of the different competencies needs further refinement, especially regarding their logical structure. They therefore propose to follow the single steps of the data pipeline as an orientation to describe the dimensions of the

DLMM (112816_Transcript_Expert1_2, n33, n34, n35, n36). Hence, matching the competencies of the preliminary model with the single steps of the data pipeline, which the Datenschule also uses during project work, helped to create a more structured definition of the dimensions for the data literacy model in close consideration of the real-world application context (112816_Transcript_Expert1_2, n36, n37, n38). After matching the steps of the data pipeline with the initial dimensions, figure 16 describes the new dimensions (112816_Transcript_Expert1_2, n38, n39, n42, n43, n44, n45, n48, n49, n50, n51, n52 n63, n64, n78, n81, n86):

Figure 16 Dimension Evolution - Iteration Phase 1

Updated Dimension Description	Previous Dimension Description
1 Ask question / Define	Data Inquiry / Data Objectives + feeling of which questions can be answered by data
2 Find	Data Sources
3 Get	TBD – new Dimension
4 Verify	Data Thinking + argue / debate through data
5 Clean	TBD – new Dimension
6 Analyze	Data Analysis
7 Visualize	Data Output
8 Communicate	Data Output – not only limited to journalistic output

The interviewees considered using the official data pipeline as dimension indicators advantageous, because it is a framework that can be easily adapted to every project that uses data somehow and that certain skills can easily be determined for each of the 8 steps (112816_Transcript_Expert1_2, n41). In the initial maturity grid, data outputs described both, visualization and communication / data storytelling, in one. However, the data practitioners repeatedly mentioned that for them, the separation between visualizations, the numeric/graphic representation of the data and communicating complements the graphics with a story and conveys a real meaning, need to be separated (112816_Transcript_Expert1_2, n40, n47, n53, n54, n55, n56, n57, n58, n59, n60, n61, n62). One aspect that the experts were unsure about and that was not yet mentioned in the model, is the ability to understand and read basic code as a competence for data literacy (112816_Transcript_Expert1_2, n79). They were not sure if this “algorithm – awareness” is somehow already anticipated in some of the steps such as find, get and analyze (112816_Transcript_Expert1_2, n79, n83, n135). In addition to that, they also agreed that understanding code can be a good indicator for the different competence levels when it comes to differentiating between a “data freshman” and more advanced individuals

(112816_Transcript_Expert1_2, n80). This fact led to the decision to not include algorithm-awareness as a separate dimension in this model.

With regards to the completeness of the model, one interesting thought was discovered. The interviewees doubt that a model like this can be complete, since this highly depends on the organizational context. They propose to offer some “instrumental variables” for each level and dimension as an orientation point, but no absolute rule that needs to be completed under any circumstances (112816_Transcript_Expert1_2, n128).

Apart from the dimensions that are included in the data pipeline that describe individual competencies, the experts suggested to include certain aspects outside the data pipeline on a more organizational level (112816_Transcript_Expert1_2, n66, n67, n70, n72, n73, n74, n75). Hence, the model is extended by two more additional dimensions and now entails ten dimensions in total, both on an organizational and an individual level. One of the organizational elements of the maturity model is Data Culture, which essentially describes that there is a basic knowledge of what data is and that there is no fear to work with data or get in touch with the topic (112816_Transcript_Expert1_2, n69, n71). In addition, considerations regarding an organizational knowledge and awareness on how data is used within the organization can be added to this dimension (112816_Transcript_Expert1_2, n76, n77). The dimension of data culture might also be a trend indicator for the future to find out whether people feel more comfortable working with data if they score higher in all the other dimensions (112816_Transcript_Expert1_2, n68). The experts further suggested a second dimension on an organizational level: Data Ethics and Security (112816_Transcript_Expert1_2, n87, n88, n89, n93). This point appeared meaningful to the experts, since they still doubt that people have an awareness of what a responsible use of data means (112816_Transcript_Expert1_2, n89). For them, this point describes the responsible use of data and a certain kind of “limited data collection” (112816_Transcript_Expert1_2, n90, n91, n92). They mention the example of collecting user data to create personal profiles and the fact that often organizations do not even question, whether they really need the data or not (112816_Transcript_Expert1_2, n90). The experts call for a reflection on when and under which circumstances data is collected, to define for which purpose it is collected / used, and to formulate organizational standards regarding an ethical and responsible use of data (112816_Transcript_Expert1_2, n90, n94, n95).

Maturity Levels

Moving away from the definition of the dimensions to the description of the maturity levels, the experts mention that there is a clear tendency regarding the maturity levels that initial levels rather obtain reading skills and step by step move forward to writing skills (112816_Transcript_Expert1_2, n96, n125, n126). In other words, this describes whether a lot of manual tasks are included when working with data or if more automated solutions are applied (112816_Transcript_Expert1_2, n154, n155, n156). This description also helps to distinguish between the four different levels. Regarding the naming of the different levels, the experts agreed on the following naming:

Level 1 – Uncertainty (112816_Transcript_Expert1_2, n108, n120)

Level 2 – Enlightenment (112816_Transcript_Expert1_2, n109, n110, n111, n112, n113, n114, n121, n122)

Level 3 – Certainty (112816_Transcript_Expert1_2, n115, n116, n123)

Level 4 – Data Fluency (112816_Transcript_Expert1_2, n117, n124, n212, n213)

In addition to that, the experts were unsure about how effectively the distinction between the different levels can be determined and what effective measures / skills are needed to move from one level to the next (112816_Transcript_Expert1_2, n104, 112816_Transcript_Expert1_2, n105, n106, n107, n118,). To address this concern, the subsequent expert interviews have been conducted to include a more technical aspect and to be able to better distinguish between the levels.

Evaluation of the Preliminary DLMM

An initial evaluation of the preliminary DLMM was included in the first interview to see if the evaluation criteria make sense and to better identify elements of improvement for the next iteration phase. In addition, Becker et al. (2009) propose an iterative evaluation throughout the development process.

Generally speaking, the evaluation criteria as proposed by Pöppelbuß and Röglinger (2011), led to a good evaluation overview of the DLMM and will be considered further for the final evaluation of the model. As can be seen in Figure 17, the overall evaluation of the preliminary DLMM was positive (112816_Transcript_Expert1_2, n197). Only minor comments regarding the descriptions of concrete skill requirements for each level and dimensions were mentioned. To meet this request, additional expert interviews, especially regarding technical requirements of the different competencies are conducted in the subsequent iteration phase.

Figure 17 Evaluation Overview Preliminary DLMM

Evaluation Criterion	Comment	Source
Correctness	Cell descriptions are appropriate, just need minor adaptations regarding more concrete skills without the demand to be complete	112816_Transcript_Expert1_2, n180, n182
Flexibility	Is high, can be easily adapted to other organizations and sizes of organizations	112816_Transcript_Expert1_2, n183, n184
Usability	Is given, especially for projects like the Datenschule or other technical consultancies to convey data literacy knowledge, helpful for external and internal assessments	112816_Transcript_Expert1_2, n186, n187, n188
Implementability	Self-assessment questionnaire: should not take too much time, if preliminary model is extended by precise skill requirements that are included in the self-assessment questionnaire, it will be easily implementable and benefit for users is clear	112816_Transcript_Expert1_2, n189, n190, n191, n192
Economic Efficiency	Low-level investment (self-evaluation should be around ~ 15 mins and openly accessible)	112816_Transcript_Expert1_2, n196

Additional Comments

Additional comments were collected throughout the interview to guarantee that new insights can be reflected on, which could not be referred to one of the codes. The experts agreed, that the model also needs to offer room for changes over time, especially when it comes to technological, organizational or social changes in the data eco-system (112816_Transcript_Expert1_2, n104, n195). Additionally, the experts mentioned that the suggested skill requirements / tool descriptions for the different levels and dimensions can probably not be universally valid, due to differing organizational resources and contexts (112816_Transcript_Expert1_2, n119, n141, n142, n144). They further highlight, that diverse teams are needed to complete successful data projects and that individual experts for specific topics rather than data-allrounder will be common in the future. Hence, there will be specialists for data content, designers, communicators, analysts etc. (112816_Transcript_Expert1_2, n216n n217). Apart from that, it is not only important that organizations can handle and use data appropriately, but also that there is an audience for data products who can interpret / understand them accordingly (112816_Transcript_Expert1_2, n214, n215).

6.2 Iteration Phase 2 – Creation of the Alpha Version

After the insights of the first development phase, it became obvious that more data had to be collected in order to better describe the different levels and dimensions. Therefore, another three expert interviews were conducted to get a better picture of the requirements for each level and corresponding dimension. All insights of the different levels and dimensions from the preliminary framework and the interviews have been incorporated in one data literacy maturity grid (see Appendix F) to get a better overview and compare the insights, to formulate the data literacy maturity questionnaire and to refine the cell descriptions of the maturity grid. The development of the data literacy questionnaire and the cell descriptions have been identified through detecting patterns, repetitions and diction interpretation of the arguments proposed by the experts. This helped to create an alpha version of the questionnaire (see Appendix G) that determines the different competence levels of the data literacy dimensions as represented in the data literacy maturity grid (see Figure 18) in a self-assessment excel tool through a synthesis of the different insights from the expert interviews.

Criticism of the Preliminary DLMM

Especially expert 3 had problems with filling out the maturity grid. This was due to the following aspects:

- A core aspect of data literacy is missing: ability to critically assess data (outputs), (120516_Transcript_Expert3, n199, n203, n208) even without a technical analysis (120516_Transcript_Expert3, n202).
- It is not necessary to be familiar with all competencies (find, get, verify, clean, analyze, visualize, communicate) to have data literacy. There is the possibility to step in at any point of the pipeline (120516_Transcript_Expert3, n209).
- The evaluation of different competence levels depends highly on the context / what you are trying to do with data. For small NGOs, excellent performance can mean something completely different than for bigger, more experienced organizations (120516_Transcript_Expert3, n210, n211, n218, n219, n220, n235).

- Problems often come from fitting people into a model, which is not for them and which does not make any sense for their context (120516_Transcript_Expert3, n227, n236, n222, n223, n224, n225, n226).

Expert 3 proposes that an alternative model should allow users to critically asses what their needs are versus what they can do (120516_Transcript_Expert3, n228), prioritize the context, and identify the skills that are needed to reach these goals (120516_Transcript_Expert3, n229, n237).

Figure 18 Data Literacy Maturity Grid - Alpha Version

	Underlying Rationale	Read →			Write
	Generic Level Descriptors	Uncertainty	Enlightenment	Certainty	Data Fluency
	General Level Descriptions	Organizations are unaware of the need for data literacy skills and have no or very vague understanding on what is required. Individuals might have a certain interest in data and work digitally, but are unsure about the different steps that exist when working with data.	Organizations are experimenting with the application of data-related topics. Describes a state where a lot about data has already been understood theoretically, but cannot be applied in many cases and has to be further trained.	Organizations perform data handling steps with confidence and have built data-driven activities into their routine processes wherever it makes sense. Generic procedures and standards on how to handle data are formalized and widespread, benefits are understood at all levels of the organization.	Organizations have established a data-informed culture throughout all levels. Data is actively used to improve processes and create workflows.
Organizational	Data Culture	Data is perceived as an ambiguous term which causes insecurities.	Data is perceived as an interesting concept and benefits are appreciated. Insecurities exist regarding use cases and what exactly to expect.	Data is not perceived as a source of insecurities, but rather understood as an enabler for progress and support for existing and planned activities. Higher management and leaders support data initiatives.	Psychological barriers of data have been brought down (insecurities, fear, resignation, etc.) and comfort around data is promoted. Higher management and project managers understand and support importance of dedicated resources (time, budget, human resources) for data handling and conversion.
	Data Ethics & Security	No awareness for guidelines that ensure confidentiality, integrity and availability of data.	Rising awareness and uncoordinated attempts to promote the importance of the responsible use of data. No defined guidelines.	Awareness of the impacts of data use. Guidelines for responsible data handling are defined and incorporated internally to activtities.	Processes are in place to ensure confidentiality, integrity, and availability of data. Only data that is necessary/relevant is collected/used. Consistent, company wide policies for secure and ethically sound data handling are constantly redefined and updated.
Individual	Ask question / Define	Lacking ability to formulate questions in order to find meaningful answers in data. No feeling about which questions can be answered by data.	Questions can be asked to data in limited number of situations and answers are provided through simple queries.	Questions to data are formulated precisely and target-oriented in order to find meaningful answers in most of the cases.	Entire projects are based on multidimensional questions. Answers to informational needs can be found consistently in data, because of the high awareness of what questions can be answered by data (no overinterpretation).
	Find	Limited understanding of possible data sources. Use of basic search engines to find data. No experience for identifying and selecting most relevant data sources.	Knowledge only limited to a few data sources. Advanced use of search engines and data requests at public institutions are common practice.	Broad understanding of different data sources, most relevant ones can be chosen from a selection of data sources. Awareness and use of data portals for specific topics.	Profound understanding of the various possible types of data sources. Assessment criteria for selecting the ones most relevant to an informational need are formulated. Ability to detect when a given problem or need cannot be solved with the existing data, and are familiar with research techniques to obtain new data (e.g. complex queries).
	Get	Data is derived from full text and used as base for further processing.	Use of downloads and data formats such as .csv .	Data can be accessed using more complex data formats (e.g. JSON, XML). Use of APIs to get data.	Access to data through sophisticated methods (e.g. automated data scrapers / scripts). Ability to convert input format into a form that can be used for further processing and analysis.
	Verify	Critical evaluation of data does not exist, data is taken at face value. Data evaluation criteria cannot be described.	Critical check of simple data quality measures.	Multiple layers of data checking are implemented in standard procedures across the organization.	Ability to do data quality assessment independently. Data evaluation criteria regarding authorship, method of obtaining and analyzing data, comparability and quality are precisely defined.
	Clean	No awareness that given data might have to be checked, cleaned or normalized. Data is further processed as is.	Awareness that given data most often is not perfect. Awareness of some data quality criteria (e.g. empty fields, duplicates) and manual fixing of errors.	Invalid records can be detected and are removed using programs that support data cleaning (e.g. OpenRefine). High awareness of data quality criteria (e.g. machine processable, empty fields, duplicate detection).	Independent ability to remove invalid records and translating all the columns to use a sane set of values through automated script. Ability to combine different datasets into a single table, remove duplicate entries or apply any number of other normalizations.
	Analyze	Bar and pie charts, simple use of data tables and basic summaries of data.	Ability to work with basic descriptive statistics. Pivot tables for aggregating information, histograms and boxplots.	Ability to work with advanced statistics (e.g. inferential view of data, linear regression, decision trees).	Full suite of machine learning tools (e.g. clustering, forecasting, boosting, ensemble learning).
	Visualize	No awareness of the multiplicity of how data can be presented. No understanding of when standard visualizations are chosen, decision based on what looks best (trial and error).	Ability to find specific outputs in accorance with information that want to be represented (e.g. in Excel).	Creation of interactive charts / dashboards, uncertainties are always visualized along with the data.	High awareness of the various forms in which data can be presented (written, numerical or graphic). Sophisticated visualizations are programed, linked, dynamic dashboards that anticipate user requests are designed.
	Communicate	Insights from data are not communicated or put into a broader context.	Limited ability to find specific outputs. Simple narrative support static visualizations.	Own projects are supported by interactive visualizations and more sophisticated narrative in a broader context. (e.g. data storytelling, conferences, talks, monthly updates, blog posts).	Ability to synthesize and communicate in ways suited to the nature of the data, their purpose and the audience (e.g data storytelling, data-driven campaigning, workshops, conferences, monthly updates, blog posts, reproducible research).

6.3 Iteration Phase 3 – Evaluation of the Alpha Version

In the last step of this study, the adapted alpha version, containing the maturity grid and an excel self-assessment tool, (see Figure 18 and Appendix G) were tested. The testings were executed together with experts from the Datenschule, as well as employees of a non-profit organization that is currently cooperating with the Datenschule, to include an organizational perspective in the model development. The cooperating NGO is active in the area of sustainable education programs at schools and is currently working on transferring school data in Germany into meaningful knowledge and stories. The evaluation included a testing of the rough self-assessment questionnaire, the representation of the results, as well as a brief feedback interrogation.

The experts evaluated the model positively. Minor changes and aspects to review can be seen in Figure 19. Especially the logical order as well as the differentiation between the different levels with the help of the self-assessment questionnaire were mentioned positively by the experts (010217_Evaluation_Expert6_7, n362, n363, n364). One point for criticism was, that some formulations on lower levels might be too negative and should be rephrased (010217_Evaluation_Expert6_7, n365). In a later version of the model, the experts further suggest to mix the answers and to avoid biased evaluations due to anticipated score distributions (1-4 for answer 1-4) (010217_Evaluation_Expert6_7, n366).

Figure 19 Testing Results Experts of the Datenschule

Evaluation Criterion	Comment	Source
Correctness	Logical order and most important aspects are covered, differentiation between different levels is very well formulated	010217_Evaluation_Expert6_7, n362, n363
Flexibility	Very good and easy to understand, especially because of tool descriptions	010217_Evaluation_Expert6_7, n364
Usability	Good, maybe pay a little more attention that formulations do not sound too negative especially in the lower levels, answers could be mixed so that level evaluation cannot be anticipated by users	010217_Evaluation_Expert6_7, n365, n366
Implementability	Can be easily used, maybe easier for the organizations if translated to German, great as a tool for a beginning partnership of the Datenschule	010217_Evaluation_Expert6_7, n367, n368
Economic Efficiency	Easy to complete, would be interesting to know how long it approximately takes to fill out the self-assessment, nice if integrated in google forms so it can be send out easily	010217_Evaluation_Expert6_7, n369

The testing of the artifact in an organizational context together with two representatives of a partnering organization of the Datenschule yielded slightly different results. Although the model was positively evaluated regarding Implementability and Economic Efficiency as can be seen in Figure 20, more criticism rose regarding the model correctness, flexibility and to some extent usability. The correctness of the model was difficult to evaluate, since the representatives are no data experts and therefore were highly insecure about this aspect. On the other hand, they mentioned that a model like this might not be complete at any point, due to the high complexity of the topic in general and the specific use cases in particular (010517_Evaluation_IP8, n379, 010517_Evaluation_IP9, n394, n395). Flexibility has been critically regarded, since the formulations of the possible answers in the self-assessment tool often appeared too complex and might be difficult to understand for beginners (010517_Evaluation_IP8, n376). Regarding the usability, the representatives interestingly had different opinions. Interview partner 8 mentioned that the usability is high, since the assessment tool was easy to use (010517_Evaluation_IP8, n377). Interview partner 9 on the other hand, felt that the usability is rather low due to the long text answers in the questionnaire in some areas (esp. Data Culture) (010517_Evaluation_IP9, n399, n402).

Figure 20 Testing Results Organizational Perspective

Evaluation Criterion	Comment	Source
Correctness	Neutral / high: difficult to evaluate, can never be complete, correctness difficult since no experts, very different use cases, hard to represent complexity	010517_Evaluation_IP8, n379, 010517_Evaluation_IP9, n394, n395
Flexibility	Neutral: improvements regarding the formulation of the questions / answers and their complexity, might be too difficult for beginners, flexible regarding technological changes, unsure evaluation since no data expert especially regarding technological changes	010517_Evaluation_IP8, n376, n380, 010517_Evaluation_IP9, n398
Usability	High: easy to use, low: text blocks in questionnaire are too big, hard to answer consistently, possible answers have to have finer granularity	010517_Evaluation_IP8, n377, 010517_Evaluation_IP9, n399, n402
Implementability	Very high: depends on where organization stands at, at beginning: might be overwhelmed, more advanced organizations: helpful to identify position, shows strengths and weaknesses which is very good	010517_Evaluation_IP8, n382, 010517_Evaluation_IP9, n400, n401
Economic Efficiency	Very high: only took ca. 15 mins, openly accessible	010517_Evaluation_IP8, n378 010517_Evaluation_IP9, n402

The testing of the solution with an organization, yielded the results as illustrated in figure 21 and 22.

Figure 21 Evaluation Results - Testing Iteration Phase 3 - Spider-web Diagram

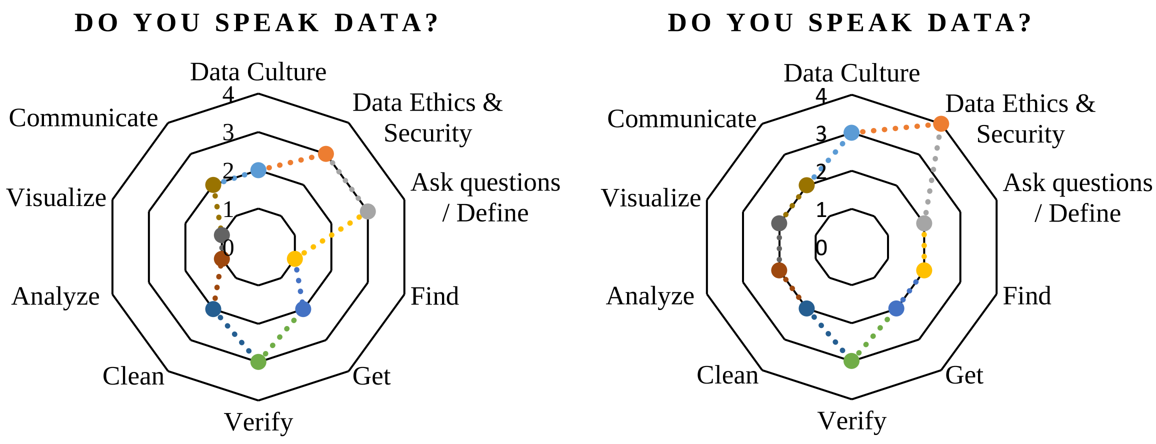


Figure 22 Evaluation Results - Solution Testing - Iteration Phase 3

Dimension	Individual 1 Level - Result	Individual 2 Level - Result	Organization Overall Average
Data Culture	2	3	3
Data Ethics & Security	3	4	4
Ask questions / Define	3	2	3
Find	1	2	2
Get	2	2	2
Verify	3	3	3
Clean	2	2	2
Analyze	1	2	2
Visualize	1	2	2
Communicate	2	2	2
Average	**2**	**2**	**2**

*Please note: Averages are rounded to full numbers for a better interpretation in the model

Both participants scored a data literacy level of two, which represents a total score of two for the organization as a whole as well. Overall, this means that the organization is experimenting with the application of data-related topics. This phase describes a state where a lot about data has already been understood theoretically, but cannot be applied in many cases and has to be further trained (see Figure 18 for a more detailed description). To get valuable insights on data literacy competencies in an organization, the tool should not only be used by single individuals in an organization, but within entire teams to get a better understanding of the overall situation and make the averages more meaningful.

Further Comments

The interview partners from the organization specifically mentioned the following points for improving the model:

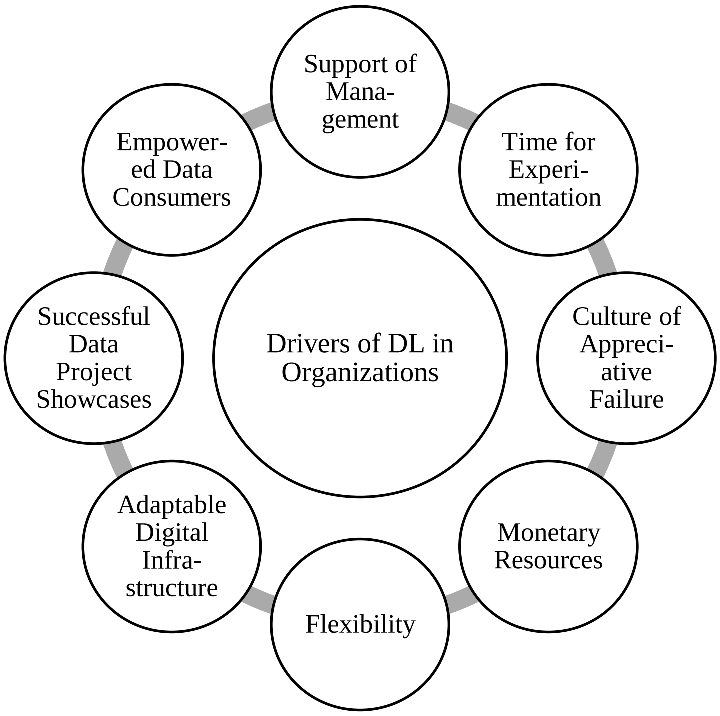
- Given answer possibilities might be too difficult, depending on which organizations are approached (010517_Evaluation_IP8, n371, 010517_Evaluation_IP9, n388, n393).
- Applying the model highly depends on how an organization is working, different skills are needed for specific purposes (010517_Evaluation_IP8, n372).
- Data Culture: given answer possibilities describe too many topics (esp. Data Culture), should be broken down into several questions (010517_Evaluation_IP9, n394, n385).
- Verify: Pay attention to clear division between organizational and individual competencies (010517_Evaluation_IP8, n370).
- Find Data: Use of own data is missing (do not combine them with external sources) (010517_Evaluation_IP9, n386, n387).
- Get: Use CRM system, internal data, multidimensional queries (010517_Evaluation_IP9, n389).
- Communicate: To funding partners: aggregated numbers, reportings / newsletters (010517_Evaluation_IP9, n390, n391).
- Model only highlights data handling and conversion, but not critical interpretation / assessment of data (outputs) (010517_Evaluation_IP8, n370, 010517_Evaluation_IP8, n374, n375, n381, in accordance with criticism of Expert 3, see Iteration Phase 2).

Reflecting on the evaluation outcome, the model was still considered to be a useful tool to help individuals and organizations to find out how people handle data and where competencies still need to be trained more intensively (010517_Evaluation_IP8, n396). According to interview partner 9, this will help to promote openness on working, using and communicating with data and better plan education programs, which perfectly corresponds with the initial purpose of the model (010517_Evaluation_IP8, n397). A rejection of the model therefore is not necessary. However, further adaptations need to be acknowledged, which will be incorporated in the beta – version.

Drivers and Barriers of Data Literacy

As a last part of the analysis, drivers and barriers of data literacy can be identified (see Figure 23) to get a better feeling of the eco-system as well as to describe indicators for facilitating to move along different maturity levels. One major aspect that promotes more data literacy in organizations is the support of management, so that team members get a feeling that data is an important topic to work on and understand it as an organization-wide approach (112816_Transcript_Expert1_2, n20, n21, n168, 120516_Transcript_Expert3, n231, 120816_Transcript_Expert4 n246, n312, 120516_Transcript_Expert3, n233). This point became even more important, when one interviewee mentioned that time is key to establish data skills (112816_Transcript_Expert1_2, n165). The experts mentioned that it is crucial to have enough time, besides daily business, to get in touch with a new topic, train certain skills, experiment and learn from mistakes (112816_Transcript_Expert1_2, n165, n166, n169, n170, n174, 120816_Transcript_Expert4 n312). This directly led to another point: room for experimenting and a culture of appreciative failure to explore the topic of data (112816_Transcript_Expert1_2, n167, n173). Apart from that, money can be an enabler for data topics in organizations as well, especially when it comes to offering trainings and acquiring additional human and technical resources (112816_Transcript_Expert1_2, n171, 120516_Transcript_Expert3, n230, n234). Regarding technical resources, the experts mentioned that not only hardware resources have to be present, but also the freedom to choose and use new software to facilitate working with data and allow for a certain digital infrastructure (112816_Transcript_Expert1_2, n172, 120516_Transcript_Expert3, n232). This also calls for a certain degree of flexibility that is needed to drive data projects in an organization and to reframe the scope and purpose of it, especially at the beginning when experience is still lacking (112816_Transcript_Expert1_2, n175, n176). In this regard, successful showcases of data projects can serve as motivators to engage in data topics (120816_Transcript_Expert4 n247, n309, n310, 120816_Transcript_Expert4 n311). Putting this topic in a broader perspective, the experts mentioned that a common understanding of data potentially being highly beneficial for society is a crucial motivator for NGOs to include data in their activities (112816_Transcript_Expert1_2, n177). Nevertheless, society also needs to be able to understand, appreciate and critically consume data products that are provided to it in order to create a fruitful data eco-system (112816_Transcript_Expert1_2, n178).

Figure 23 Drivers of Data Literacy in Organizations



7 Discussion

The results of the data collection process have been integrated into the beta-version of the data literacy maturity model, which can be seen in the maturity grid in Figure 24 and the evaluation questionnaire (see Appendix H).

Figure 24 Data Literacy Maturity Grid Beta – Version

Underlying Rationale		**Read** →	→	→	**Write**
Generic Level Descriptors		**Uncertainty**	**Enlightenment**	**Certainty**	**Data Fluency**
General Level Descriptions		Organizations are unaware of the need for data literacy skills and have no or very vague understanding on what is required. Individuals might have a certain interest in data and work digitally, but are unsure about the different steps that exist when working with data.	Organizations are experimenting with the application of data-related topics. Describes a state where a lot about data has already been understood theoretically, but cannot be applied in many cases and has to be further trained.	Organizations perform data handling steps with confidence and have built data-driven activities into their routine processes wherever it makes sense. Generic procedures and standards on how to handle data are formalized and widespread, benefits are understood at all levels of the organization.	Organizations have established a data-informed culture throughout all levels. Data is actively used to improve processes and create workflows.
Organizational	**Data Culture**	Data is perceived as an ambiguous term which causes insecurities.	Data is perceived as an interesting concept and benefits are appreciated. Insecurities exist regarding use cases and what exactly to expect.	Data is not perceived as a source of insecurities, but rather understood as an enabler for progress and support for existing and planned activities. Higher management and leaders support data initiatives.	Psychological barriers of data have been brought down (e.g. insecurities, fear, resignation) and comfort around data is promoted. Higher management and project managers understand and support importance of dedicated resources (time, budget, human resources) for data handling and conversion.
Organizational	**Data Ethics & Security**	No awareness for guidelines that ensure confidentiality, integrity and availability of data.	Rising awareness and uncoordinated attempts to promote the importance of the responsible use of data. No defined guidelines.	Awareness of the impacts of data use. Guidelines for responsible data handling are defined and incorporated internally to activities.	Processes are in place to ensure confidentiality, integrity, and availability of data. Only data that is necessary is collected/used. Consistent, companywide policies for secure and ethically sound data handling are constantly redefined and updated.
Individual	**Ask question & Define**	Lacking ability to formulate questions to find meaningful answers in data. No feeling about which questions can be answered by data.	Questions can be asked to data in limited number of situations and answers are provided through simple queries.	Questions to data are formulated precisely and target-oriented to find meaningful answers in most of the cases.	Entire projects are based on multidimensional questions. Answers to informational needs can be found consistently in data, because of the high awareness of what questions can be answered by data (no overinterpretation).

	Find	Limited understanding of possible data sources. Use of basic search engines to find data. No experience for identifying and selecting most relevant data sources.	Knowledge only limited to a few data sources. Advanced use of search engines, use of internal data sources and data requests at public institutions are common practice.	Broad understanding of different data sources, most relevant ones can be chosen from a selection of data sources. Awareness and use of data portals for specific topics.	Profound understanding of the various possible types of data sources. Assessment criteria for selecting the ones most relevant to an informational need are formulated. Ability to detect when a given problem or need cannot be solved with the existing data, and knowledge about research techniques to obtain new data (e.g. complex queries).
	Get	Data is derived from full text and used as base for further processing.	Use of downloads and data formats such as .csv . Often use of internal programs to access data (e.g. CRM).	Data can be accessed using more complex data formats (e.g. JSON, XML). Use of APIs to get data.	Access to data through sophisticated methods (e.g. automated data scrapers / scripts). Ability to convert input format into a form that can be used for further processing and analysis.
	Verify	Critical evaluation of data does not exist, data is taken at face value. Data evaluation criteria cannot be described.	Critical check of simple data quality measures.	Multiple layers of data checking are implemented in standard procedures across the organization.	Ability to do data quality assessment independently. Data evaluation criteria regarding authorship, method of obtaining and analyzing data, comparability and quality are precisely defined.
	Clean	No awareness that given data might have to be checked, cleaned or normalized. Data is further processed as is.	Awareness that given data most often is not perfect. Awareness of some data quality criteria (e.g. empty fields, duplicates) and manual fixing of errors.	Invalid records can be detected and are removed using programs that support data cleaning (e.g. OpenRefine). High awareness of data quality criteria (e.g. machine processable, empty fields, duplicate detection).	Independent ability to remove invalid records and translating all the columns to use a sane set of values through an automated script. Ability to combine different datasets into a single table, remove duplicate entries or apply any number of other normalizations.
	Analyze	Bar and pie charts, simple use of data tables and basic summaries of data.	Ability to work with basic descriptive statistics. Pivot tables for aggregating information, histograms and boxplots.	Ability to work with advanced statistics (e.g. inferential view of data, linear regression, decision trees).	Full suite of machine learning tools (e.g. clustering, forecasting, boosting, ensemble learning).
	Visualize	No awareness of the multiplicity of how data can be presented. No understanding of when standard visualizations are chosen, decision based on what looks best (trial and error).	Ability to find specific outputs in accordance with information that want to be represented (e.g. in Excel).	Creation of interactive charts / dashboards, uncertainties are always visualized along with the data.	High awareness of the various forms in which data can be presented (written, numerical or graphic). Sophisticated visualizations are programed, linked, dynamic dashboards that anticipate user requests are designed.
	Commu-nicate	Insights from data are not communicated or put into a broader context.	Limited ability to find specific outputs. Simple narrative support static visualizations / key numbers (e.g. reporting to funding partners, newsletters).	Own projects are supported by interactive visualizations and more sophisticated narrative in a broader context. (e.g. data storytelling, conferences, talks, monthly updates, blog posts).	Ability to synthesize and communicate in ways suited to the nature of the data, their purpose and the audience (e.g data storytelling, data-driven campaigning, workshops, conferences, monthly updates, blog posts, reproducible research).

	Assess & Interpret	Data outputs are used at face value without questioning their correctness and message.	Growing awareness for critically assessing data outputs and interpreting the results. Insecurities regarding what exactly to pay attention to.	Data outputs and results are interpreted confidently and critically. Evaluation criteria are internalized.	Data outputs and results are consistently questioned and challenged, interpretation extents the obvious and information are successfully translated into actionable knowledge.

Throughout the course of this study, the data literacy maturity model emerged from a preliminary maturity grid with seven DL competence dimensions (1 Data Sources, 2 Data Thinking, 3 Data Inquiry, 4 Data Outputs, 5 Data Objectives, 6 Data Culture, 7 Data Analysis) to a maturity model that describes data literacy in non-governmental organizations through 11 competencies, both on an organizational level (1 Data Culture, 2 Data Ethics & Security) as well as an individual level (3 Ask questions & Define, 4 Find, 5 Get, 6 Verify, 7 Clean, 8 Analyze, 9 Visualize, 10 Communicate, 11 Assess & Interpret). It is distributed on four competence levels (Uncertainty, Enlightenment, Certainty, Data Fluency) as presented in Figure 24. The beta-version of the maturity model contains a maturity grid, which is complemented by a self-assessment excel tool (Appendix H).

The development of the data literacy maturity model showed that it is a) a quick and easy to use tool for, (b) an insightful exploration of data literacy competencies at different levels and (c) a way to indicate current data literacy competencies in an organization, which (d) offers options for an organization to better plan data projects and trainings for the future. It therefore answers the research question of how to describe data literacy in a maturity model for non-governmental organizations. However, the analysis yielded results that, when set into a broader perspective, extend the limits of the research question. Five aspects that critically assess the development of the data literacy maturity model are extracted in the following.

Data Processors vs. Data Consumers – a promising Insight for describing DL

Overall, this research resonates with other research that is concerned to find a suitable way for describing data literacy (Bhargava and D'Ignazio, 2015; Calzada Prado and Marzal, 2013; Data-Pop Alliance, 2015; Gray et al., 2012) and confirms the difficulties to find a universal description. It turned out that DL allows for different interpretations, especially distinguishing between processing / handling data and using / interpreting data (especially data outputs). This is one of the most interesting insights of this study regarding the problem in this field that do not yet address this differentiation explicitly. At the same time, traditional literacy earlier was described as "the ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts." (UNESCO, 2004: 13), which implies that this differentiation is not destined by the original interpretation of literacy. Hence, future research might explore if this separation is necessary in the area of data literacy, or if all competencies can be combined in one definition.

Interestingly, elements of the interpretation of data literacy in the suggested maturity model to some extent confirm the definition of Wolff et al. (2016) (see Chapter 2), which describes data literacy as: asking and answering questions, ethical use of data, select, clean, analyze, visualize, critique and interpret, and communicate stories. However, the researcher identified additional aspects that are missing in Wolff's description of data literacy in an organizational context: data culture, as well as individual competencies, such as get and verify. Hence, a more complete and precise picture can be illustrated with the insights from this research for the study object of non-governmental organizations. Nonetheless, this also contributes to the widely-held argument that it is difficult to agree on one definition for data literacy when considering different application domains (Bhargava and D'Ignazio, 2015; Calzada Prado and Marzal, 2013; Data-Pop Alliance, 2015; Gray et al., 2012). At the same time, the findings of this study suggest that it is necessary to formulate and agree on a common basis for further conversations on data literacy, to break down the insecurities attached to the term and to make future research more comparable.

Context is Key

Throughout the analysis, the importance of considering the application context of the model has been mentioned several times. This shows that empowerment through data can only be guaranteed, if organizations and individuals have the freedom to choose which direction they want to follow and what competencies are meaningful to them. At the same time, it turned out that data literacy is a topic that means different things to different people under varying contexts. It therefore became difficult to represent this multiplicity in a model that is known for simplifying the study object to a fraction of its real complexity, especially when it comes to differing application scenarios (de Bruin et al., 2005; McCormack et al., 2009). To respect this sensitive fact, the DLMM does not represent an exclusive solution, but rather offers option values for planning and stimulating adapted action. One way could be to modularize the content of the DLMM and only regard special topics of interest when it comes to data literacy to foster the discussion in organizations. On the other hand, the DLMM offers a way to operationalize data literacy in organizations through its applied descriptions of the different dimensions and levels, rather than focusing on general traits. It therefore provides an option to break down vague definitions, and in this regard, can especially help data learners to accomplish their mission.

Transferability to Other Domains

Although this research was set out to explore data literacy in non-governmental organizations, the clear influence on this organizational context is hard to evaluate. The experts mentioned that the DLMM can easily be transferred to other domains, which implies a certain level of generalizability of the model in general. Nevertheless, the argument that the organizational context matters when evaluating data literacy competences was repeatedly mentioned by different experts as well as the team members of the NGO. Hence, this might implicitly demonstrate the specifics of NGOs on the topic of data literacy and represents the importance of respecting the multitude of NGOs. A lot of social change organizations are not only new to the topic of data literacy, but also need data for different purposes, due to differing topics, objectives and visions. The examination of a more differentiated view on social change organizations and data literacy can therefore lead to promising future research projects.

Flexibility in Exchange of Completeness

As already mentioned, the DLMM should offer a high level of adaptability for different organizational contexts and thus cannot map a complete picture of data literacy. Since this has not been the purpose of this study, the model rather offers a suggestion and option values to follow and should be interpreted as a model that is offering guidance, orientation and no absolute commandments. This also corresponds with existing research, which describes that maturity models need to stay flexible regarding their changing technological, social and organizational environment (Mettler and Rohner, 2009) and can therefore never represent a complete picture. Especially under these circumstances, it was surprising that the model was rejected by one of the interview partners, who mentioned that it does not respect different contexts and simply tries to fit something into a model that is not made for it 1(20516_Transcript_Expert3, n227). To circumvent this criticism, future research might develop a system of modularizing the DLMM, so that organizations can individually define their objectives and answer questions from a pool of questions for the different competencies, while respecting their organizational context. The tool would thereby be made customizable for a broader audience and serve more specific needs.

Do We need it All? – Opportunities for Improved Prioritization

The study, and especially the evolution from seven to eleven data literacy competencies, showed that data literacy is diverse. Although the representation of data literacy in a maturity model was challenged during this study, the findings suggest that the model was perceived as a good way to raise awareness on the topic and can help to better understand what data literacy is and what can be expected from it. It therefore contributes to an existing research gap that was described earlier (Data-Pop Alliance, 2015; Wolff et al., 2016). Nevertheless, it became evident that not all skills might be necessarily needed to become data literate. It turned out that individuals and organizations would rather concentrate on specific competencies that are described in the whole model, which might foster the creation of new, specialized data roles for data handling skills in organizations that have expertise in a specific dimension of data literacy. However, still in recognition of this element, the data literacy maturity model could be used to evaluate organizations through integrating different persons in the evaluation process, which depicts that the application of the model might extend the potential as presented in this study. The DLMM would thereby allow a better prioritization of efforts and optimize the activity portfolio for data literacy in organizations.

8 Conclusion

This study contributes to the general question of how to describe data literacy for non-governmental organizations in a maturity model. With the help of a structured review of literature, a preliminary data literacy maturity grid has been created. After that, action design research has been used, to refine the details of the framework and to design a complementing self – evaluation excel tool. Through exploratory expert interviews and a solution testing in an organizational setting throughout three iteration phases, the maturity model grew more and more detailed at each phase of the development process. The major research contribution lies in the development of the maturity grid and the translation into an easy-to-use excel tool that helps to evaluate current data handling and conversion skills and to better prepare for the future. This research suggests a description of data literacy based on 11 competencies, both on an organizational (1-2) and an individual level (3-11): 1 Data Culture, 2 Data Ethics & Security, 3 Ask questions & Define, 4 Find, 5 Get, 6 Verify, 7 Clean, 8 Analyze, 9 Visualize, 10 Communicate, 11 Assess & Interpret. These dimensions are described across four different competence levels: 1 Uncertainty, 2 Enlightenment, 3 Certainty, 4 Data Fluency. This study was set out to explore the definition of data literacy in a maturity model, which represents an original contribution to the research domain. Although the findings of this research suggest that representing data literacy in a maturity model can be a helpful way to foster discussion in organizations, especially regarding the evaluation of current competencies and a plan for the future, it was challenged by the fact that it tends to simplify the topic under investigation and cannot represent the multiplicity of organizational contexts in the non-governmental sector. However, it was not within the scope of this thesis to offer a "fits-all" solution, but rather a proposal of option values as a guideline for the still opaque topic of data literacy as described initially. Moreover, it can help to better discuss the bigger picture of data literacy competencies and thus make future educational endeavors more scalable. By doing that, it can help non-governmental organizations to catch up with data practitioners from science and businesses (Desouza and Smith, 2014). Overall, the research set the stage for further promising research directions around the topic of data literacy in NGOs. Additionally, the suggested model can potentially be used to evaluate and explore data literacy in other areas such as businesses and therefore sets a foundation for further education programs in this area.

8.1 Limitations

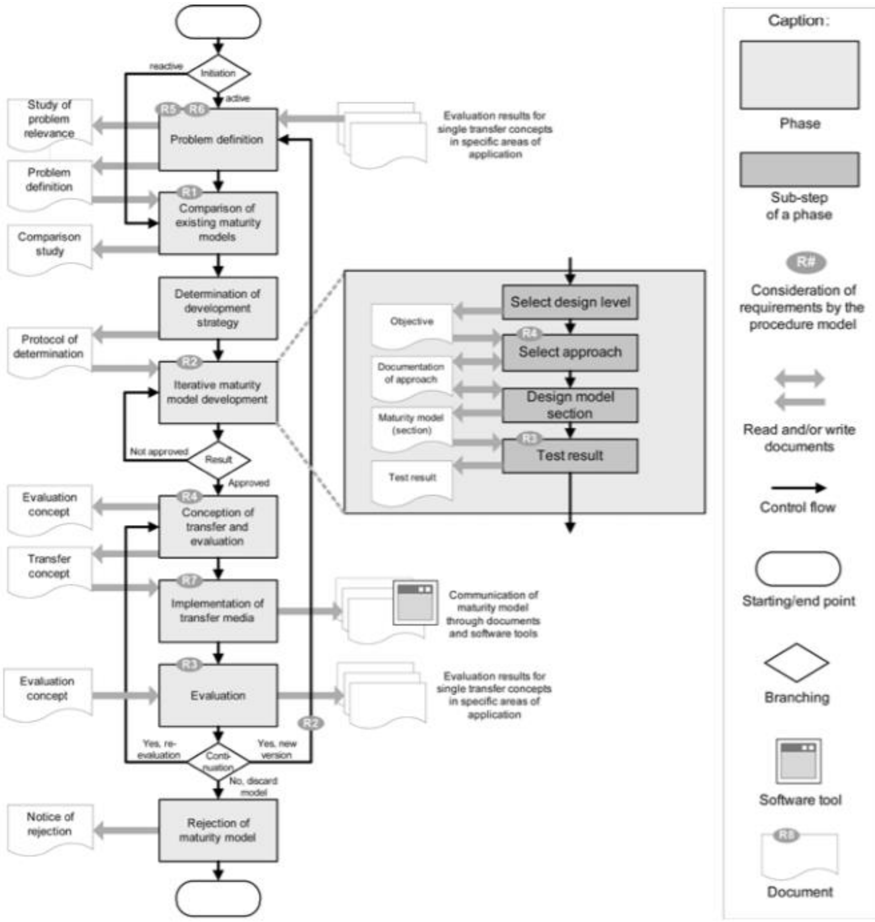
Although the research has reached its aim to explore and describe data literacy in a maturity model for NGOs, unavoidable limitations need to be acknowledged. First, the period of time dedicated to this study limits what is revealed, especially regarding the ambitious development process that contained three iteration phases. This led to the conscious decision to focus the data collection process to one specific case. As a consequence, the ability to draw generalizations for NGOs operating in other fields is restricted. Nevertheless, since the purpose of this study was an initial exploration of data literacy in NGOs, the study sets out to be a starting point for future research in this respect. Second, the decision to represent data literacy in a maturity model was a new approach that was best investigated with the exploratory character of qualitative research methods such as expert interviews. A main shortcoming of this approach is the subjective influence regarding data collection and the analysis of the results. However, by combining a deductive and inductive code approach, the degree of subjectivity in this study can be limited, since findings are based on pre-established theories. Apart from that, the study highly depends on the natural setting of the study objects, which might cause difficulties in reproducing the study regarding its findings. Nonetheless, a strict methodology has been followed that guarantees the reproducibility of this study. One major limitation of this study is that research on data literacy in general is opaque and the representation in a maturity model has been un-touched until now. The interpretation of the data and the translation into a maturity model therefore highly depend on the subjective interpretation of the researcher and could have yielded different results by guaranteeing intercoder reliability, which was not possible due to the time limit. Another shortcoming of the study is that the influence of the organizational context and the study object of non-governmental organizations could not have been extracted precisely, although the sample frame has been strictly limited to non-governmental organizations only. Although this discovery was only made after the data collection, one possibility to address this problem would have been to include an explicit question regarding NGOs in the semi-structured interviews with data experts and practitioners.

8.2 Where do we go from here? Implications for Future Research

The development of a maturity model for data literacy in NGOs opened various new research directions for the future, both for the application and further refinement of the model and regarding the definition of data literacy. One potential research lead is the exploration of a potential benchmarking function of the model between different organizations. What are differences between learning organizations and organizations that already have a certain expertise in data handling and conversion topics? This can be an exciting way to further refine the model in varying organizational contexts. Moreover, examining the correlations between the different competences that are described in the DLMM can lead to promising research streams. It would be interesting to learn about the connection and influence of the different dimensions and which recommendations / insights can be drawn from them. While the study object of the Datenschule helped to explore the topic of data literacy in the context of social change organizations, further research will need to be performed to settle the reliability of the results. This study could apply the DLMM to non-governmental organizations with different focuses (human rights, environment, gender equality, etc.) to create a global picture of data literacy competencies within non-governmental organizations of different domains. This can turn out to be a highly interesting research focus, since the findings of this study suggest that the big contextual difference between non-governmental organizations can have a dramatic impact on the interpretation of data literacy. However, not only the differences within the non-governmental sector can yield interesting research, but also the influence and understanding of other domains such as businesses or governments as well as society as a whole can create emerging research streams for future exploration. Apart from that, future research could contribute to investigate the meaning of data literacy from a user / interpreter and processor view and determine whether a differentiation between the two is necessary.

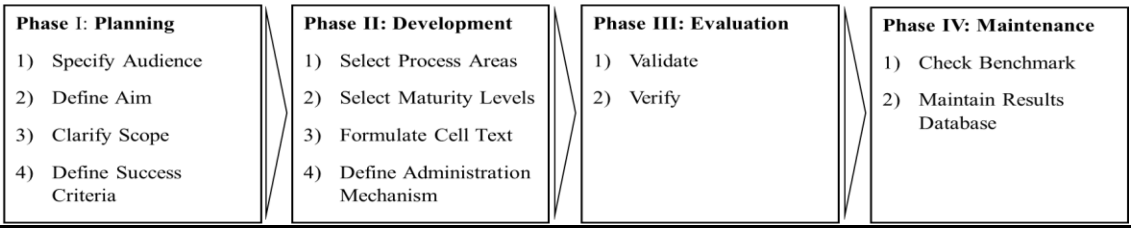
Appendices

Appendix A: Procedure Model for Developing Maturity Models



Based on Becker et al. (2009)

Appendix B: Roadmap to Develop New and Evaluate Existing Maturity Grids



Phase	Decision Points	Decision Options
Phase 1 Planning	1) Specify Audience	Users (identify all potential stakeholders) Improvement entity (e.g. teams, organization, process or product)
	2) Define Aim	Raise awareness or best practice benchmark
	3) Clarify Scope	Generic (e.g. energy management) or domain-specific (e.g. energy management in construction)
	4) Define Success Criteria	High-level requirements (e.g. usability, usefulness) Specific requirements
Phase II Development	1) Select Process Areas (components and theoretical framework)	E.g. Reference to established body of knowledge; Literature review; Expert knowledge; Defining goals
	2)Select Maturity Levels (underlying rationale)	E.g. Existence and adherence to a structured process; Alteration of organizational structure; Emphasis on people; Emphasis on learning
	3) Formulate Cell Text	Type of formulation: prescriptive or descriptive Information source: Synthesizing viewpoints from future users or comparing practices of several organizations Formulation mechanism: Inductively generated from descriptions of practice or deducted from underlying rationale
	4) Define Administration Mechanism	Focus on the process of assessment (e.g. face-to-face interviews, workshops) or focus on end results (e.g. survey)
Phase III Evaluation	1) Validate	Correspondence between author's intent and user's understanding Correctness of results
	2) Verify	Correspondence with Requirements specified
Phase IV Maintenance	1) Check benchmark (and adjust descriptions in cells)	If applicable
	2) Maintain Results Database	If applicable
	3) Document and Communicate	Audience-specific

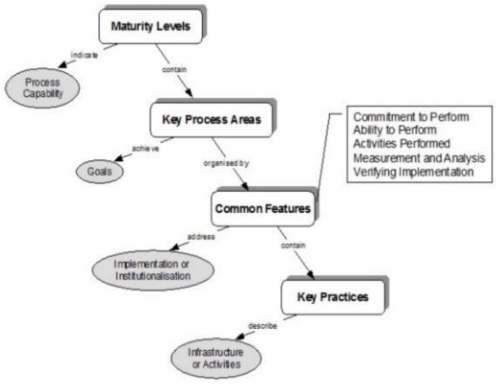
Based on Maier et al. (2012)

Appendix C: Origins of Maturity Thinking

Measurement Categories	**Stage 1: Uncertainty**	**Stage 2 Awakening**	**Stage 3: Enlightenment**	**Stage 4 Wisdom**	**Stage 5: Certainty**
Management understanding and attitude	No comprehension of quality as a management tool. Tend to blame quality department for "quality problems".	Recognizing that quality management may be of value but not willing to provide money or time to make it happen.	While going through quality improvement program learn more about quality management; becoming supportive and helpful.	Participating. Understand absolutes of quality management. Recognize their personal role in continuing emphasis.	Consider quality management as an essential part of the company system.
Quality organization status	Quality is hidden in manufacturing or engineering departments. Inspection probably not part of the organization. Emphasis on appraisal and sorting.	A stronger quality leader is appointed but main emphasis is still on appraisal and moving the product. Still part of manufacturing or other.	Quality department reports to top management; all appraisal is incorporated and manager has role in management company.	Quality manager is an officer of company effective status reporting and preventive action. Involved with customer affairs and special assignments.	Quality manager on board of directors. Prevention is main concern. Quality is a thought leader.
Problem handling	Problems are fought as they occur, no resolution, inadequate definition; lots of yelling and accusations.	Teams are set up to attack major problems. Long-range solutions are not solicited.	Corrective action communication established. Problems are faced openly and resolved in an orderly way.	Problems are identified early in their development. All functions are open to suggestion and improvement.	Except in the most unusual cases, problems are prevented.
Cost of quality as % of sales	Reported: Unknown Actual: 20%	Reported: 3% Actual 18%	Reported: 8% Actual: 12%	Reported: 6.5% Actual: 8%	Reported: 2.5% Actual: 2.5%
Quality improvement actions	No organized activities. No understanding of such activities.	Trying obvious motivational short-range efforts.	Implementation of a multi-step program with thorough understanding and establishment of each step.	Continuing the multi-step program and starting other pro-active / preventive product quality initiatives.	Quality improvement is a normal and continued activity.
Summary of company quality posture	"We don't know why we have problems with quality."	"Is it absolutely necessary to always have problems with quality."	"Through management commitment and quality improvement we are identifying and resolving our problems."	"Defect prevention is a routing part of our operation."	"We know why we do not have problems with quality."

Quality Management Maturity Grid based on Crosby (1980)

Optimizing	Continuous process improvement is enabled by quantitative feedback from the process and from piloting innovative ideas and technologies
Managed	Detailed measures of the software process and product quality are collected. Both the software process and products are quantitatively understood and controlled.
Defined	The software process for both management and engineering activities is documented, standardized, and integrated into a standard software process for the organization. All projects use an approved, tailored version of the organization's standard software process for developing and maintaining software
Repeatable	Basic project management processes are established to track cost, schedule, and functionality. The necessary process discipline is in place to repeat earlier successes on projects with similar applications
Initial	The software process is characterized as ad hoc, and occasionally even chaotic. Few processes are defined, and success depends on individual effort and heroics



Capability Maturity Model based on Paulk (1993)

Appendix D : Interview Guidelines

Datenschule Practitioners Interview Guideline – Data Literacy Maturity Model (DLMM) Content

Purpose

- Clearly define stakeholders
- Check general composition of the model based on given criteria
- Collect data on the description of the individual levels (cell text)
- Duration approximately 2 hours

1. Problem Definition and Planning

1.1 How would you specify the audience / stakeholders for a maturity model that measures data literacy capabilities for your purposes?

1.2 Which improvement entity do you think is most important regarding the model? (e.g. teams, organizations, process, product, etc.)

1.3 What do you consider more important for the model: raising awareness for data literacy or best practice benchmarking between organizations?

2. Development

2.1 What do you think about the following process areas to describe and evaluate data literacy capabilities?

Competence	Description
Data Sources Know how to find data	track down sources of existing data, know how to collect data if it does not exist yet
Data Thinking Apply critical thinking skills to data	ability to do data quality assessment, contextualizing specific information to other aspects
Data Inquiry Ask questions to the data and find answers	ability to ask questions to data and ultimately find answers as one of the goals of data literacy trainings
Data Outputs Find specific outputs (stories or visualizations) in data	importance of finding stories and other journalistic outputs
Data Objectives Use it to advance one’s goals	link between data and action was evident
Data Culture Feel comfortable around data and working with it	promoting comfort around data (and bringing down the psychological barriers that exist between people and data)
Data Analysis Do statistical analysis with data	ability to work with basic statistics

2.2 Is there anything you are missing and how would you describe it?

2.3 If you were to describe different capability levels for data literacy, which ones would you choose? Would you prefer another description to the given ones? If yes, which ones?

Level 1	Acclimation		Uncertainty		Naive	
Level 2	Early Competencies		Awakening		Novice	
Level 3	Late Competencies		Enlightenment		Normalized	
Level 4	Proficiency		Certainty		Natural	

2.4 Formulate cell text.

2.4.1 Please briefly describe the most critical aspects of each dimension in accordance with the experience you made during your project work .

What would you say are the primary activities of each data literacy dimension?

Reflect on highly effective to highly ineffective observable behaviors relevant to the dimension under consideration.

2.4.2 Please take a look at the filled maturity grid and categorize the respective descriptions into the following categories:

Agree	**Incomplete**	**Disagree**

Please briefly specify, in case you have chosen the categories *incomplete* and disagree for certain descriptions, what was missing / wrong.

2.5 Based on your opinion, what could be drivers and barriers of maturation when it comes to data literacy?

Maturity models should explicate the underpinning theoretical foundations of evolution and change with respect to the class of entities under investigation. This includes among other things information about the way change typically happens in the respective application domain as well as about drivers and barriers of maturation.

3. Evaluation

3.1 What do you think about the following success criteria for the evaluation of the data literacy maturity model (after the design phase is completed)?

Evaluation Criteria	**Is the evaluation criteria adequate / meaningful?**		**Comment**
	Yes	**No**	
Correctness			
Flexibility			
Usability			
Implementability			
Economic Efficiency			

3.2 Is there an evaluation criteria that you are missing? If yes, which one?

4. Closing

Is there anything that is still unclear or that you would like to add?

Thank you for your participation ☺

Interview Guideline Expert Interviews

Purpose of the Focus Group

- Check general composition of the model based on given criteria, especially technical requirements
- Collect data on the description of the individual levels (cell text)
- Duration approximately 1 hour

1. Development

1.1 What do you think about the following areas to describe and evaluate data literacy capabilities?

	Competence	Description
Organizational	**Data Culture**	Promoting comfort around data (and bringing down the psychological barriers that exist between people and data).
	Data Ethics & Security	Processes that are in place to ensure confidentiality, integrity, and availability of data is adequately protected.
Individual	**Ask question / Define**	Ability to ask adequate questions to data and ultimately find answers
	Find	Track down sources of existing data, know how to collect data if it does not exist yet
	Get	Describes gaining access to data or generating fresh data as well as conversion of different input formats
	Verify	Apply critical thinking skills to data. Ability to do data quality assessment, contextualizing specific information to other aspects
	Clean	Removing invalid records and translating columns to use a sane set of values
	Analyze	Ability to work with statistics and other analytical methods
	Visualize	Ability to represent findings in appropriate visual outputs
	Communicate	Importance of finding stories and communicating them to the targeted audience

1.2 Is there anything you are missing and how would you describe it?

1.3 Formulate cell text.

Please briefly describe the most critical aspects of each dimension in accordance with the experience you made during your project work. What would you say are the primary activities / characteristics of each data literacy dimension?

Please reflect on highly effective to highly ineffective observable behaviors relevant to the dimension under consideration.

Unsatisfactory Performance	Marginal Performance	Fully Competent Performance	Excellent Performance

1.4 Based on your opinion, what could be drivers and barriers of maturation when it comes to data literacy?

2. Closing

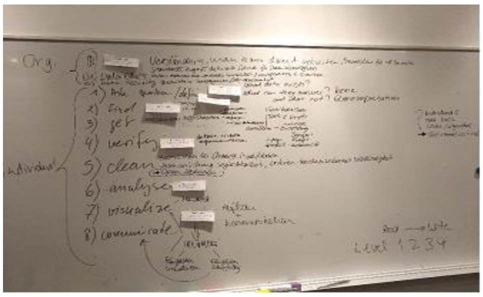
Is there anything that is still unclear or that you would like to add?

Thank you for your participation ☺

Appendix E: Codings of all Documents – Model Development

n	Document	Code	Segment
1	112816_Transcript_Expert1_2	Problem Definition\Stakeholder	super viele Anwendungen finden. Das kann man glaube ich auf sehr viele Sachen übertragen.
2	112816_Transcript_Expert1_2	Problem Definition\Stakeholder	NGOs bzw. gemeinnützige Vereine
3	112816_Transcript_Expert1_2	Problem Definition\Stakeholder	auch auf andere Organisationsformen noch übertragen
4	112816_Transcript_Expert1_2	Problem Definition\Stakeholder	NGOs zusammen, die ein grundlegendes Datenverständnis haben. Das bedeutet, dass sie nicht sozusagen damit arbeiten können, aber dass sie zumindest erkannt haben was das Potenzial für ihre Arbeit, für ihre Organisation, für ein Projektteam, was das Potenzial von Daten sein könnte. Die in irgendeiner Form mit Fakten, Beweisen und Zahlen arbeiten und etwas nachprüfen wollen. Also das Klassische ist diese Monitoring Sache. NGOs, die in irgendeiner Weise Fact-Checking machen oder Monitoring machen.
5	112816_Transcript_Expert1_2	Problem Definition\Stakeholder	NGOs, die ein grundlegendes Datenverständnis haben, aber natürlich in bestimmten Themenwolken mit uns zusammen, also zum Beispiel Bildung.
6	112816_Transcript_Expert1_2	Problem Definition\Stakeholder	Ich glaube, dass das Maturity Modell etwas ist, was nicht für alle letztendlich gelten kann, weil du musst ja ein gewisses Verständnis davon haben was für ein Potenzial es gibt und dass es sich eher auf das digitale Arbeiten bezieht.
7	112816_Transcript_Expert1_2	Problem Definition\Stakeholder	NGOs, die digitalen Content produzieren, veröffentlichen und kommunizieren und in irgendeiner Form digital mit anderen Leute in Kontakt treten, das sind glaube ich die, für die so ein Modell sinnvoll ist oder wo man das gut anwenden könnte zumindest.
8	112816_Transcript_Expert1_2	Problem Definition\Stakeholder	es geht wirklich darum Daten für politische Themen oder gesellschaftsrelevante Probleme zu nutzen und damit dieses Fact-Checking zu machen und das sind so die NGOs
9	112816_Transcript_Expert1_2	Problem Definition\Stakeholder	nicht mit analog arbeitenden NGOs zusammenarbeiten
10	112816_Transcript_Expert1_2	Problem Definition\Stakeholder	einen externen Nutzen, also es ist nicht, dass man jetzt interne Prozesse optimieren möchte? IP1: Naja doch, weil erstmal gucken wir uns an, wer ist überhaupt geeignet dafür, den mir angehen können und wo wir sagen können, hier gibt es Potenzial, hier können wir mit unseren Maßnahmen mit Workshops usw. ansetzen
11	112816_Transcript_Expert1_2	Problem Definition\Stakeholder	Anfang einer NGO Partnerschaft haben wir es so gemacht, dass wir uns auch wirklich konkret anschauen, wie lagern die eigentlich intern ihre Daten, wie sind die gespeichert, mit was für Programmen arbeiten sie? Und bisschen so diesen Workflow sich anzugucken, wie teilen sie Daten innerhalb der Organisation? Das sind natürlich schon so Sachen, die intern gerichtet sind
12	112816_Transcript_Expert1_2	Problem Definition\Improvement Entity	internen Kompetenzen aufzubauen ist super wichtig
13	112816_Transcript_Expert1_2	Problem Definition\Improvement Entity	du fängst quasi bei einer einzelnen Person innerhalb einer Organisation an und das baut sich dann halt auf. Also du kannst, ich glaube nicht, dass das sozusagen von oben nach unten vorausgesetzt werden kann bestimmte Prozesse, gerade was Arbeiten mit Daten angeht, definiert werden kann.
14	112816_Transcript_Expert1_2	Problem Definition\Improvement Entity	Ich glaube das ist eher etwas was stetig, bottom-up aufgebaut wird

15	112816_Transcript _Expert1_2	Problem Definition\Improvement Entity	Wo man sagt, ok, es gibt hier eine Person, die versteht bestimmte grundlegende Datenarbeitsschritte, sag ich jetzt mal und ist anerkannt bei den anderen Teams, oder wird um Meinungen gebeten. Es gibt dann also so bestimmte Workflows wo man sagen kann: ok, ich habe hier Daten oder ich geh zu der Person oder die Person ist beteiligt in einzelnen Projekten und kann da das Wissen einbringen, oder es gibt vielleicht eine definierte Rollenverteilung in einem Team, der eine ist also Designer, der andere ist Entwickler, oder Journalist und dort kommt man dann zu, dass man darüber spricht, wie man Daten einbetten kann und das steigert sich dann immer weiter bis eine Organisation Workflows für wie lagern wir Daten, wie managen wir Daten, was passiert wenn das Projekt vorbei ist? Was machen wir dann mit den Daten?
16	112816_Transcript _Expert1_2	Problem Definition\Improvement Entity	Also ich glaube man muss aufjedenfall so ein Modell bei Personen ansetzen ersteinmal
17	112816_Transcript _Expert1_2	Problem Definition\Improvement Entity	ok, wie sind da eigentlich die Data Literacy Skills vorhanden? Und dass man dann von da aus aggregieren kann. Also in dem Team habe ich die und die Stärken, weil im Endeffekt sind es ja erstmal die einzelnen Personen, die an Daten arbeiten. Und daraus kann man dann quasi auf die höheren Ebenen Rückschlüsse führen.
18	112816_Transcript _Expert1_2	Problem Definition\Improvement Entity	Genau, also ich glaube man fängt bei der Einzelperson an
19	112816_Transcript _Expert1_2	Problem Definition\Improvement Entity	Das ist natürlich auch das einfachste, weil man dann gut abfragen kann, was sind deine Skills, was ist dein Verständnis, was nutzt du bereits? Solche Sachen und das dann sozusagen bisschen größer denkt in Teams oder Abteilungen usw. Aber ich glaube es ist total wichtig dieses Basic Data Knowledge zu vermitteln an Einzelpersonen, weil dann jeder in dem Team das gleiche Verständnis hat und sozusagen du die Person selbst stärkst, die Rolle der datenarbeitenden Person innerhalb einer Organisation stärken musst.
20	112816_Transcript _Expert1_2	Drivers and Barriers of Data Literacy	Das bedeutet, dass die Person, was man mit Daten erarbeitet, muss im Team-Konsens anerkannt sein, aber du brauchst natürlich sozusagen auf der Management- oder CEO – Ebene, brauchst du Rückenwind dafür, dass du das auch machen darfst. Das du dir bestimmte Datenskills letztendlich antrainieren darfst.
21	112816_Transcript _Expert1_2	Drivers and Barriers of Data Literacy	Das ist wirklich wichtig und das merken wir auch bei uns in dem Projekt. Dadurch, dass es diesen Rückhalt durch die Geschäftsführung so stark gibt, sagt man auch: hey das ist wichtig für uns, setzt euch damit auseinander und das ist dann natürlich auch einfacher für euch.
22	112816_Transcript _Expert1_2	Problem Definition\Improvement Entity	was wir halt machen ist schon teamgebunden, aber zielt eher auf die Einzelskills der Personen ab
23	112816_Transcript _Expert1_2	Problem Definition\Improvement Entity	wenn man so ein grundlegendes Verständnis in einem Team hat, was Daten sind, oder wie man mit Daten arbeitet, könne sich überhaupt erst Prozesse innerhalb einer Organisation etablieren.
24	112816_Transcript _Expert1_2	Problem Definition\Improvement Entity	Das schafft also auch eine ganz andere Kommunikationskultur und Basis innerhalb einer Organisation, aber ich glaube das kann man erst machen, wenn man bei den Einzelpersonen dieses grundlegende Verständnis hast was es bedeutet.
25	112816_Transcript _Expert1_2	Problem Definition\Model Purpose	sehr stark davon abhängt in welchen Ländergrenzen man denkt
26	112816_Transcript _Expert1_2	Problem Definition\Model Purpose	Deshalb würde es mich voll interessieren, dass erstmal für eine Organisation zu machen. Und zu sagen: das hier sind die Kriterien, das sind die einzelnen Stufen und diese weiter ausarbeiten, weil ich glaube man kann das dann gut auf andere Organisationen ausweiten, weil ich bezweifle, zumindest im NGO Bereich
27	112816_Transcript _Expert1_2	Problem Definition\Model Purpose	NGO Bereich ist es halt ein neues Thema und da muss man sozusagen kleiner ansetzen, wenn man sich Organisationen anguckt

28	112816_Transcript_Expert1_2	Problem Definition\Model Purpose	muss man sich Gedanken machen und klar wenn du dann sagst, ok, wir gucken uns vielleicht irgendwie Organisationen an, die schon mit Daten arbeiten und das Veröffentlichen und da ganz viele Sachen schon erkannt haben und für dich nutzen, dann klar wäre das natürlich super interessant zu sagen, als so eine weiterführende, oder besser so einen Ausblick, wo man sagt, hier ist das Modell, wie wird das anhand von einer lernenden Organisation formuliert und das sieht so und so aus für eine Organisation, die das bereits nutzt und daraus könnte man dann sicherlich auch eine eigene Arbeit drüber schreiben.
29	112816_Transcript_Expert1_2	Problem Definition\Model Purpose	Aber ersteinmal würdest du schon sagen, dass es darum geht das Thema erstmal breiter zu machen? IP1: Ja, genau erstmal überhaupt das Thema Daten. Also weg von diesem ganz Abstrakten: was sind Daten.
30	112816_Transcript_Expert1_2	Problem Definition\Model Purpose	Quasi ein Common Knowledge zu definieren und darauf dann aufzubauen.
31	112816_Transcript_Expert1_2	Problem Definition\Model Purpose	Ja, also das kann ich so unterschrieben, das sehe ich ähnlich.
32	112816_Transcript_Expert1_2	Dimensions	alles was bei Description steht passt total gut dazu
33	112816_Transcript_Expert1_2	Dimensions	Das sind definiert Kompetenzen sozusagen, die man haben sollte. Ich fand ein bisschen die Ordnung durcheinander.
34	112816_Transcript_Expert1_2	Dimensions	Weil Data Output ist zum Beispiel etwas, was ganz weit hinten ist. Und ganz vorne braucht es noch so ein paar andere Kompetenzen.
35	112816_Transcript_Expert1_2	Dimensions	Die Frage ist natürlich auch so ein bisschen, ob man das anhand der Data Pipeline auf. Die Schritte wie ich mit Daten arbeite?
36	112816_Transcript_Expert1_2	Dimensions	Genau, also ich glaube auch man müsste das so ein bisschen vielleicht die Data Pipeline in diese Kompetenzen, oder andersrum diese Kompetenzen, die teilweise mit der Data Pipeline korrespondieren, dass man die der Data Pipeline zuordnen, weil die finde ich halt methodisch einfach total gut, wenn man halt sagen kann, relativ klar, welche Kompetenzen genau damit gemeint sind.
37	112816_Transcript_Expert1_2	Dimensions	Ja, wir machen die etwas ausführlichere Variante der Pipeline. Also hier: (see picture). Genau und jetzt können wir ja die Punkte von dir etwas ordnen und der Pipeline zuordnen.
38	112816_Transcript_Expert1_2	Dimensions	
39	112816_Transcript_Expert1_2	Dimensions	1)Ask question / define 2) Find 3) Get 4) Verify 5) Clean 6) Analyze 7) Visualize 8) Communicate
40	112816_Transcript_Expert1_2	Dimensions	Find ich auch gut, dass die letzten beiden Punkte getrennt sind.
41	112816_Transcript_Expert1_2	Dimensions	Das ist die offizielle Data Pipeline und das ist sozusagen das Gerüst was eigentlich für jedes Projekt was in irgendeiner Form Daten nutzt übertragbar ist. Man kann eigentlich bei jedem 1 bis 8, kann man ganz genau sagen, oder relativ genau sagen welche Skills dazu gehören. Und ich würde vielleicht nochmal probieren, diese Kompetenzen, dazu zu ordnen. Dass man sagt, Punkte 1-8 sond sozusagen die Data Pipeline und die Kompetenzen, die hier so ein bisschen anders beschrieben sind, ordnet man darunter oder macht Ergänzungen.

42	112816_Transcript_Expert1_2	Dimensions\Individual\Find	also Data Sources, know how to find data, das wäre find, das ist 2.
43	112816_Transcript_Expert1_2	Dimensions\Individual\Verify	Data Thinking würde ich dann bei verify eigentlich unterordnen.
44	112816_Transcript_Expert1_2	Dimensions\Individual\Verify	Ja, genau das hab ich hier auch. Know how to verify data.
45	112816_Transcript_Expert1_2	Dimensions\Individual\Verify	bei Data Thinking, da hab ich noch: being able to argue through and about data and engage in debates. Genau also nicht nur sozusagen, die Quelle zu benennen, sondern tatsächlich mit Daten zu argumentieren. Also eigentlich schon was Konkretes. Du nutzt Statistik oder Zahlen, oder Beweise, Fakten um damit zu argumentieren. Oder halt Sachen zu widerlegen. Das finde ich ist auch Data Thinking für mich.
46	112816_Transcript_Expert1_2	Dimensions\Individual\Verify	Ja, genau. Ich hätte sonst gedacht, dass könnte man sonst auch…in importance of finding stories and other journalistics output. So ein bisschen, was mach ich, wie setze ich das wieder in einen Kontext so ein bisschen. Aber man kann es auch unter verify einordnen.
47	112816_Transcript_Expert1_2	Dimensions\Individual\Communicate	Das eine ist eine Frage zu formulieren und das andere ist storytelling zu machen. Das ist glaube ich etwas, was schon fortschrittlicher ist. Da musst du schon viel mehr. Da musst du Daten im Kontext visualisieren, miteinander verknüpfen können, da musst du Fragen stellen können, mehrere.
48	112816_Transcript_Expert1_2	Dimensions\Individual\Ask question / Define	Also Data Inquiry, ask question s tot he data and find answers, das ist aufjedenfall Nummer eins. Ask questions oder define questions
49	112816_Transcript_Expert1_2	Dimensions\Individual\Ask question / Define	Ich würde bei dem Ask questions, define…also was ich glaube was wichtig ist, ist zu wissen was für Fragen habe ich und worauf können mir meine Daten eigentlich Antworten geben. Das kommt natürlich noch nicht bevor ich meine Daten gefunden habe, aber das ist mir nochmal aufgefallen, weil gerade wenn man nicht soooo häufig mit Datensätzen arbeitet und da quasi Informationen herausziehen oder Aussagen möchte, habe ich häufig erlebt, dass Leute sagen: das und das und das möchten wir erkennen, aber man muss eigentlich ganz klar sehen, dass die Daten, die ich habe, nur eine ganz spezielle Frage beantworten und dafür ein Gefühl zu haben ist wichtig.
50	112816_Transcript_Expert1_2	Dimensions\Individual\Ask question / Define	Aber das könnte ja auch bedeuten, dass es ein höheres Kompetenzlevel bedeutet. Also dass man die Multidimensionalität der Daten erkennt und dementsprechend dann auch Fragen stellt.
51	112816_Transcript_Expert1_2	Dimensions\Individual\Ask question / Define	Ja, stimmt auch, genau.
52	112816_Transcript_Expert1_2	Dimensions\Individual\Ask question / Define	Also so dass man sich halt keine Sachen ausdenkt oder hineininterpretiert. IP2: Genau, dieses überinterpretieren.
53	112816_Transcript_Expert1_2	Dimensions	Ok, dann haben wir Data Outputs ist ja irgendwas was….genau Stories oder Visualisierungen in Data.
54	112816_Transcript_Expert1_2	Dimensions	Das bedeutet du hast deine Kompetenzen nicht nur die Story oder die Frage zu finden, sondern deine Kompetenz ist ja eigentlich die Zahlen zu einer Story zu verarbeiten oder eben zu einer Visualisierung zu verarbeiten.
55	112816_Transcript_Expert1_2	Dimensions	Ich würde es auch wirklich nochmal trennen vom Visualisieren, also das Storyfinden und erzählen.
56	112816_Transcript_Expert1_2	Dimensions	Jaaa.
57	112816_Transcript_Expert1_2	Dimensions\Individual\Visualize	ch finde den Punkt visualisieren ganz häufig brauchen praktische Skills und Wissen. Also ein Pie Chart oder so. Das funktioniert nicht für 30 verschiedene Kategorien oder solche Sachen.
58	112816_Transcript_Expert1_2	Dimensions	Also deine Fähigkeiten deine Daten zu visualisieren und Storys zu erzählen. Das finde ich auch einen wichtigen Punkt, weil man denkt halt, Visualisierung = Storytelling, aber eigentlich sind die Begleittexte viel mehr Storytelling als diese zwei Kuchendiagramme, die du da reinbaust.
59	112816_Transcript_Expert1_2	Dimensions\Individual\Communicate	die Begleittexte viel mehr Storytelling als diese zwei Kuchendiagramme, die du da reinbaust

60	112816_Transcript_Expert1_2	Dimensions	Und du kannst ja häufig, als es gibt ja auch ganz viele Design Agenturen, die sich darauf spezifizieren Infografiken zu erstellen oder wirklich irgendwie schöne Diagramme zu bearbeiten eher in so eine Richtung, aber das heisst dann auch noch nicht, dass sie wirklich Geschichten mit Daten erzählen. Oft beim Visualisieren ist es dieses handwerkliche Zeugs, was als Skills mit reinkommt.
61	112816_Transcript_Expert1_2	Dimensions\Individual\Visualize	Oft beim Visualisieren ist es dieses handwerkliche Zeugs, was als Skills mit reinkommt.
62	112816_Transcript_Expert1_2	Dimensions\Individual\Communicate	Aber würdet ihr dann sagen, dass Storytelling dann eher bei communicate reinkommt? IP1 + 2: Ja, genau
63	112816_Transcript_Expert1_2	Dimensions	Data Objectives – link between data and action is evident.
64	112816_Transcript_Expert1_2	Dimensions	Ja, also das war ein Punkt wo ich mir nicht ganz sicher war, wo. Also ich meine, was meinen wir damit? Link between data and action was evident.
65	112816_Transcript_Expert1_2	Dimensions	Ja, genau wenn man das so als Metaebene, wozu mache ich das eigentlich alles nimmt, genau dann würde man das wahrscheinlich auch bei ask a question einbringen.
66	112816_Transcript_Expert1_2	Dimensions\Organizational\Data Culture	Oder man macht das wirklich so ein bisschen außerhalb von der Data Pipeline. Und packt das so ein bisschen zusammen mit Data Culture, weil da weiß ich noch nicht genau wie man das in den Prozess, wie arbeite ich mit Daten direkt reinbringe.
67	112816_Transcript_Expert1_2	Dimensions\Organizational\Data Culture	Genau und Data Culture ist ja irgendwie so ein bisschen Überbau, oder? Zu dem Ganze. Dass man sagt, in jedem einzelnen Schritt, mit jedem einzelnen Schritt bin ich einigermaßen vertraut, hab das vielleicht mal selbst ausprobiert, oder weiß was darin dann meine Schwächen und Stärken sind…
68	112816_Transcript_Expert1_2	Dimensions\Organizational\Data Culture	IP2: Also eine interessante Frage wäre halt auch, abgesehen wie man das dann aufbaut, zu gucken, ist das so wenn ich bei dem Maturity Framework ziemlich hoch score, viele hohe Werte habe in den verschiedenen Dimensionen, fühle ich mich dann auch direkt besser im Umgang mit Daten? Oder kann es vielleicht auch sein: Ich kann das, aber irgendwie mach ich das trotzdem nicht gerne. IP1: Ja, ja, genau. IP2: Ja, also das wäre interessant zu sehen. Wahrscheinlich würde man da einen Trend erkennen.
69	112816_Transcript_Expert1_2	Dimensions\Organizational\Data Culture	Ne, ich fand das eigentlich ganz gut, ich habe nur noch kurz überlegt: Data Culture, könnte das so ein bisschen gegen gesellschaftliche….Also es gibt z.B. ein Allgemeinmodell: Daten klar, dazu muss man Informatik studiert haben. Ob sozusagen, Data Culture auch so das Verständnis ist, dass man damit auch arbeiten kann ohne dass man Informatik studiert hat oder ohne dass man irgendwie ein Wetterforscher ist und nur der damit Wetterdaten was macht. Also so bestimmte Annahmen auch.
70	112816_Transcript_Expert1_2	Dimensions\Organizational\Data Culture	Dass bedeutet da nochmal das Verständnis der Daten, was ich habe, was ich in der Arbeit nutze. Also wenn das sowas übergeordnetes ist, dann wäre das für mich so ein Punkt 0,
71	112816_Transcript_Expert1_2	Dimensions\Organizational\Data Culture	Also so ein ganz einfach: ich weiß was Daten sind, ich habe keine Angst davor.
72	112816_Transcript_Expert1_2	Dimensions\Individual	Was ich sonst noch dachte, wäre vielleicht eine Möglichkeit, dass man sagt, das hier sind eher Skills, die die Person angeht
73	112816_Transcript_Expert1_2	Dimensions\Organizational	und so Sachen wie die Datenkultur oder Data Culture, dass das auf Organisationsebene stattfindet.
74	112816_Transcript_Expert1_2	Dimensions\Organizational	Wie ist in unserer Organisation quasi die Kultur mit Daten? Ist das vielleicht häufig ein Gesprächsthema? Wird bei uns Datenschutz sehr hoch geschrieben? Also wie ist so das Umfeld was halt dann die Organisation hat?
75	112816_Transcript_Expert1_2	Dimensions\Organizational	Also das wäre eigentlich sozusagen nicht die Data Pipeline, sondern es gibt innerhalb der Organisation strategische Schritte, die beschreiben wie mit Daten gearbeitet wird.

76	112816_Transcript_Expert1_2	Dimensions\Organizational	Also wenn man sich sozusagen einen Lebenszyklus von Daten denkt von ich sammele die, ich lagere sie, welche Leute haben darauf Zugriff? Wie werden die geteilt? Über welche Kanäle? Download oder nicht?
77	112816_Transcript_Expert1_2	Dimensions\Organizational	Aber das ist sozusagen sowas Verschriftliches gibt, wo man sieht: wie wird damit umgegangen? Wie wird z.B. mit sensiblen Daten umgegangen? Gibt es dafür ein Bewusstsein innerhalb der Organisation? Es hat ja nicht unbedingt was damit zu tun wie arbeite, was sind die einzelnen Schritte, die ich da abklappern muss um ein Datenprojekt aufzubauen, sondern dass es so Organisationswissen ist.
78	112816_Transcript_Expert1_2	Dimensions\Individual\Analyze	Ich hatte Data Analysis und das ist natürlich analyse,
79	112816_Transcript_Expert1_2	Dimensions	Wo wir das hier schon so grob aufgeteilt haben. Und dann hatte ich nochmal sowas: ability to use and understand basic code. Ist halt sozusagen, ist das hier auch etwas wo man sagen muss es gehört da irgendwie zu Data Literacy dazu? Also das ist mir eher so am Rande als großes Fragezeichen gekommen, da bin ich mir sehr unsicher, da Daten ja auch verarbeitet werden in Applicationen und Ähnliches, ist das natürlich so ein, überspitzt gesagt Algorithmus-Awareness, gehört das dazu?
80	112816_Transcript_Expert1_2	Dimensions	Ich glaube das könnte man gut ansetzen wenn man dann unterscheiden möchte zwischen einem Data Freshman und jemandem, der irgendwie natural ist oder so. Wenn man da nochmal Abstufungen macht.
81	112816_Transcript_Expert1_2	Dimensions\Individual\Get	Get data – Mach ich das nur so dass ich mir etwas aus einem Datenportal ziehe oder kann ich mir Scraper schreiben
82	112816_Transcript_Expert1_2	Dimensions\Individual\Visualize	auch bei Visualisierung. Kann ich einfach mit Excel ein Diagramm machen, oder kann ich auch in mein HTML Code einbetten
83	112816_Transcript_Expert1_2	Dimensions	Also ich finde, dass das auch zu einem gewissen Grad dazu gehört. Dass man Code verstehen kann. Also ganz ganz basic, oder wie Code aufgebaut ist.
84	112816_Transcript_Expert1_2	Dimensions\Individual\Verify	Das ist natürlich auch etwas: Selbstkontrolle, ich weiß was mit den Informationen anzufangen.
85	112816_Transcript_Expert1_2	Dimensions\Individual\Verify	Das wäre dann wieder in Richtung verify, kritisches Nachdenken darüber
86	112816_Transcript_Expert1_2	Dimensions\Individual\Communicate	Also ich würde das nicht auf Journalisten beziehen. Also bei Data Outputs. Da haben wir other journalistic Outputs. Ich glaube School of Data arbeitet viel mit Journalisten und bei einigen Mitgliedern ist Datenjournalismus ein großes Thema, aber ich glaube, dass man das streichen kann.
87	112816_Transcript_Expert1_2	Dimensions\Organizational\Data Ethics and Security	Ich habe noch zwei Punkte, die ich total interessant und wichtig finde, die da noch nicht auftauchen und zwar: Data Ethics.
88	112816_Transcript_Expert1_2	Dimensions\Organizational\Data Ethics and Security	Also nicht ganz an den Anfang, aber Punkt 01, eher auf Organisationsebene.
89	112816_Transcript_Expert1_2	Dimensions\Organizational\Data Ethics and Security	Und auch Data Security. Weil ich glaube die muss man einzeln behandeln, weil das eben aus menschenrechtlicher Sicht total relevant ist. Und ich mir da sehr sehr unsicher bin, ob wir da verstanden haben was das eigentlich heißt.
90	112816_Transcript_Expert1_2	Dimensions\Organizational\Data Ethics and Security	Und da gehört sowas hin wie, so ein Bewusstsein wie Datensparsamkeit, was heißt das eigentlich? Es ist total fesh irgendwelche Nutzergeschichten per Formular abzufragen als Organisation, weil es ist doch toll wenn ich so ein Personenprofil habe, aber geht das mit meinen Werten einher? Was heißt das sozusagen auf Datenebene?
91	112816_Transcript_Expert1_2	Dimensions\Organizational\Data Ethics and Security	also dieses: resonsible data, data ethics, was da so reinspielt.
92	112816_Transcript_Expert1_2	Dimensions\Organizational\Data Ethics and Security	Ja, genau.

93	112816_Transcript _Expert1_2	Dimensions\Organizational\Data Ethics and Security	Data Security eher, also was du meintest die Konsumenten von daten und die Sparsamkeit und Data Ethics eher, die Methoden, die man benutzt um zum Beispiel Daten zu analysieren, dass die transparent sind? IP1: Ja, ich glaube es ist etwas was viel stärker sozusagen auf Organisationsebene
94	112816_Transcript _Expert1_2	Dimensions\Organizational\Data Ethics and Security	Beispiel Oxfam, die haben eine responsible Data Policy als Organisation für sich definiert, wo Grundwerte drin stehen, die definieren, wie Oxfam als Organisation mit Daten arbeitet und auf welche Basis sie da Wert legen
95	112816_Transcript _Expert1_2	Dimensions\Organizational\Data Ethics and Security	wir teilen die Daten innerhalb der Organisation und so, aber gleichzeitig auch nicht alle Daten, sondern da ist auch wieder ein ethischer Umgang damit.
96	112816_Transcript _Expert1_2	Maturity Levels	Genau, und was natürlich ganz schön auffällt bei diesem, das hier was wir gerade machen und halt sozusagen dem gesamten Maturity Modell ist, dass es eigentlich, dass es sozusagen read and write skills sind im Großen und Ganzen. Also erst entwickelst du reading skills und dann entwickelst du writing skills. Genau und das dazwischen ist dann sozusagen …..ja, also das ist dann so ein Pfeil von oben nach unten. Weil reading wäre sozusagen das Verständnis usw. ich kann das lesen, ich kann mich damit auseinander setzen, ich kann das kritisch hinterfragen und writing ist: ich mach jetzt selber was und ich glaube das ist so ein Gefälle.
97	112816_Transcript _Expert1_2	Maturity Levels	Also ich glaube man müsste diese Level nochmal irgendwie definieren bevor man die mit einem Begriff versieht.
98	112816_Transcript _Expert1_2	Maturity Levels\Level 1	Also ich hatte überlegt, Level 0 wäre eigentlich so eine analoge Phase. Also wo du eigentlich Leute erstmal pushen musst damit sie erkennen, dass es Daten gibt und dass es irgendwie Zukunftsrelevanz hat.
99	112816_Transcript _Expert1_2	Maturity Levels\Level 1	Also ich find das eigentlich ganz gut diese Unsatisfactory Performance,
100	112816_Transcript _Expert1_2	Maturity Levels\Level 1	Damit legt man halt auch irgendwie so eine Art Standard
101	112816_Transcript _Expert1_2	Maturity Levels\Level 1	Das heißt es ist eigentlich sozusagen eine dateninteressierte Person, eine digital arbeitende Person, die aber halt noch nicht so ganz weiß
102	112816_Transcript _Expert1_2	Maturity Levels\Level 1	die halt nicht genau weiß: Wo anfangen? Genau ich glaube das ist so der typische Fall.
103	112816_Transcript _Expert1_2	Maturity Levels\Level 1	Genau, dass wir halt von Leuten angesprochen werden, die sagen: super, interessant ich würde das total gern nutzen, aber ich weiß nicht wo ich anfangen kann, bis halt, das ist so der Knackpunk, ob excellent Performer, ob das halt, wer das dann ist.
104	112816_Transcript _Expert1_2	Additional Comments	Genau, da ist halt die Frage, kann man das an..Kann ich das dann wirklich an bestimmten Sachen festmachen? Das wäre glaube ich….Wenn ich dann sowas…wie kann ich diese Entscheidung, diese Unterschiede wirklich treffen? Ist es eine Selbsteinschätzung, kann ich halt bestimmte Fragen stellen und sagen: wenn du das und das kannst, dann bist du auf Level 4 oder 3 und da muss man dann natürlich auch sehen, dass das auch ein Wandel ist. Also vielleicht ist das was man in zehn Jahren oder in fünf Jahren können muss…werden dann ganz andere Skills, die man wieder für die einzelnen Level festlegen muss.
105	112816_Transcript _Expert1_2	Maturity Levels	Also das aufeinander aufbauende das fand ich total gut. Ich glaube man müsste es nochmal etwas detaillierter machen, dass man genau die Skills benennt. Also ich glaube es macht das so ein bisschen gröber zu fassen für die einzelnen Bereiche….und das war für mich total schlüssig das aufzubauen, so wie du das gemacht hast, ich habe ein paar Notizen.
106	112816_Transcript _Expert1_2	Maturity Levels	Aber genau, da wo es sozusagen drauf aufbaut, ist das so ein bisschen besser, also vague understanding of possible Data Sources. Was konkret heißt das? Heißt das, ich habe verstanden, dass es Datenportale sozusagen gibt? Oder hab ich jetzt…ja sozusagen wie umfangreich geht da sozusagen in dieser Ebene, die man betrachtet, wie umfangreich ist es und dann glaube ich ist es auch einfacher die zuzuordnen, zu dein einzelnen Level.

107	112816_Transcript _Expert1_2	Maturity Levels	Ja, oder halt bei Data Thinking: Critical Evaluation is vaguely defined, data quality is not assessed consistently. Also dann glaube ich muss man sich auch erstmal klar machen: Was heißt den dann Datenqualität, also was für Items sollte man da checken?
108	112816_Transcript _Expert1_2	Maturity Levels\Level 1	Level 1 Uncertainty finde ich ganz gut. Also beschreibt den Zustand ziemlich gut. Ich habe schon ein bisschen was begriffen aber ich weiß nicht wo ich anfangen soll, und bin in allen Schritten einfach komplett unsicher oder kenne diese Schritte einfach gar nicht. Das finde ich ist irgendwie Level 1.
109	112816_Transcript _Expert1_2	Maturity Levels\Level 2	Level 2….genau, Awakening fand ich irgendwie…hat irgenwie, war nicht so richtig…
110	112816_Transcript _Expert1_2	Maturity Levels\Level 2	Das klingt halt ganz schön
111	112816_Transcript _Expert1_2	Maturity Levels\Level 2	Ja, aber ich glaube Level 2 da passt so das Enlightenment schon besser.
112	112816_Transcript _Expert1_2	Maturity Levels\Level 2	Dass man sagt: ah ich hab schon ganz viel verstanden, auf den Metaphase, kann das jetzt vielleicht noch nicht total anwenden, oder so realization irgendwie sowas. Also ich hab Sachen begriffen und jetzt kann ich angehen, die einzelnen, mir anzutrainieren.
113	112816_Transcript _Expert1_2	Maturity Levels\Level 2	Ja, oder ich find auch so das dritte. Also Novice, so quasi ein bisschen Neuling, oder Freshman vielleicht. Das also ne, das ist ja so ein bisschen: ich hab das Verständnis, ich bin ein Anfänger, aber irgendwie…
114	112816_Transcript _Expert1_2	Maturity Levels\Level 2	Ja, ich glaube man kann beides nehmen für Level 2. Also sowohl, ich finde Novice einfach nicht so ein schöner Begriff, oder…ja.
115	112816_Transcript _Expert1_2	Maturity Levels\Level 3	Genau, und bei Level 3 habe ich wieder aus Level 4 genommen. Da finde ich Certainty eigentlich ganz gut.
116	112816_Transcript _Expert1_2	Maturity Levels\Level 3	Also man geht mit den einzelnen Schritten, sehr selbstsicher um, weiß sozusagen richtig einzuschätzen was man kann, was man nicht kann…also man muss ja nicht alles können, aber das finde ich ist so ein Kriterium Level 3 war für mich Certainty einfach mit.
117	112816_Transcript _Expert1_2	Maturity Levels\Level 4	Die beste Beschreibung für Level 4 wäre dann für mich wirklich Data Fluent, bedeuted, die Arbeitsschritte sind drin, ich kann sie im Schlaf…genau.
118	112816_Transcript _Expert1_2	Maturity Levels	Ja, ich meine das ist halt auch so ein bisschen die Frage, bis wohin geht's? Weil wo sage ich, du hast es erreicht, es gibt natürlich noch super viel mehr Sachen, also muss ich irgendwie binärcode in hexideximal umrechnen können, oder ist das halt, kann man auch sagen, hier ist vorher schon die Grenze und alles drüber ist dann gut.
119	112816_Transcript _Expert1_2	Additional Comments	Ich mein was man sich natürlich auch überlegen kann, ist dann so ein bisschen wen man das dann abfragt, so ein Double Check zu machen, auf der einen Seite: wie verordne ich mich selbst? Sehe ich mich als Anfänger, als Experte und sonst was und dann nochmal konkret abzufragen, ok: kannst du denn das und das so bisschen zwischen: wir legen ein paar Kriterien fest und wie sieht derjenige sich eigentlich selbst? Und dann ist es natürlich auch irgendwie interessant wenn man dadurch dann zumindest zu einer Testversion. Wenn man so ein Maturity Modell entwirft und das dann irgendwie testen würde mit ein paar Organisationen und dann auch sieht: stimmt das jetzt eigentlich überein, müsste ich viellecht meine Kriterien nochmals überdenken, oder haben wir hier doch den Fall, dass sich Leute immer falsch einschätzen, oder halt auch mal richtig einschätzen? Also das das wirklich miteinander korrespondiert und das würde dann auch einen guten Indikator dafür geben, dass das Modell wirklich passend ist.
120	112816_Transcript _Expert1_2	Maturity Levels\Level 1	Level 1 wäre Uncertainty
121	112816_Transcript _Expert1_2	Maturity Levels\Level 2	das Zweite wäre Enlightenment oder Realization
122	112816_Transcript _Expert1_2	Maturity Levels\Level 2	IP2: Dann wprde ich aber eher sagen Enlighenment
123	112816_Transcript _Expert1_2	Maturity Levels\Level 3	Dann Level 3 Certainty
124	112816_Transcript _Expert1_2	Maturity Levels\Level 4	Level 4 Data Fluency

125	112816_Transcript_Expert1_2	Maturity Levels	Genau und die Dimensionen wären ja von Reading Skills zu Writing Skills. Da finde ich das auch total schön, weil das passt halt einfach total gut. Level 1 bis Level 4.
126	112816_Transcript_Expert1_2	Dimensions	Also ich finde das erstmal total gut das Allgemeiner zu fassen. Selbst wenn wir sozusagen die Skills irgendwie versuchen so klar wie möglich zu definieren, sind die Grenzen sozusagen fließend. Deswegen finde ich so eine allgemeinere Definition, die darüber hängt total wichtig und darunter listet man vielleicht einige Kompetenzen irgendwie auf, die besonders relevant sind irgendwie für dann diese Punkte. Weil ich glaube vollständig, weiß ich nicht, wie vollständig man das machen kann.
127	112816_Transcript_Expert1_2	Additional Comments	Weil ich glaube vollständig, weiß ich nicht, wie vollständig man das machen kann. IP2: Das geht wahrscheinlich auch gar nicht.
128	112816_Transcript_Expert1_2	Dimensions	HS: Hab ich auch in der Arbeit geschrieben. Kommt auch sehr auf den Kontext and. IP1: Ja, genau das kommt ja auch darauf an wie groß und umfangreich diese Projekte sind und auf welcher Ebene du die machst. IP2: Also alles was man dann da drunter listet, kann man ja auch als Instrumental Variables beschreiben. Quasi Beispiele, oder wichtige Punkte, die als Instrumente herhalten um das einordnen zu können.
129	112816_Transcript_Expert1_2	Dimensions	Ich fand eigentlich alles agree, weil es eben sozusagen diese übergeordnete etwas allgemeingefasste Ebene war.
130	112816_Transcript_Expert1_2	Dimensions\Individual\Find	Data Sources habe ich dann bei Level 1 does not collect data.
131	112816_Transcript_Expert1_2	Dimensions\Individual\Verify	Genau, bei Data Thinking habe ich z.B. genau Organisation bei Level 1: Organisation stellt Daten immer selbst her also Evaluationen, Umfragen, Interviews, aber weiß halt nicht um das Potenzial oder weitere Nutzungformen dieser Daten das wäre vielleicht nochmal sowas was bei Level 1 ist.
132	112816_Transcript_Expert1_2	Dimensions\Individual\Ask question / Define	Data Inquiry – ability to find patterns hab ich nochmal. Also nicht so, dass man eine Frage oder so formulieren kann, sondern ein kleines bisschen darüber hinaus. Ich habe eine Frage und die Daten beantworten in irgendeiner Form diese Frage und erkenne ich Muster oder Tendenzen entlang meiner Fragestellung.
133	112816_Transcript_Expert1_2	Dimensions\Individual\Ask question / Define	IP2: Und quasi das Gegenteil: Abweichungen. Muster und Abweichungen, Trends.
134	112816_Transcript_Expert1_2	Dimensions\Organizational\Data Ethics and Security	know/ awareness for Data Security / Data Ethics. Ich publiziere z.B. irgendwas, eine Karte mit Daten, aber ich habe mir eigentlich nicht wirklich Gedanken darüber gematcht was daran hängt, tu es aber schon. Hab halt irgendwie, irgendwelche statistischen Sachen auf meiner Website, weiß ich jetzt nicht.
135	112816_Transcript_Expert1_2	Dimensions	Ja, also die Frage ist halt, also ich bin mir manchmal unsicher muss man find und get, muss man das trennen? Oder ist das quasi dann so ein bisschen von wegen: read and write? Also ich kann auf der einen Seite Sachen finden und kann's dann auch bekommen? Aber eigentlich glaube ich schon, dass das eine gute Struktur ist um das halt zu analysieren.
136	112816_Transcript_Expert1_2	Dimensions	IP1: Die zwei Literaturbeispiele fand ich auch echt gut. HS: Welche Beispiele? IP1: Die zwei nach Hillson und Argyris und Schön.
137	112816_Transcript_Expert1_2	Dimensions\Individual\Find	Also ja, ich hab hier nochmal so ein paar weiterführende Data Sources, also wäre dann halt find, wissen wer Daten sammelt generell, können Daten in Texten erkennen, also du hast irgendwie eine Website und kannst daraus definieren was das für Informationen sind, selbst wenn die halt irgendwie in Textblöcken oder so verwandelt sind.
138	112816_Transcript_Expert1_2	Dimensions\Individual\Ask question / Define	Level 4 habe ich dann bei Data Objectives, bei Level 4 Link between data and action is evident, das nochmal konkreter verfassen und irgendwas bei 4 müsste ersichtlich sein, dass es workflows gibt. Also dass man generell bei vier, dass man das nochmal nicht nur sozusagen auf einzelne Skills bezieht, sondern, dass dieser Aspekt flüssiges Arbeiten von einem zum nächsten Schritt, vor allem was auf Organisationsebene sehr entscheidend ist, diese Workflows zu verbessern und effizienter zu machen und Zeit sparen, oder

139	112816_Transcript _Expert1_2	Additional Comments	Aber das ist ja…wenn wir an einem Projekt arbeiten, dann schauen wir ja auch wie teilen wir das auf, was muss zuerst geschehen und was passiert danach, wer übernimmt die Verifizierung, wer die Visualisierung und so weiter…Obwohl man dann natürlich auch wieder so ein bisschen auf Organisationsebene ist, glaube ich. Weil es dann ja wieder das Zusammenspiel zwischen einzelnen Individuen sein kann. Also ich glaube da muss man einfach so ein bisschen gucken wie kann ich das irgendwie gut in dem Fragebogen, wenn er sich nur an Individuen richtet, wie kann ich das da gut abfragen.
140	112816_Transcript _Expert1_2	Additional Comments	Ich würde das tatsächlich auch erstmal so auf Individuenebene übertragen und dann sagen: ok das hier sind Gemeinsamkeiten, die für eine Organisation gelten oder so. Also du brauchst ja dann so einen kleinsten, gemeinsamsten Nenner, wo du sagen kannst, wo sich die Organisation als Ganzes befindet. Genau, also ich glaube genau diese Skills und was für uns dann halt weiterführend letztendlich noch gefragt ist, sind die Maßnahmen. Also weil wenn wir das dann definiert haben, können wir ja sagen: wenn das und das die wichtigsten Skills sind, dann sind dafür die und die Maßnahmen und diese Skills irgendwie oft zu dran zu arbeiten oder zu trainieren
141	112816_Transcript _Expert1_2	Additional Comments	Ich mein auch wenn es um diese Workflows geht, könnte man auch sagen: ok, vielleicht sind wir jetzt an einem Punkt, wir teilen uns auf, aber wir haben noch kein wirkliches Management Tool dafür
142	112816_Transcript _Expert1_2	Additional Comments	IP1: Tools sind glaube ich dafür auch interessant. Also nicht nur: ich nutze jetzt irgendwelche Tools, sondern gibt es überhaupt so einen gewissen, nutzt eine Organisation in bestimmte Abteilungen oder Aufgabenbereichen Tools, um strukturierter zu kommunizieren, Projektmanagement zu machen…..automatisiert irgendwelche Sachen abzufragen statt ständig händisch zu recherchieren.
143	112816_Transcript _Expert1_2	Dimensions\Individual\Cl ean	nutze ich vielleicht openrefine um Daten zu cleanen
144	112816_Transcript _Expert1_2	Additional Comments	IP1: Und was dann sozusagen allgemeingültig ist. Weil wenn man sich die Data Pipeline anguckt: natürlich gibt es ein Tool für einen Schritt was sich gut eignet, was aber eine andere Person anders nutzt, oder ein anderes Tool nutzt und die sind ja auch meistens austauschbar. Und ja das ist ja auch eine andere Diskussion. Aber dass man sagt, dass es Prozesse gibt, oder Challenges innerhalb einer Organisation, eben auf übergreifender Ebene gibt, wo man eben Tools nutzt und einsetzt als Hilfestellung, zur Prozessoptimierung, zum Workflowoptimieren einsetzen.
145	112816_Transcript _Expert1_2	Dimensions\Individual\Ge t	Ok also, ich meine mit get wenn man da anfängt Daten zu bekommen wenn es quasi… IP1: Also du siehst quasi eine Tabelle online irgendwie und machst dir Gedanken, wie kann ich diese Tabelle jetzt abholen? Ist es copy/paste oder… IP2: ….screenshot und abtippen.
146	112816_Transcript _Expert1_2	Dimensions\Individual\Ge t	IP1: Genau, oder kann ich das downloaden oder kann ich dazu eine IFG Anfrage stellen, weil ich weiß, dass Behörden diese Form von Daten mir zusenden müssen, gibt es sozusagen…nutze ich eine Plattform um Daten irgendwo anders anzufragen, weiß ich grundsätzlich an welche Stelle ich mich wenden muss, die diese Daten sammeln? Damit ich die überhaupt erst anfragen kann?
147	112816_Transcript _Expert1_2	Dimensions\Individual\Fi nd	Aber das wäre ja fast auch so ein bisschen im Bereich find. IP1: Ja, stimmt das ist etwas blurry. IP2: Ja, also wenn ich so sage: gewisse Date, dass diese Anfragen, wo frage ich an, das gehört eher zu find.
148	112816_Transcript _Expert1_2	Dimensions\Individual\Ge t	live ticker, irgendwas passiert und ich möchte das automatisiert anzeigen lassen in meiner Email Nachricht und welche Tools nutze ich dafür, damit mir diese Informationen rausgefiltert werden vielleicht. Oder halt vielleicht würde man sich darüber hinaus: Wie komme ich an, ich bin besonders interessiert an Social Media Daten, wie komm ich denn jetzt an die Social Media Daten? Nutze ich Tools oder nutze ich ein Script, ich glaube so groß ist der Bereich gar nicht.

149	112816_Transcript _Expert1_2	Dimensions\Individual\Get	Ich glaube das ist eher so das Bewusstsein dafür was es halt alles gibt, wo ich anfragen kann und wie mir das sozusagen übersetzt wird, also als Text, Buch, Schrift, Tabelle sowas.
150	112816_Transcript _Expert1_2	Dimensions\Individual\Get	IP2: Ja, und ich glaube wirklich, dass wenn man dann zum Schluss weiß, wie ich mir ein Scraper schreiben kann oder solche Sachen um mir das zu suchen, ist das schon ein Indikator dafür, dass man sehr gut damit umgehen kann. Also ich mein das ist ja quasi ein Script wie man an Daten rankommt.
151	112816_Transcript _Expert1_2	Dimensions\Individual\Get	IP1: Das wäre dann schon wirklich sehr fortschrittlich. IP2: Ja, genau das wäre dann schon so das höchste Level was man erreichen könnte.
152	112816_Transcript _Expert1_2	Dimensions\Individual\Get	HS: Und Daten anfragen wäre dann schon eher niedrgier? IP1: Ja, genau.
153	112816_Transcript _Expert1_2	Dimensions\Individual\Get	IP2: Obwohl man muss auch sagen, wenn jemand weiß, wie man sich ein Script schreibt, das heißt nicht unbedingt, dass man über das IGF Bescheid weiß.
154	112816_Transcript _Expert1_2	Maturity Levels	IP1: Also man kann das sozusagen so formulieren, dass man dieses, mach ich das händisch, oder mach ich das nicht händisch, oder mach ich das automatisiert und wie komplex ist das was ich als Tool nutze um das zum Beispiel automatisiert abzufragen, vielleicht sind das so die….
155	112816_Transcript _Expert1_2	Maturity Levels	IP2: Ja, dass du sozusagen, ich mach das händisch oder ich benutze ein Tool, oder ich schreibe mir quasi selber meine Tools, oder vielleicht auch wenn ich da sowas wie ScrapingHub benutze, das ist dann ja wirklich to write. IP1: Ja, genau. Deswegen, eigentlich dieses Read à Write nicht von Oben nach Unten, sondern von links nach rechts.
156	112816_Transcript _Expert1_2	Maturity Levels	IP1: Ne, ich glaube tatsächlich eher von rechts nach links, weil zum Beispiel dieses get kannst du ja super technisch komplex gestalten, aber du kannst natürlich auch auf anderen Wegen deine Daten zusammen sammeln.
157	112816_Transcript _Expert1_2	Dimensions\Individual\Clean	P1: Clean, also ganz grundsätzlich glaube ich, weiß ich, also das Bewusstsein dafür, dass die Daten, die ich bekomme, dass die überhaupt gecleant werden müssen, dass sie in irgendeiner Weise vereinheitlicht werden müssen, gerade wenn ich eine Fragestellung habe und zwei Tabellen oder so habe, wie schaffe ich, also vereinheitlichen Daten auf einen kleinsten gemeinsamen Nenner zu bringen, damit sie vergleichbar sind so ein bisschen, also cleaning und weil wie gesagt, dass kommt wahrscheinlich eher so ein bisschen weiter hinten wenn man eben mit zwei Tabellen oder so…mehreren Tabellen arbeitet. Also so von der Kompetenz ein bisschen weiter hinten, aber clean, ich weiß halt, dass Daten, die ich bekomme, sind nicht perfekt und ich muss sie aufbereiten, ich weiß, was Kriterien sind, z.B. Maschinenlesbarkeit, Vereinheitlichung.
158	112816_Transcript _Expert1_2	Dimensions\Individual\Clean	IP2: Ja, und dann auch Vollständigkeit, ist ein wichtiger Punkt. IP1: Vollständigkeit, genau.
159	112816_Transcript _Expert1_2	Dimensions\Individual\Clean	bestimmte Datenformate und weiß wie das aufgebaut ist und danach cleane ich, oder bis hin zu Details. Ich weiß, das eine ist sozusagen die Awareness für die Kriterien: Daten müssen maschinenlesbar sein, müssen vollständig sein, etc. und der nächste Schritt ist eigentlich: ich kann das dann auch noch runterbrechen, bei Maschinenlesbarkeit weiß ich, ich brauche einheitliche Spalten und Zeilen.
160	112816_Transcript _Expert1_2	Dimensions\Individual\Clean	IP2: Ja, genau also auch dass ich so ein Verständnis habe: Es könnten Leerzeichen in den Spalten sein, also etwas, was ich nicht sehe.
161	112816_Transcript _Expert1_2	Dimensions\Individual\Clean	IP1: Also wenn ich einmal weiß, welche Kriterien es gibt und auf welche Besonderheiten ich achten muss, dann kann ich sie später natürlich auch besser cleanen.
162	112816_Transcript _Expert1_2	Dimensions\Individual\Clean	IP2: Genau, dann halt auch der Datentyp. Man kennt das ja aus Excel: du hast irgendwelche Zahlen, die als Datum formatiert sind, oder so. Und das zu erkennen und zu wissen: ah das passt hier gerade nicht überein, weil ich kann die Sachen nicht zusammenzählen, weil meine vier Spalten das falsche Datenformat haben. Und das dann auch wieder beheben zu können.

163	112816_Transcript _Expert1_2	Dimensions\Individual\Clean	HS: Also würdest du auch wieder dieser Logik folgen: Read: man erkennt das erstmal und Write: man ist dann auch in der Lage selbst daran was zu ändern. IP2: Ja, genau.
164	112816_Transcript _Expert1_2	Problem Definition	IP1: Also Eco-System: ist für mich super entscheidend und ich glaube das ist so der Rahmen um halt Individuen entweder zu motivieren, oder zu demotivieren.
165	112816_Transcript _Expert1_2	Drivers and Barriers of Data Literacy	Für mich also die wichtigsten Aspekte sozusagen innerhalb eine Organisation, ist Zeit, also das bedeutet: ich habe neben meinem Tagesgeschäft habe ich Zeit mir Sachen als Weiterbildung oder was auch immer anzulesen, anzutrainieren und auch auszuprobieren. Ich habe sozusagen in meinem normalen Organisationsworkflow habe ich Punkte wo ich sagen kann, hey kann ich irgendwie mal Daten mit angucken, oder so. Bei der Recherche, bei der Kommunikation, bei der Evaluation…
166	112816_Transcript _Expert1_2	Drivers and Barriers of Data Literacy	IP1: Zeit deswegen, weil wenn man anfängt, braucht man Zeit, weil man aus Fehlern lernen muss.
167	112816_Transcript _Expert1_2	Drivers and Barriers of Data Literacy	Das ist halt so das Zweite: Experimentierraum. Also in irgendeiner Form eine Kultur entwickeln, wo es dann erlaubt ist Fehler zu machen, erlaubt ist zu experimentieren, erlaubt ist Fragen zu stellen und sich überhaupt mit einem neuen Umfeld oder Thema auseinander zu setzen.
168	112816_Transcript _Expert1_2	Drivers and Barriers of Data Literacy	IP2: Und dann auch ganz klar, wenn gerade von der Führungsebene, oder höhere Positionen, das dann auch erlaubt wird oder gefördert wird.
169	112816_Transcript _Expert1_2	Drivers and Barriers of Data Literacy	IP1: Wertschätzung, Wertschätzung dann eigentlich für, dafür, dass du eben etwas Neues reinbringst und dass du dich bewegen kannst und dass du ein safe environment hast. Das bedeutet also nicht, dass du Angst haben musst: ich habe 1000 andere Sachen zu tun und mich dann damit gar nicht auseinandersetzen kann.
170	112816_Transcript _Expert1_2	Drivers and Barriers of Data Literacy	Oder ich mach Fehler und es wird was veröffentlicht und ich werde alleine dafür, also muss Rechenschaft dafür zu leisten, weil ich mich getraut habe etwas Neues auszuprobieren.
171	112816_Transcript _Expert1_2	Drivers and Barriers of Data Literacy	Klar, Geld hilft vielleicht Tools, weitere personelle Resources zu bekommen, Weiterbildungen und total grundsätzlich digitale Geräte..das haben wir ganz häufig mit Behörden…da gibt es sicher Parallelen, wenn du auf deinem Gerät keine Programme installieren kannst, dann bist du einfach, dann stimmt deine Umgebung nicht dafür um ein bisschen Experimentierraum zu haben.
172	112816_Transcript _Expert1_2	Drivers and Barriers of Data Literacy	Hardware muss da sein und ich muss eigenmächtig sagen können: brauch ich auch Programm, brauch ich ein Tool und dann hab ich halt keine Hürden, das jetzt zu besorgen und es kostet mich jetzt nicht noch mehr Aufwand, sondern ich mach das dann einfach.
173	112816_Transcript _Expert1_2	Drivers and Barriers of Data Literacy	Verständnis dafür, dass datengestütze Projekte auch einfach ins Nichts verlaufen können
174	112816_Transcript _Expert1_2	Drivers and Barriers of Data Literacy	Und dass wenn bestimmte Daten nicht vorhanden sind oder das Projekte dann dadurch sterben müssen, oder dürfen sozusagen. Weil das veraltet ist oder was auch immer, der plötzlich wird da bei Verwaltungen dieses Thema nicht mehr weiter gefördert oder was auch immer. So dieses Bewusstsein dafür, dass es auch mal ins Nichts führen kann
175	112816_Transcript _Expert1_2	Drivers and Barriers of Data Literacy	IP1: Genau, mn muss auch eine gewisse Flexibilität haben.
176	112816_Transcript _Expert1_2	Drivers and Barriers of Data Literacy	Oh ich habe mir das und das gedacht, und so soll das Projekt sein, aber dann haben wir die Daten geholt und dann haben wir festgestellt: oh das geht so nicht. Dann brauchen wir die Flexibilität um das Projekt neu zu gestalten oder umzufassen und die Fragestellung zu überarbeiten und zu überprüfen.

177	112816_Transcript _Expert1_2	Drivers and Barriers of Data Literacy	IP2: Also dann auch auf gesellschaftlicher Ebene. Also ich glaube natürlich Driver ist natürlich, wenn man so ein Verständnis herrscht, dass offenen Daten für die Zivilgesellschaft wichtig sind und das ist für NGOs natürlich auch ein super Ansporn um Daten zu nutzen
178	112816_Transcript _Expert1_2	Drivers and Barriers of Data Literacy	Das ist dann ja wieder dieses, was passiert wenn ich sozusagen als Organisation das geschafft habe Produkte zu veröffentlichen, datengestützte Produkte, sauberes Arbeiten, gehen wir mal davon aus: wir haben eine perfekt digitalarbeitende, datengetriebene NGO, aber die Gesellschaft ist einfach kein Data-fluent Consumer sozusagen und kann damit nicht wirklich was anfangen, oder man ist halt dann so ein Vorreiter und gibt es ja auch…im internationalen Bereich gibt es ja solche Vorreiter.
179	112816_Transcript _Expert1_2	Evaluation\Correctness	Ok, also correctness, also ob das Modell angewendet werden kann und ob die verschiedenen Level aufeinander aufbauen. HS: Genau, also correctness auch, ob der Inhalt korrekt ist und es richtig beschrieben ist.
180	112816_Transcript _Expert1_2	Evaluation\Correctness	IP1: Genau, also ich glaube, wenn wir diese einzelnen Schritte nehmen und das unterteilen auf Organisationsebene und wissen, uns was bei jeder einzelnen Person da sein muss, was dann so ein Organisationswissen ausmacht, dann glaube ich ist das total gut. Ich finde die Beschreibung, die du hier drin hast, finde ich super. Ich würde wie gesagt nur noch mal ein paar kleine Ergänzungen machen, aber darüber haben wir ja schon gesprochen und darunter dann nochmal etwas konkreter zu fassen, was das denn für Skills sein könnten, ohne den Anspruch, dass es vollständig ist, aber beispielhaft.
181	112816_Transcript _Expert1_2	Maturity Levels\Level 3	Aber bei Level 3 hatte ich mich gefragt, ob diese Person nicht eigentlich noch ein bisschen mehr als das was hier beschrieben wird. Um sie nochmal zu Level 2 und Level 4 viel stärker abzubilden.
182	112816_Transcript _Expert1_2	Evaluation\Correctness	IP1: Ja, aber ansonsten finde ich das total correct.
183	112816_Transcript _Expert1_2	Evaluation\Flexibility	IP2: Beim nächsten Punkt: ich glaube auch, dass man es sehr gut anpassen kann für verschiedene Organisationen oder für verschiedene Größen von Organisationen.
184	112816_Transcript _Expert1_2	Evaluation\Flexibility	Aber ja, Flexibilität ist gegeben, weil du dann an diesen Kriterien der Organisation oder Projektteams oder ganz egal, ob wir von einer Organisation mit 100 Menschen sprechen oder einer kleineren Organisation sprechen, die dann nur zwei kleine Projekte haben. Ich denke aber, dass es für beide gut anwendbar ist.
185	112816_Transcript _Expert1_2	Evaluation	IP2: Ich glaube auch, dass wenn man das noch etwas miterfasst, so die Organisationscharakteristika. Wenn man dann vergleichen möchte, dass man halt sieht: ok gibt es irgendeinen Punkt wo man Richtlinien hat, die vorgeben, welche Programme man installieren kann oder es andere Punkte gibt, wo du vom Management dann Infrastruktur vom Teilen von Daten oder bei kleineren Gruppen eben sagt: wir speichern das alles auf GoogleDrive oder so
186	112816_Transcript _Expert1_2	Evaluation\Usability	IP2: Ich glaube Usability: aufjedenfall, das kann man sogar für Datenschule und so und alle, die eigentlich so bisschen in Richtung technische Beratung oder den Leuten Wissen vermitteln möchte im technischen Bereich, da kann man das super benutzen
187	112816_Transcript _Expert1_2	Evaluation\Usability	IP1: Ich glaube es funktioniert super auf zwei Ebenen, zum einen wenn Externe, so wie wir dahin gehen uns sagen: hey wir wollen euch jetzt an ein neues Thema heranführen und zu eurem Wissenstand und daraus können wir dann Maßnahmen ergreifen, aber dann auch so für die interne Nutzung, wenn du eine Organisation hast und auf der anderen Seite, dass du als Organisation unabhängig von den Externen so deinen Stand anzeigen lassen kannst und eine Selbsteinschätzung machen kannst: Wo stehen wir gerade, was sind Bereiche, die wir strategisch verbessern können.
188	112816_Transcript _Expert1_2	Evaluation\Usability	Gerade das so als Selbsttest anzubieten ist eine super Sache.

189	112816_Transcript_Expert1_2	Evaluation\Implementability	Und ich glaube jetzt bei Implementability ist da noch so ein bisschen die Frage, wie kann man es dann gut aufbereiten diesen Fragebogen selbst, also das es halt nicht pro Person zwei Stunden dauert.
190	112816_Transcript_Expert1_2	Evaluation\Implementability	IP1: Ja, absolut.
191	112816_Transcript_Expert1_2	Evaluation\Implementability	IP2: Aber generell glaube ich, mit diesem Stufenmodell hier und der Data Pipeline und eigentlich das was du schon gemacht hast, wenn man das noch so ein bisschen mit den konkreteren Punkten verknüpft, sehe ich da keine Probleme. Und vor allen Dingen so ein Fragebogen ist dann auch nichts wie eine Umfrage oder so, die einen am Telefon nervt, oder so in einem Callcenter, sondern das ist ja was, was ja bestimmt eine bewusste Entscheidung ist: das würde ich hier jetzt selber für meine Organisation wissen und man erkennt was man als Organisation noch so braucht.
192	112816_Transcript_Expert1_2	Evaluation	IP1: Genau, also was halt, wo es halt dann auch noch interessant ist, ist die Punktezusammensetzung, wo du da sagst: ok da landest du mit diesen Skills anhand der Kriterien.
193	112816_Transcript_Expert1_2	Evaluation	IP2: Und ja genau, diese konkrete Punktezusammensetzung ist wirklich etwas, was kontext- und zeitabhängig ist.
194	112816_Transcript_Expert1_2	Evaluation	IP1: Aber das ist ja egal, weil es geht ja darum sozusagen ganz egal in welchem Level du dich befindest, dass du dich dann da verorten kannst und sagen kannst: Ok das sind die Skills wo ich mich noch weiter entwickeln kann und ich weiß nicht, ich sehe das tatsächlich nicht unbedingt.
195	112816_Transcript_Expert1_2	Evaluation	Ich glaube was man aufjedenfall mit bemerken könnte bei den konkreten Punkten, wenn die dann ausgearbeitet sind, dass das in 5 Jahren dann nicht mehr so aktuell ist.
196	112816_Transcript_Expert1_2	Evaluation\Economic Efficiency	IP1: Und das ist ja auch ein low-level Investment. Klar du hast ja eine Person, die sich 15 Minuten an einen Fragebogen setzt.
197	112816_Transcript_Expert1_2	Evaluation	Ne, also bei mir auch nicht. Es ist gut durchdacht und ja sehr schön schon ausgearbeitet.
198	112816_Transcript_Expert1_2	Evaluation	Obwohl vielleicht noch einen Gedanken: Das habe ich vorhin schon erwähnt, das mit den demographischen Abfragen, das würde ich noch mit überlegen. Vielleicht auch besondere Formen: was macht ein Verein aus etc. Da gibt's vielleicht noch bestimmte Sachen, die beachtet werden müssen, wenn ich so ein Maturity Framework entwickle. IP1: Altersstruktur.
199	120516_Transcript_Expert3	Dimensions	I think one big thing that is missing for me is, which is probably the core of what I understand data literacy to be is the ability to think critically and assess data. And I don't see that anywhere here. So I think that is probably missing, so from all the work I have done, you might be able to technically find or get or verify on the technical side, but if you can't critically assess what it means than there is no point in it.
200	120516_Transcript_Expert3	Dimensions\Individual\Verify	ng of the data pipeline is what you teach when you going to technical skills that are needed to work with data. Verify is this point of, for example knowing…seeing an image and being able to run it through Google Image Search and reverse Image search it and that kind of thing or getting a data set and being able to, like knowing what to check.
201	120516_Transcript_Expert3	Dimensions\Individual\Verify	That is what it says is in terms of…it says it's this data centric governing and knowing that you need governmental data portal and type in these things and make sure that it is there.
202	120516_Transcript_Expert3	Dimensions	Whereas, what I am talking about is more of an umbrella term for what is…like say someone presents you, say someone communicates these findings to you in a graph that says, you know 21782 Syrian were killed last year. You would need to understand that this is probably not true. And that is very different. That is the ability to assess visualizations and assess analogies and assess communications even without a technical analysis.

203	120516_Transcript _Expert3	Dimensions	I don’t know, I would say critical thinking is for me the kind of key of data literacy, so I guess first. ‘Cause you can have…you can be able to do all of the things you listed, but if you cannot do that, you did not really get that much.
204	120516_Transcript _Expert3	Dimensions\Organizational\Data Culture	you can be completely comfortable around working with data, but in your organization if you don’t have buy-in from senior management, if you don’t have time and budgets and project managers who understand that you need this time and the software capabilities and the access….and if not that you don’t really have a data culture.
205	120516_Transcript _Expert3	Dimensions\Organizational\Data Culture	So I’d say comfort around data is actually kind of different from what I understand as data culture that would actually help to work with it.
206	120516_Transcript _Expert3	Drivers and Barriers of Data Literacy	you can be completely comfortable around working with data, but in your organization if you don’t have buy-in from senior management, if you don’t have time and budgets and project managers who understand that you need this time and the software capabilities and the access….
207	120516_Transcript _Expert3	Dimensions	And I mean, I guess on the others, find, get, verify, clean, analyze, visualize, communicate, that is pipeline that we used for years. I mean you don’t have to be able to do all of them in any way to actually have data literacy, but I don’t know, as we teach it, you can step in at any level, and any step in the pipeline that kind of things.
208	120516_Transcript _Expert3	Additional Comments	For me, as I said that this kind of critical thinking around data is actually how I would define data literacy, more than any of the things you have listed here.
209	120516_Transcript _Expert3	Additional Comments	And I mean, I guess on the others, find, get, verify, clean, analyze, visualize, communicate, that is pipeline that we used for years. I mean you don’t have to be able to do all of them in any way to actually have data literacy, but I don’t know, as we teach it, you can step in at any level, and any step in the pipeline that kind of things.
210	120516_Transcript _Expert3	Additional Comments	Yea..I think it depends a little bit on what you are trying to do. So for example if I’m working with human rights defenders who are looking at getting data, their kind of measure of unsatisfactory to excellent is hugely different from when I would be working with people whose main field is like open data.
211	120516_Transcript _Expert3	Maturity Levels	Yea..I think it depends a little bit on what you are trying to do. So for example if I’m working with human rights defenders who are looking at getting data, their kind of measure of unsatisfactory to excellent is hugely different from when I would be working with people whose main field is like open data.
212	112816_Transcript _Expert1_2	Additional Comments	Was ich total schön finde ist die Metapher im englischen wird z.B. viel über Data Fluency gesprochen und ich finde das ist ein sehr sehr bezeichnender Begriff dafür, weil das eben etwas ist, was du wie eine Sprache lernst. Du lernst quasi das gleiche wie eine Sprache. Du fängst klein an, dann hast du deine Vokabeln, dann lernst du deine Vokabeln, dann übst du die Grammatik und bis dann tatsächlich bis zu einem ganz flüssigen Gebraucht und dann hast du nochmal so ein Level wo du sagst: Fachkenntnisse z.B. Business English. Hier haut man dann die ganzen krassen Dinger raus.
213	112816_Transcript _Expert1_2	Additional Comments	Und ich finde dieses Data Fluency das passt viel viel stärker.
214	112816_Transcript _Expert1_2	Additional Comments	Dass man sagt man braucht natürlich, sonst hat man dann daten-getriebene Produkte entwickelt oder daten-getriebene Use-Cases oder was auch immer, aber du hast eigentlich niemanden, der das tatsächlich lesen kann, der damit was anfangen kann. Und dann wäre halt auch nochmal dieses Gegenstück zu sagen, wie sieht das denn eigentlich mit der Audience aus? Wer sind diejenigen, die diese ganzen Produkte von den NGOs konsumieren sollen? Was muss man denen den geben damit sie die Produkte richtig lesen und verstehen können, dass das tatsächlich auch in so Action – Phase übersetzt wird. Weil du möchtest ja eben eine Aktion hervorrufen durch diese ganzen Produkte, die du da machst oder Kampagnen.

215	112816_Transcript _Expert1_2	Additional Comments	IP2: Bzw das auch wieder kritisch hinterfragen so wie quasi NGOs, oder wir den beibringen Daten, Fakten, Zahlen kritisch zu hinterfragen oder was du sagtest, das so darzustellen: hier haben wir die Daten her, das gibt's noch, daher haben wir unsere Erkenntnisse. Dass die Audience dann auch wieder so kritisch hinterfragen kann.
216	112816_Transcript _Expert1_2	Additional Comments	ok es gibt eine bestimmte Rollenverteilung. Also es gibt neue Rollenverteilungen innerhalb der Organisation, keine Ahnung. Z.B. hat man dann inhaltliche Experten, Editoren oder man hat einen Designer oder einen Entwickler im Team, oder man hat einen Kommunikationsmenschen im Team. Also daran, nach diesen Data Fluency Modellen, oder Vorstellungen, dass man da halt sozusagen auch Rollen einzelner Personen innerhalb der Organisation anders definiert oder zuteilt.
217	112816_Transcript _Expert1_2	Additional Comments	Vielleicht noch ein letzter Gedanke dazu. Was wir immer wieder bemerken, auch bei uns im Team, deswegen ist unser Team auch so wahnsinnig divers. Deswegen funktioniert OKF auch nicht nur als eine Entwicklerbude, oder halt nur Geisteswissenschafter oder so, das funktioniert halt nicht. Deswegen muss, ich glaube so gute datengestütze Projekte oder Produkte brauchen echt ein diverses Team.
218	120516_Transcript _Expert3	Maturity Levels	IP3: Yea, because, I mean for example, you said marginal performance was being able…what was the example you gave, being able to…what was it…design surveys. And require data from surveys, for some people is highly complex and there are entire research courses on survey science and not trivial at all. So, that being marginal is like….it depends very much on what you are trying to do, I guess. Or for example visualization, if you are working in an organization where you can pay for access to Tableau or you know, data visualization platforms, then your excellent performance might be unsatisfactory performance, because you can't code in D3.
219	120516_Transcript _Expert3	Additional Comments	IP3: Yea, because, I mean for example, you said marginal performance was being able…what was the example you gave, being able to…what was it…design surveys. And require data from surveys, for some people is highly complex and there are entire research courses on survey science and not trivial at all. So, that being marginal is like….it depends very much on what you are trying to do, I guess. Or for example visualization, if you are working in an organization where you can pay for access to Tableau or you know, data visualization platforms, then your excellent performance might be unsatisfactory performance, because you can't code in D3.
220	120516_Transcript _Expert3	Additional Comments	Because it looks like, from my experience, it depends very much on what people's aims are. If I am working with data with an aim on producing a rigorous and robust statistical analysis, then I look at this and I am like: Ok, I can do everything, but then like a statistician looks at my work and will be like: no, you did not do your survey design correctly. Whereas, when I am working in a small NGOs and I am just getting into working with data and I think, I can do surveys, it is like, as good as it needs to be for my purpose. There is an element of needing to know of what the aim is and what the data is going to be used for before you can…yes before you can prescriptive about what the categories are, if you get me.
221	120516_Transcript _Expert3	Dimensions\Organizational\Data Culture	I guess data culture, I mean that no one has computers or no one has digital literacy or no one understands, you know they just don't have access to that kind of thing and don't, they don't even know that it would be useful for them potentially.
222	120516_Transcript _Expert3	Additional Comments	I don't know, I think being general is actually not that helpful, because as I said if you have a small NGO that, it matters much more especially in the practical and NGO sense it matters so much more what you're saying that you are doing.
223	120516_Transcript _Expert3	Additional Comments	I think this whole thing needs a whole thing of having a layer of what they are saying what they are doing.
224	120516_Transcript _Expert3	Additional Comments	What are they trying to do and taking much more into account what are the aims of the people that you are assessing actually are,

225	120516_Transcript _Expert3	Additional Comments	Well I think really, for me it is so context specific
226	120516_Transcript _Expert3	Additional Comments	I think that is probably my biggest learning after years of working on data literacy stuff that all of this depends, it changes dramatically depending on the context.
227	120516_Transcript _Expert3	Additional Comments	So, like communication could be a really great infographic that you can, that is like five megabytes to download that looks beautiful. But if you want people who are meant to be looking at that, who don't have access to fast internet, it actually is completely unsatisfactory, because they cannot download the image. And I think lots of the problems that I have seen have come around people trying to fit in a model that isn't for them. That just makes no sense to them in their context and that is a something we have to be really really respectful of. Like recognize that not everything in this place can fit in to a general model, but that the context needs to lead.
228	120516_Transcript _Expert3	Additional Comments	What do you think could be a good alternative for such a model to educate people more on the topic? IP3: I guess something that allows them to critically assess what their needs are versus what they can do.
229	120516_Transcript _Expert3	Additional Comments	Yeah, I think watch more the context, prioritize the context what the organizations are trying to do and then assessing them according to that would make much more sense and is also more useful. So, finding a model of what are your aims, what are you trying to do, what are your needs in order to reach those things and what are the skills that you need in order to reach that and how are you according to those needs.
230	120516_Transcript _Expert3	Drivers and Barriers of Data Literacy	I mean drivers of most things can be quite funders-driven, so if the funders are saying the NGO should work more with data, then this might be a really good incentive.
231	120516_Transcript _Expert3	Drivers and Barriers of Data Literacy	I think everything you need is buy-in from the manager level, I mean you need a strategy that fits in the broader organizational strategy and then you need the resources to do what you are trying to do and unless you don't have all of these three, it's unlikely that you are getting successful in the organizational transformation.
232	120516_Transcript _Expert3	Drivers and Barriers of Data Literacy	But then, I guess you also need the digital peace, infrastructure to use, hardware and software, like access to the things you need to access the data, like a digital infrastructure
233	120516_Transcript _Expert3	Drivers and Barriers of Data Literacy	I guess data transformation or data literacy part is more around building in processes that it is only one project that uses data in that cool way, but rather, you know, it is an organizational-wide, institutional approach.
234	120516_Transcript _Expert3	Drivers and Barriers of Data Literacy	you have a look at civil society and what drives their data literacy, it would be especially funders that are important, needs from beneficiaries that they are working with
235	120516_Transcript _Expert3	Additional Comments	Yes, I think the biggest thing I would say is prioritizing contexts and that people have very different contexts. That data literacy in Germany for a big NGO could mean something wholly different for a big NGO in another country
236	120516_Transcript _Expert3	Additional Comments	Yes, I wonder whether your difficulty around the topic is because generalizations on the topic are not useful.
237	120516_Transcript _Expert3	Additional Comments	Yes, I think something like that could be really helpful, as long as it does not try to be too general so that everything has to fit in to the same model, or into the same categories in order to work with the model. And so that it could help to think through the questions and issues that they face in their context that leaving space for adaptation and flexibility.
238	120816_Transcript _Expert4	Dimensions\Organizational\Data Culture	dieser Begriff Data Culture ist ziemlich weit

239	120816_Transcript_Expert4	Dimensions\Organizational\Data Culture	Ich würde es etwas enger sehen. Schwierig. Ich hatte jetzt in dem was ich jetzt hier habe, hatte ich das so genommen, dass man sich Experten dazu holt sozusagen um eine Datenkultur intern zu kreieren. Das ist aber eigentlich nur ein Aspekt von dem Ding. Das Andere ist irgendwie wie zeigt sich das in der NGO selbst und mir ist da noch kein…damit hab ich aufjedenfall so meine Probleme das irgendwie zu definieren, was eigentlich data culture ist.
240	120816_Transcript_Expert4	Dimensions\Organizational\Data Ethics and Security	Data Ethics ist so das responsible Thema
241	120816_Transcript_Expert4	Dimensions	Ne, also das ist eigentlich ganz gut. Genau…ja ansonsten passt das alles so.
242	120816_Transcript_Expert4	Dimensions	HS: Und auch von den einzelnen Schritten? IP4: Ja, genau, das sind ja die klassischen Sachen. Teilweise ist es so, dass du Sachen überspringst. Also wenn die Daten schon clean sind, die du aus einem Datenportal geholt hast, dann brauchst du kein Datacleaning mehr.
243	120816_Transcript_Expert4	Dimensions	teilweise finde ich die Übergänge schwierig in dieser Data Pipeline, weil klassisch würde man von oben nach unten durchgehen, das ist aber oft nicht der Fall. Man macht dann eher…man überlegt sich: was möchte ich gerne machen, dann findet man ein paar Sachen, dann analysiert man ein paar Sachen und dann findet man neue Sachen und schmeißt das sozusagen um. Also ist das ganz gut, dass du das so nebeneinander nochmal hast. Aber es deckt sonst aufjedenfall alles ab, was man so macht.
244	120816_Transcript_Expert4	Dimensions\Organizational\Data Culture	was heißt Datenkultur, heißt das ich stelle eigene Sachen online? Ich glaube das wäre ein großer Aspekt.
245	120816_Transcript_Expert4	Dimensions\Organizational\Data Culture	auch so Transparenz. Ich arbeite transparent, weil ich meine Sachen auch so verfügbar mache, oder nutze ich Daten um Reports zu machen, oder…um einige Sachen zu unterfüttern und dann das dritte wäre das mit den Experten sozusagen. Dass ich mir Experten ins Team hole, oder Externe, die mit Daten arbeiten können und dann in diesem Zusammenhang arbeiten. Das würde dann darein fallen denke ich.
246	120816_Transcript_Expert4	Drivers and Barriers of Data Literacy	Ja, vielleicht auch von der Managementebene, dass es halt unterstützt wird. Das haben wir ja zum Beispiel bei der Kooperation mit der deutschen Bahn, da war's auch so, also das ist keine NGO, sondern ein Unternehmen, aber es kam halt von oben so die Ansage ja wir machen jetzt Open Data, aber es hat sich halt noch überhaupt nicht in den verschiedenen Abteilungen verbreitet.
247	120816_Transcript_Expert4	Drivers and Barriers of Data Literacy	Ja, aber war dann ganz cool. Es gab einen Hackathon und da wurden Anwendungen gebaut und die anderen Abteilungen haben gesehen: ahh da passiert ja was. Und jetzt kommt auch immer mehr und mehr und die machen jetzt jedes Jahr einen Hackathon wo es dann jedes Jahr neue Daten gibt.
248	120816_Transcript_Expert4	Dimensions\Organizational\Data Culture	Also ich hatte quasi bei Data Culture hatte ich im ersten Schritt: no employees with experience and data, das wäre also so der Anfangspunkt.
249	120816_Transcript_Expert4	Dimensions\Organizational\Data Culture	Genau dann holen sie sich quasi Experten, die Datenwissen haben und die dann ins Team reinkommen, das wäre dann so die erste Stufe sozusagen (Level2)
250	120816_Transcript_Expert4	Dimensions\Organizational\Data Culture	Die zweite Stufe (Level3) , da ist super kritisch, dass man die Experten, die man rangeholt hat, dass die im Team eingebunden sind sozusagen. Ich glaube das ist eine riesige Challenge.
251	120816_Transcript_Expert4	Dimensions\Organizational\Data Culture	Datenteam was, wenn man was braucht, aber die Challenge ist wirklich, dass auch die, die in diesen Datenthemen sind quasi an der normalen Sitzung teilnehmen und da wirklich in diesem Workflow drin sind, das finde ich wichtig
252	120816_Transcript_Expert4	Dimensions\Organizational\Data Culture	Die höchste Stufe wäre dann: central repository with searchable information.

253	120816_Transcript_Expert4	Dimensions\Organizational\Data Culture	Naja also man kann quasi online einfach nach Wissen suchen. Und alles Wissen was irgendwie in der NGO, oder im Unternehmen vorhanden ist, ist irgendwie so gespeichert, dass man es auch wiederfindet.
254	120816_Transcript_Expert4	Dimensions\Organizational\Data Ethics and Security	Genau. Dann Data Ethics, Data Ethics habe ich mal so mit responsible Data übersetzt.
255	120816_Transcript_Expert4	Dimensions\Organizational\Data Ethics and Security	Genau, also gerade bei dem Security spielen ja auch noch mehrere Sachen mit rein.
256	120816_Transcript_Expert4	Dimensions\Organizational\Data Ethics and Security	Einmal hast du den Datenschutz, was schon eher so responsible Data ist, aber du hast auch, dass wenn du jetzt in diesem Prozess bist, die Sicherheit, dass du beim cleanen z.B. die Daten nicht änderst.
257	120816_Transcript_Expert4	Dimensions\Organizational\Data Ethics and Security	Genau Level 1 wäre dann: no Ethics strategy at all, also dass es gar kein Bewusstsein dafür gibt, dass es relevant sein könnte.
258	120816_Transcript_Expert4	Dimensions\Organizational\Data Ethics and Security	Dann im ersten Schritt (Level2), dass man sich mit dem Thema bisschen beschäftigt hat, dass es das gibt, dass es auch Stories gibt.
259	120816_Transcript_Expert4	Dimensions\Organizational\Data Ethics and Security	Genau, dass man sich über Datenschutz ein bisschen Gedanken gemacht hat, was man jetzt veröffentlichen kann und was nicht, was vielleicht auch wiederherzustellen ist.
260	120816_Transcript_Expert4	Dimensions\Organizational\Data Ethics and Security	Es gibt bestimmte Datenschutzbestimmungen und was man releasen kann
261	120816_Transcript_Expert4	Dimensions\Organizational\Data Ethics and Security	im letzten Schritt wäre es dann so eine responsible Data Strategy. Oxfam hat das zum Beispiel in Deutschland, ein sehr beliebtes Beispiel. Die haben sich dann wirklich dieses Thema genommen und dann wirklich Guidelines für geschrieben. Das wäre quasi so die höchste Stufe irgendwie.
262	120816_Transcript_Expert4	Dimensions\Organizational\Data Ethics and Security	Das ist irgendwie auch ein Thema…also ich finde das extrem schwer..so responsible Data.
263	120816_Transcript_Expert4	Dimensions\Individual\Ask question / Define	define. Ein erster Schritt wäre erstmal, dass man gar nicht weiß, wie man rangeht und so Fragen formuliert und stellt.
264	120816_Transcript_Expert4	Dimensions\Individual\Ask question / Define	ersten Schritt (Level2) lernt man quasi simple Fragen zu stellen und stellt sich ein Datenset vor und überlegt sich: ok wie könnte ich jetzt Sachen da rausbekommen? Also ganz einzelne Punkte, ganz spezifische Punkte nimmt
265	120816_Transcript_Expert4	Dimensions\Individual\Ask question / Define	im zweiten Schritt (Level3) würde man dann halt wenn man ein Projekt hat, eine komplexe Fragestellung entwickeln, die dann mehrere Aspekte abdeckt und vielleicht auch mehrere Datensätze abdeckt.
266	120816_Transcript_Expert4	Dimensions\Individual\Ask question / Define	Die höchste Stufe wäre für mich dann. Ach ne das hab ich da unten. Genau, dass man wirklich ein datengestriebenes Projekt entwickelt und das irgendwie kommuniziert.
267	120816_Transcript_Expert4	Dimensions\Individual\Ask question / Define	Genau, also im ersten Schritt hast du erstmal die Daten und stellst dann die Fragen und weiter hinten ist es dann so, dass man Fragen stellt und dann schaut, ob es Daten gibt und die dann auch anfragt gegebenenfalls. Genau und das ist dann so die hohe Kunst, dass man wirklich vorher…also die Kunst ist wirklich bei so einem Datenprojekt, dass wenn man einmal angefangen hat das Ding zu bauen, dass man nicht denkt: jetzt kann man noch das machen und dann kann man das noch machen und am Ende kann keiner mehr was damit anfangen.
268	120816_Transcript_Expert4	Dimensions\Individual\Find	finden von Daten. Hatte ich jetzt im ersten Moment, also ich bin da eher technisch rangegangen. Im ersten Schritt, das klassische ist so Nutzung von Suchmaschinen.
269	120816_Transcript_Expert4	Dimensions\Individual\Find	Da bin ich einfach davon ausgegangen, dass das jeder kann.

270	120816_Transcript _Expert4	Dimensions\Individual\Fi nd	Ok ich will irgendwas wissen dann benutze ich Google und versuche mich da durchzuklicken und genau diese erste Enlightenment Stufe wäre quasi, dass man diese advanced Suche von Google benutzt und dann sagt: Ok, ich führe das jetzt in Anführungszeichen, das soll nicht mit drin sein, ich kann nach Dateitypen suchen, das ganze Zeug und genau das zweite wäre dann das IFG also Freedem of Information Act.
271	120816_Transcript _Expert4	Dimensions\Individual\Fi nd	Das ist glaube ich auch relativ niedrig. Also einfach Sachen bei Behörden anfragen und genau.
272	120816_Transcript _Expert4	Dimensions\Individual\Fi nd	Bei dem dritten Schritt benutzt man dann wirklich Datenportale und dann kommt man so bisschen von der Google Suche weg und sagt: ich benutze lieber Portale, ich schaue, was da so alles da ist, welche spezifischen Themen und kann diese Sachen dann auch nutzen.
273	120816_Transcript _Expert4	Dimensions\Individual\Fi nd	Genau und dann hier hinten im höchsten Schritt, wäre das auch so ein bisschen mit bei Data Ethics und Security dabei, dass man eigene Daten verfügbar macht. Eigentlich wäre der höchste Schritt dann, dass man Datensets verknüpfen kann. Man sucht sich also verschiedene Quellen und matcht die dann irgendwie und kann darauf neues Wissen generieren.
274	120816_Transcript _Expert4	Dimensions\Individual\Ge t	HS: Und was sagst du zur Generierung von neuen Daten? Surveys zum Beispiel?
275	120816_Transcript _Expert4	Dimensions\Individual\Ge t	IP4: Eigentlich eher get würde ich sagen.
276	120816_Transcript _Expert4	Dimensions\Individual\Ge t	Dann get. Wäre quasi im ersten Schritt wieder, dass man mit verschiedenen Datenformaten zum Beispiel nicht umgehen kann und dann nur aus Volltexten Sinn rausziehen kann und sich dann eigene Sachen baut.
277	120816_Transcript _Expert4	Dimensions\Individual\Ge t	Und dann erster Schritt (Level 2) wäre dann erstmal: Nutzung von Downloads und csv. Das ist so das einfachste Datenformat, dass man damit umgehen kann und da das in Excel importieren kann.
278	120816_Transcript _Expert4	Dimensions\Individual\Ge t	Dann im zweiten Schritt, dass man komplexere Datenformate wie JSON oder XML nimmt und, dass man auch APIs benutzen kann und diese abfragen kann.
279	120816_Transcript _Expert4	Dimensions\Individual\Ge t	der höchste Schritt dabei wäre für mich, dass man so ?? hat in seinen Anwendungen, dass man so Updates hat wenn neue Daten verfügbar sind. Man bekommt dann sozusagen Bescheid von seinem System: da gibt's jetzt neue Daten und füg die doch mal ein in deine Anwendung, dass man die quasi updatet.
280	120816_Transcript _Expert4	Dimensions\Individual\Ve rify	Verifizieren. Genau, da wäre ein erster Schritt wieder, dass man einfach die Daten nimmt, die man irgendwie findet, was damit macht, aber gar nicht weiß, wo die genau herkommen.
281	120816_Transcript _Expert4	Dimensions\Individual\Ve rify	Im zweiten Schritt wäre dann, dass man die Quellen checkt und guckt, dass man Datensätze aus mehreren Quellen bekommt und guckt was da wirklich dahinter steht. Vielleicht auch von wem die Sachen erstellt wurden, also da hätte man dann dieses biased Data
282	120816_Transcript _Expert4	Dimensions\Individual\Ve rify	Genau, dann im nächsten Schritt, dass man die Sachen, die man wirklich benutzt verfügbar macht in seinem Projekt und sagt: ich hab die und die Quellen benutzt, findet ihr hier und hier könnt ihr selber checken.
283	120816_Transcript _Expert4	Dimensions\Individual\Ve rify	Ich hatte hier noch live Änderungen der Daten. Ich find das sehr schön. Es gibt so eine Visualisierung der Goldmedaille bei olympischen Spielen und da kannst du quasi live in die Daten reingehen, die ändern und dann ändert sich die Visualisierung. Genau, das fand ich ziemlich cool, weil man dann auch selber als Benutzer mit den Zahlen spielen kann. Bin mir aber nicht genau sicher, ob das zu verify gehört.
284	120816_Transcript _Expert4	Dimensions\Individual\Ve rify	Aber genau, eigentlich wäre es so das Verfügbarmachen der Quellen und dass man die selber nachprüfen kann.

285	120816_Transcript _Expert4	Dimensions\Individual\Verify	Also für dich ist verify, die Quellen zu hinterfragen, also kritisch hinterfragen? IP4: Genau, also warum sind die Daten erstellt wurden und dass man sieht wo die Daten herkommen. Das ist mega wichtig und wird auch oft vergessen.
286	120816_Transcript _Expert4	Dimensions\Individual\Clean	clean da wäre der erste Schritt man hat gar kein Verständnis dafür und man benutzt die Sachen einfach, versucht dann irgendetwas daraus zu lesen.
287	120816_Transcript _Expert4	Dimensions\Individual\Clean	Irgendwann kommt man dann darauf: ok, die Sachen sind nicht ordentlich formatiert, das kann ich gar nicht ordentlich benutzen. Das erste was man so als Tool benutzt, ist wohl Excel. Dass man die Sachen dann händisch ändert sozusagen. Man weiß dann was das Problem ist… HS: Also ein paar Qualitätsktierien sind dann schon definiert? IP4: Ja, genau. Also zum Beispiel, dass das Datum gleich geschrieben ist usw. HS: Keine leeren Zellen und so. IP4: Genau. Und dann Duplikate natürlich. Genau und dass man das dann ändert.
288	120816_Transcript _Expert4	Dimensions\Individual\Clean	Der nächste Schritt wäre dann so ein Tool wie OpenRefine, was dann bisschen mächtiger ist und wo man dann wirklich auch Sachen, die ähnlich geschrieben sind, clustern kannst.
289	120816_Transcript _Expert4	Dimensions\Individual\Clean	Genau, der letzte Schritt, also der höchste Schritt wäre es dann, dass man so eine Data Pipeline hat. Man automatisiert dann quasi diesen Schritt des cleanens. Ich hole mir die Daten und das läuft durch ein Script durch und dann sind die Sachen clean und ich kann die einfach weiter benutzen.
290	120816_Transcript _Expert4	Dimensions\Individual\Analyze	Analysieren. Das erste wäre: man hat eine Tabelle, man benutzt ganz ganz einfache Sachen wie ich such mir jetzt einfach mal alle meine Werte, die ich in der Tabelle habe auf, das kann halt irgendwie jeder machen.
291	120816_Transcript _Expert4	Dimensions\Individual\Analyze	Im ersten Schritt (Level 2) würde man dann Pivot Tabellen benutzen wo man Sachen aggregiert und kompakt darstellt, dass man weiß, worum geht's eigentlich in dem Datensatz.
292	120816_Transcript _Expert4	Dimensions\Individual\Analyze	Genau, also einfach unter der Spalte steht quasi die Summe und ja, weiß aber auch nicht genau, was man da genau herauslesen will. Ist aber so das, was man so als erstes macht.
293	120816_Transcript _Expert4	Dimensions\Individual\Analyze	Genau, dann Pivottabellen, dass man so Spreadsheets benutzt und was da so alles geht.
294	120816_Transcript _Expert4	Dimensions\Individual\Analyze	Use of sophisticated Tools, vielleicht kann man da auch Methods sagen. Mein Lieblingsbeispiel ist wirklich sowas wie Standardabweichung. Damit kann der Otto-Normalverbraucher so nichts anfangen, aber das hilft total gut um Outlier zu sehen zum Beispiel und das kann man ja auch mit Excel machen.
295	120816_Transcript _Expert4	Dimensions\Individual\Analyze	der höchste Schritt wäre dann sowas wie Machine Learning Ansätze. Man gibt so ein Vokabular mit rein und anhand der Daten lernt die Machine sozusagen in welche…also welche Sachen man zum Beispiel klassifizieren kann. Also dass man die Sachen nicht mehr selber analysieren muss, sondern dass das alles automatisch passiert.
296	120816_Transcript _Expert4	Dimensions\Individual\Analyze	statistischen Modellen bisschen mit aufgebaut. Weil wir haben zum Beispiel dann auch so Vorhersagen oder sowas gemacht. Weiß nicht, ob das zu detailliert ist. IP4: Genau, aber das wäre dann auch eher in diesem höchsten Level dann.
297	120816_Transcript _Expert4	Dimensions\Individual\Visualize	dann Visualisieren. Das ist dann, dass man am Anfang nicht weiß wofür / welcher Visualisieriungstyp das eigentlich ist und probiert dann so ein bisschen rum und was gut aussieht, das nimmt man dann.
298	120816_Transcript _Expert4	Dimensions\Individual\Visualize	bei dem zweiten Schritt, ist dann, dass man so Excel zum Beispiel benutzt oder Infograms zum Visualisieren, dass man dann auch weiß, wenn ich jetzt hier Zeitreihen Daten habe, dann benutze ich ein Liniendiagramm und wenn ich irgendwelche Kategorien habe, dann benutze ich ein Diagramm. Und dass ich z.B. Kreisdiagramme

			immer keine gute Idee ist. Also wenn man immer nur 2 oder 3 Kategorien hat, dann ist das ok, aber wenn man mehrere hat, dann sieht man die feinen Unterschiede nicht mehr.
299	120816_Transcript_Expert4	Dimensions\Individual\Visualize	Dann der dritte Schritt wäre dann so, dass man seine eigenen Visualisierungen baut und noch mehr macht. Es gibt da so ein schönes Ding, die haben die Bilder von Friedrich Wilhelm IV visualisiert und quasi auf einer Timeline angeordnet auch so mit Balken.
300	120816_Transcript_Expert4	Dimensions\Individual\Visualize	HS: Das ist dann Level4? IP4: Ja, genau das ist dann Level 4, weil die ist technisch extrem anspruchsvoll.
301	120816_Transcript_Expert4	Dimensions\Individual\Visualize	HS: Bei 3. wäre das dann auch so ein bisschen Dashboards, vielleicht? Also ein paar komplexere Visualisierungen. IP4: Ja, Dashboards, Karten sowas. Genau das ist eigentlich ganz schön, so CartoDB.
302	120816_Transcript_Expert4	Dimensions\Individual\Visualize	Ja, genau, das ist dann auch maximal einfach zu benutzen. Und das letzte Level wäre dann wirklich, dass man sich sein eigenes Zeig baut. So Netzwerke bauen zum Beispiel, das ist auch immer schwer.
303	120816_Transcript_Expert4	Dimensions\Individual\Communicate	Erster Schritt wäre dann, dass man gar nicht richtig kommuniziert, vielleicht ein paar mal twittert oder so.
304	120816_Transcript_Expert4	Dimensions\Individual\Communicate	Der erste Schritt (Level 2) wäre dann wenn man das in Blogpost verpackt, also so statische Visualisierungen, dann das anreichert
305	120816_Transcript_Expert4	Dimensions\Individual\Communicate	ein dritter Schritt wäre dann, dass man eigentlich ein eigenes Projekt hat mit interaktiven Visualisierungen mit denen man interagieren kann.
306	120816_Transcript_Expert4	Dimensions\Individual\Communicate	Der letzte Schritt wäre dann, dass man wirklich Campaining macht mit daten-getriebenen Projekten, die man umsetzt
307	120816_Transcript_Expert4	Dimensions\Individual\Communicate	HS: Was sagst du so zum Thema storytelling? IP4: Ja, genau das spielt da so ein bisschen mit rein. An sich ist das hier mit drin.
308	120816_Transcript_Expert4	Maturity Levels	Ich hatte eher so drei Level, also das vierte Level war immer so man kann nichts. Ne, aber das passt aufjedenfall. Das hier hat super viel geholfen (die Beschreibungen von Argyris und Schön) das war cool, daran hab ich mich ja auch so dran lang gehangelt.
309	120816_Transcript_Expert4	Drivers and Barriers of Data Literacy	Also was so ein bisschen Segen und Fluch gleichzeitig ist, sind tolle Beispiele. Das ist cool um Leute anzufüttern. Also das ist möglich, das kann man machen, das haben NGOs gemacht.
310	120816_Transcript_Expert4	Drivers and Barriers of Data Literacy	Gleichzeitig ist es aber auch, dass wenn sie dann mal anfangen und sehen, das ist alles gar nicht so einfach, dass sie dann auch wieder ziemlich abgeschreckt sind wenn es nicht so cool aussieht.
311	120816_Transcript_Expert4	Drivers and Barriers of Data Literacy	Was wirklich wichtig ist, ist dass man ganz basic anfängt.
312	120816_Transcript_Expert4	Drivers and Barriers of Data Literacy	Ja, genau das ist wichtig. Wo wir bei der Datenschule drauf achten wenn wir diese Workshops machen, dass immer irgendwie einer vom Management drin sitzt. Es muss schon unterstützt seind und man braucht auch Freiraum um sich mit den Sachen beschäftigen zu können. Vor allem auch zeitlich. Ja, das wäre das so.
313	122316_MaturityGrid_Expert5	Dimensions\Organizational\Data Culture	No concern for and no employees with significant data science background.
314	122316_MaturityGrid_Expert5	Dimensions\Organizational\Data Culture	employees with data analytics background work in prominent positions.
315	122316_MaturityGrid_Expert5	Dimensions\Organizational\Data Culture	Project managers are fully aware of the need to collect, store and act on data. Data analysis is on the roadmap.

316	122316_MaturityGrid_Expert5	Dimensions\Organizational\Data Culture	Data analysis/insights are implemented and optimized at all levels of the workflow. Data literacy is an important skill for most employees even at a senior level.
317	122316_MaturityGrid_Expert5	Dimensions\Organizational\Data Ethics and Security	data are handled (if at all) in an ad hoc manner with no concern for security or ethical dilemma
318	122316_MaturityGrid_Expert5	Dimensions\Organizational\Data Ethics and Security	some employees and projcts are aware of security issues; uncoordinated attempts to store data securely.
319	122316_MaturityGrid_Expert5	Dimensions\Organizational\Data Ethics and Security	internal consistent, company wide policies for secure and ethically sound data handling.
320	122316_MaturityGrid_Expert5	Dimensions\Organizational\Data Ethics and Security	consistent, company wide policies for secure and ethically sound data handling are constantly redefined and updated. Senior employees are responsible for the enforcement of these rules.
321	122316_MaturityGrid_Expert5	Dimensions\Individual\Ask question / Define	Ability to ask simple questions
322	122316_MaturityGrid_Expert5	Dimensions\Individual\Ask question / Define	Ability to ask complex questions that could be answered by simple data queries.
323	122316_MaturityGrid_Expert5	Dimensions\Individual\Ask question / Define	Entire projects are based on multidimensional questions that need complex data queries and multiple iterations to resolve.
324	122316_MaturityGrid_Expert5	Dimensions\Individual\Ask question / Define	Entire projects are based on multidimensional questions that need complex data queries and multiple iterations to resolve. No single person can hanlde these inquries.
325	122316_MaturityGrid_Expert5	Dimensions\Individual\Find	Google
326	122316_MaturityGrid_Expert5	Dimensions\Individual\Find	Knowledge of more specialized search engines.
327	122316_MaturityGrid_Expert5	Dimensions\Individual\Find	Knowledge of public databases and more sophisticated search engines.
328	122316_MaturityGrid_Expert5	Dimensions\Individual\Find	Complex database queries
329	122316_MaturityGrid_Expert5	Dimensions\Individual\Get	no awareness of different data formats, information is derived from full text and used as based for further processing
330	122316_MaturityGrid_Expert5	Dimensions\Individual\Get	use of downloads and data formats such as .csv
331	122316_MaturityGrid_Expert5	Dimensions\Individual\Get	use of more complex data formats such as JSON and XML, use of APIs
332	122316_MaturityGrid_Expert5	Dimensions\Individual\Get	automatic system updates that new data exists and can be integrated
333	122316_MaturityGrid_Expert5	Dimensions\Individual\Verify	data are taken at face value
334	122316_MaturityGrid_Expert5	Dimensions\Individual\Verify	Critical check of simple data quality measures
335	122316_MaturityGrid_Expert5	Dimensions\Individual\Verify	Multiple layers of data checking are implemented in standard procedures across the company.
336	122316_MaturityGrid_Expert5	Dimensions\Individual\Verify	Best practices exist, employees are trained in data verification procedures and priorities.
337	122316_MaturityGrid_Expert5	Dimensions\Individual\Clean	data are taken as is
338	122316_MaturityGrid_Expert5	Dimensions\Individual\Clean	Simple dealing with duplicates, missing values and extreme outliers
339	122316_MaturityGrid_Expert5	Dimensions\Individual\Clean	Sophisticated duplicte detection, string manipulation tools, contingency tables, conditional plots,

340	122316_MaturityGrid_Expert5	Dimensions\Individual\Clean	Data cleaning is regarded as a challengning modeling tasks. Outliers are defined as residuals, the full power of data visualization is put to task.
341	122316_MaturityGrid_Expert5	Dimensions\Individual\Analyze	bar and pie charts, simple scatter plots, basic summaries of data
342	122316_MaturityGrid_Expert5	Dimensions\Individual\Analyze	pivot tables, multidimensional tables and summaries, histograms, boxplots, still mainly descriptive.
343	122316_MaturityGrid_Expert5	Dimensions\Individual\Analyze	Inferential view of the data, full account of uncertainties, linear regression, decision trees,
344	122316_MaturityGrid_Expert5	Dimensions\Individual\Analyze	full suite of machine learning tools, clustering, forecasting, boosting, ensemble learning
345	122316_MaturityGrid_Expert5	Dimensions\Individual\Visualize	bar and pie charts, simple scatter plots, basic summaries of data
346	122316_MaturityGrid_Expert5	Dimensions\Individual\Visualize	pivot tables, multidimensional tables and summaries, histograms, boxplots, still mainly descriptive.
347	122316_MaturityGrid_Expert5	Dimensions\Individual\Visualize	interactive charts, uncertainties are always visualized along with the data
348	122316_MaturityGrid_Expert5	Dimensions\Individual\Visualize	Linked dynamic dashboards, anticipating user requests
349	122316_MaturityGrid_Expert5	Dimensions\Individual\Communicate	no need recognized to communicate beyond the employee
350	122316_MaturityGrid_Expert5	Dimensions\Individual\Communicate	Simple narratives, no attempt at serious legacy code
351	122316_MaturityGrid_Expert5	Dimensions\Individual\Communicate	conferences, talks, monthly updates, blogs
352	122316_MaturityGrid_Expert5	Dimensions\Individual\Communicate	workshops, conferences, talks, monthly updates, blogs, reproducible research is taken seriously !!
353	010217_Evaluation_Expert6_7	Evaluation	nach vision der organisation fragen
354	010217_Evaluation_Expert6_7	Evaluation	Hier könnte man auch darüber nachdenken, eine Auswahl vorzugeben. Z.b. Entwicklungshilfe, Jungendarbeit etc.. Wäre aber wohl eher für eine aggregierte Auswertung im Anschluss interessant.
355	010217_Evaluation_Expert6_7	Evaluation	Idee: Ggf. könnte man auch noch fragen, an wie vielen Standorten die Organisation tätig ist. (oder ob sie international arbeitet). Ggf. wichtig um etwas über Data Culture aussagen zu können / Oder einen Indikator dafür zuhaben ob sich Organisation auf Grund mehrerer Standorte in ihren Data Literacy Skills stark unterscheiden? () Hast du bestimmt schon gemacht, ansonsten könnte man sich auch bestimmt Fragebögen aus der Organisationsanalyse als "Best Practices" anschauen
356	010217_Evaluation_Expert6_7	Dimensions\Individual\Find	Da Englisch würde ich hier die Abkürzung FOI benutzen
357	010217_Evaluation_Expert6_7	Dimensions\Individual\Get	vorallem pdf genutzt und selbst erstellt?
358	010217_Evaluation_Expert6_7	Dimensions\Individual\Get	Vielleicht könnte das Wort 'downloads' hier für Verwirrung sorgen?. Also besteht die Abgrenzung zum vorherigen darin, dass keine digitalen Quellen verwendet werden? Wenn ja, vielleicht vorher noch: Information is rarely processed digitally?
359	010217_Evaluation_Expert6_7	Dimensions\Individual\Analyze	Hier ggf. zu beginn noch einfügen: einfügen. "Exploration of relationship between variables in a dataset." Vielleicht etwas einfacher zu verstehen als Inferential view of the data....
360	010217_Evaluation_Expert6_7	Dimensions\Individual\Visualize	evtl überlegen, wo man excel/libre office als analysetool einordnet, weil damit am häufigsten gearbeitet wird
361	010217_Evaluation_Expert6_7	Dimensions\Individual\Communicate	evtl noch: bauen narrative online auf die website ein

362	010217_Evaluation _Expert6_7	Evaluation\Correctness	finde ich soweit alles schlüssig und das relevanteste ist drin :)
363	010217_Evaluation _Expert6_7	Evaluation\Correctness	Ich denke auch, die wichtigsten Aspekte sind enthalten. Die Abstufungen zwischen den einzelnen Levelstufen finde ich sehr gelungen
364	010217_Evaluation _Expert6_7	Evaluation\Flexibility	sehr gut - klar verständlich v.a. durch einzelne toolbeschreibungen
365	010217_Evaluation _Expert6_7	Evaluation\Usability	auch gut, evtl noch darauf achten, dass die formulierungen nicht zu negativ klingen - gerade, wenn man die orgs bittet, sich dazu zu positionieren (bei den niedrigen wissensstufen in den einzelnen kategorien (find, get ...)
366	010217_Evaluation _Expert6_7	Evaluation\Usability	Hier hatte ich noch die Idee, dass man die Aussagen vielleicht nicht hierarchisch aufbaut sondern durchmischt. Dadurch kann derjenige der den Fragebogen ausfüllt nicht gleich die "Wertigkeit" der Antwort durchschauen.
367	010217_Evaluation _Expert6_7	Evaluation\Implementability	können wir gut nutzen :) evtl. für die orgs einfacher, wenn alles auf deutsch ist
368	010217_Evaluation _Expert6_7	Evaluation\Implementability	Ja als Einstufung zu Beginn einer Partnerschaft wäre das für uns großartig!!
369	010217_Evaluation _Expert6_7	Evaluation\Economic Efficiency	gut machbar, interessant wäre zu wissen, wie lange das ausfüllen durch orgs dauert. wäre sicher gut das dok in einen google fragebogen einzbauen, dann könnten wir den einfach digital versenden
370	010517_Evaluation _IP8	Dimensions\Individual\Verify	Also hier wird gefragt: welche Methoden nutzt du für die kritische Überprüfung von Daten und hier in der Antwort ist in der Organisation implementiert….ich finde das wiederspricht sich ein bisschen. Also entweder bin ich das, der das macht, oder eben auf Organisationslevel.
371	010517_Evaluation _IP8	Evaluation	Ja, genau. Je nachdem mit was für Organisationen ihr da macht, würde ich glaube die Antwortmöglichkeiten nochmal etwas vereinfachen. Bzw. noch etwas anpassen.
372	010517_Evaluation _IP8	Evaluation	Ja, es kommt tatsächlich glaube ich auch total drauf an wie man arbeitet. Was braucht man. Weil das natürlich irgendwie gerade wenn es dann ums visualisieren oder analysieren geht, dann brauchst du ja tatsächlich auch Skills. Also das ist dann ja nicht nur ein Verständnis für Daten und wo finde ich die, was kann ich mit denen machen. Da geht's dann ja wirklich ums richtige arbeiten, programmieren SPSS und so. Ja, das braucht man dann ja nicht immer und überall. Da braucht man aber eben andere Sachen, dass man es gut darstellen kann, dass man es gut interpretieren kann und so was.
373	010517_Evaluation _IP8	Evaluation	Also ich glaube, dass ich mein Verständnis und die Interpretation von Daten sehr gut ist, aber das direkte arbeiten damit, eher schwierig ist. Bzw. das kritische hinterfragen und zu überlegen: wie sind die jetzt auf den Weg gekommen und was bedeutet das?
374	010517_Evaluation _IP8	Evaluation	: Also für dich ist das auch noch so ne Kompetenz dieses Verstehen und Interpretieren? IP8: Ja, total. Das finde ich jetzt vor allem für den Laien sozusagen.
375	010517_Evaluation _IP8	Evaluation	Gerade weil ja Daten und Ergebnisse von Studien, wird ja immer mehr benutzt um irgendwelche Forderungen zu untermauern und dann finde ich es super wichtig um zu fragen: Ok, aber ist es überhaupt richtig, das Ergebnis und kann man das überhaupt in Beziehung setzen, oder nicht?
376	010517_Evaluation _IP8	Evaluation\Flexibility	Ich glaube da (bei Flexibilität) geht noch bisschen was. Aber es benutzt halt viel Fachsprache und ich finde die Ansprache in den Fragen und die Antwortmöglichkeiten ist etwas unterschiedlich.
377	010517_Evaluation _IP8	Evaluation\Usability	Nutzerfreundlich ist es schon. War ja jetzt wirklich einfach.
378	010517_Evaluation _IP8	Evaluation\Economic Efficiency	Ja, also das hat ja jetzt irgendwie zehn Minuten gedauert.

379	010517_Evaluation _IP8	Evaluation\Correctness	Puhhh…das ist echt schwierig. Weil vollständig und korrekt kann ich eigentlich nicht einschätzen. Vollständig kannst du glaube ich nie sein. Und korrekt finde ich jetzt für mich sehr schwer zu bewerten. Würde ich dann mit Neutral bewerten eher.
380	010517_Evaluation _IP8	Evaluation\Flexibility	Ich denke so technologische Entwicklung schon, aber bei verschiedenen Organisationen, ich glaube dazu muss man die Fragen nochmal ein bisschen anpassen. Je nachdem wo man hingeht.
381	010517_Evaluation _IP8	Evaluation	Es ist halt ein großer Unterschied zwischen Gewinnung und Nutzung der Daten. Und bei uns ist eher der Fokus auf der Nutzung. Und die Fragen fand ich jetzt eher auf die Gewinnung der Daten ausgerichtet. Und so hatte ich schnell das Gefühl, das passt jetzt nicht so richtig und ich musste was beantworten, was ich gar nicht so richtig weiß.
382	010517_Evaluation _IP8	Evaluation	

	Sehr hoch	Hoch	Neutral	Gering	Sehr gering	Kommentar
Richtigkeit			x			
Flexibilität			x			
Nutzerfreundlichkeit		x				
Durchführbarkeit	x					
Wirtschaftlichkeit	x					

383	010517_Evaluation _IP9	Evaluation	Das ist voll schwierig zu sagen (Evaluierung der eigenen Daten Fähigkeiten), woran man das festmachst. Das möchtest du ja gerade herausfinden, aber ich kann dir gar nicht sagen wie die Gesellschaft im Allgemeinen vielleicht bin ich schon fortgeschritten, aber einfach, weil die Allgemeinheit wenig weiß, aber das willst du ja quasi rausfinden.
384	010517_Evaluation _IP9	Dimensions\Organizational\Data Culture	Hier ist es irgendwie etwas schwierig (Datenkultur), eher das oder das. Weil man keine Zwischenschritte gehen kann. Also hier bis zu Hälfte, da etwas weniger, also wo man sagen kann: trifft eher zu, trifft weniger zu. Die einzelnen Sachen aufschlüsseln, weil es ja so ein großer Text ist, und ich muss jetzt dem großen Text zustimmen, was etwas schwierig ist. Und muss jetzt dem ganzen Text zustimmen und stufe uns jetzt etwas besser ein als wir wahrscheinlich eigentlich sind.
385	010517_Evaluation _IP9	Dimensions\Organizational\Data Ethics and Security	Also hier ist es jetzt einfacher (Data Ethics and Security), das geht alles in eine Richtung und ich weiß was damit gemeint ist.
386	010517_Evaluation _IP9	Dimensions\Individual\Find	Bei uns ist es jetzt ja auch so, dass wir viel eigene Daten generieren (Daten finden). Also welche Schulen mit uns zusammenarbeiten, das wird ja schon dargestellt und veröffentlicht. Dafür bekommen wir dann ja auch die Fördergelder. Das wäre dann sowas wie Punkt 4? Obwohl in Verbindung mit verschiedenen Datenquellen und komplexen Abfragen, das ist es dann halt auch wieder doch nicht. Das weiß ich nicht genau.
387	010517_Evaluation _IP9	Dimensions\Individual\Find	Ja, genau. Also wir nutzen eigene Datenquellen, aber verbinden die nicht mit anderen Quellen.
388	010517_Evaluation _IP9	Evaluation	Ja, also dann könnte ich mir auch vorstellen, dass wenn man Anfänger ist, dann ist das schon sehr komplex.
389	010517_Evaluation _IP9	Dimensions\Individual\Get	Also wir haben so ein CRM System wo ganz viele Daten hinterlegt sind und wenn ich mir da Daten herausziehen möchte, dann ist es ja schon so, also es geht so um ich möchte alle Schulen aus Brandenburg, die schon an diesem und jenem Programm teilgenommen haben und kann daraus dann eine Excel-Tabelle ziehen. Das ist jetzt kein komplexes Datenformat, oder? Das ist ja einfach eine Excel-Tabelle, die ich daraus ziehen kann.
390	010517_Evaluation _IP9	Dimensions\Individual\Communicate	Also wir kommunizieren oft an den Fördergeldgeber häufig kommunizieren. Das geht dann so Teilnehmerzahlen, oder weiß ich auch nicht.

391	010517_Evaluation_IP9	Dimensions\Individual\Communicate	Ja, genau. Es ist eher Reporting. Also manchmal vielleicht auch noch in einem Newsletter damit es nach außen getragen wird, aber mehr machen wir dann da auch nicht.
392	010517_Evaluation_IP9	Evaluation	Ja, aber wenn du diese Frage jetzt hier stellst, dann muss ich ja tatsächlich wissen, ob ich zwischen 2 oder 3 bin oder 1 und 2.
393	010517_Evaluation_IP9	Evaluation	HS: Und gab es im Modell Überraschungen, die an Kompetenzen gefragt wurden? IP9: Ne, eigentlich nicht, für mich nicht. Wo ich jetzt schon so 1,2 Mal drüber nachgedacht habe. Aber für jemanden, der noch nie damit Kontakt hatte, der hat auch gar keine Vorstellung darüber was man mit Daten alles machen kann.
394	010517_Evaluation_IP9	Evaluation\Correctness	Richtigkeit…weil ich ja auch kein Experte bin, kann ich irgendwie schlecht einschätzen.
395	010517_Evaluation_IP9	Evaluation\Correctness	Ja, genau ok. Ich würde das schon hoch sehen, aber so Kleinigkeiten, wo man das noch etwas besser bzw. gibt es ja auch so viele verschiedene Menschen, die Daten nutzen unter verschiedenen Gesichtspunkten, die sehr schwierig ist das abzubilden und dann wird es auch zu komplex.
396	010517_Evaluation_IP9	Evaluation	Du meinst das Modell im Allgemeinen? Um herauszufinden wie Menschen mit Daten arbeiten? Ich glaube das ist definitiv wichtig, weil da kann man ja genau nochmal gucken: wo fehlen Informationen, solche Studien haben ja auch immer Auswirkungen auf Bildung oder was man noch intensiver betreiben könnte.
397	010517_Evaluation_IP9	Evaluation	Aber solche Sachen helfen so allgemein eine Offenheit zu fördern und dass man darüber spricht und schaut wo man Verbesserungen ansetzen kann und deswegen ist es aufjedenfall eine gute Idee.
398	010517_Evaluation_IP9	Evaluation\Flexibility	Das kann ich jetzt ehrlich gesagt gar nicht so sagen (Flexibilität), auf technologischer Ebene bin ich eben gar nicht so drin im Datenthema, also ob sich Data Literacy in die und die Richtung entwickelt, ob das Modell dann noch ausreicht. Da kenn ich mich einfach gar nicht aus und kann da dann auch keine Aussage zu treffen.
399	010517_Evaluation_IP9	Evaluation\Usability	Nutzerfreundlichkeit finde ich tatsächlich etwas schwierig, weil es diese ganzen Textblöcke sind und das kann ich irgendwie nicht.
400	010517_Evaluation_IP9	Evaluation\Implementability	Also Durchführbarkeit finde ich dann schon auch…eher gut.
401	010517_Evaluation_IP9	Evaluation\Implementability	Ja, das kommt auch so drauf an wo die Organisation schon steht. Wenn sie noch ganz am Anfang steht, dann ist sie völlig überfordert, wenn man nicht weiß wo man anfangen kann. Wenn man sich aber schon positionieren kann und wir achten darauf, aber wir könnten das noch verbessern, dann wäre das schon sehr gut. Das zeigt ja auch Schwächen und Stärken ganz gut. Das fand ich bei den einzelnen Punkten auch sehr gut, da bin ich nicht so gut, aber da habe ich irgendwie schon was erreicht.
402	010517_Evaluation_IP9	Evaluation\Economic Efficiency	Genau, das ist ja schon was offen Zugängliches. Also dann eher Wirtschaftlichkeit hoch (kostenlos).
403	010517_Evaluation_IP9	Evaluation	Was ich mir noch vorstellen könnte, wenn ich das auswerte. Bei diesen Sachen so allgemeingültige Sachen zu sagen. Dazu müsste es irgendwie kleinschrittiger sein. Also dazu müsstest du jeden einzelnen Satz zur Abstimmung machen.
404	010517_Evaluation_IP9	Evaluation	

	Sehr hoch	Hoch	Neutral	Gering	Sehr gering	Kommentar
Richtigkeit		x				
Flexibilität			x			
Nutzerfreundlichkeit				x		
Durchführbarkeit	x					
Wirtschaftlichkeit	x					

Appendix F: Maturity Grid - All Experts

		Read			Write
Generic Level Descriptors	**EXP1_2**	**Uncertainty**	**Enlightenment**	**Certainty**	**Data Fluency**
Underlying Rationale	**Argyris and Schön (1978)**	Describes situations, where the problem is not yet addressed and there is no engagement with the related activity. The connection between the factor and the overall topic of the MM is not reflected on	Corresponds with the first type of learning suggested by Argyris and Schön, and points out that some action is changed to correct a certain behavior. Other than that, tasks are carried out as usual.	Is based on Argyris' and Schön's second type of learning: people at this stage, modify their actions, as well as thinking critically about existing norms, procedures, policies, and objectives that govern their actions. This means that the general situation is taken into consideration.	It corresponds to the third type of learning and signifies a stage, where an organization is aware of the influence of a given factor and continuously checks whether the way things are handled or setup is still appropriate for the given situation. At this stage, learning does not only happen when a mistake needs to be corrected: participants have the general mindset of continuously adjusting and improving the situation.
General Level Description	**Hillson (1997)**	Organizations on Level 1 are unaware of the need for data literacy skills and have no or very vague understanding on what is required. There is no structured approach to dealing with data-related topics. Processes are reactive with no or little attempt to learn from the past or prepare for future data-driven possibilities.	Organizations are experimenting with the application of data-related topics, usually through a small number of nominated individuals, but has no structured generic processes in place. Although, there are aware of the potential benefits of managing and using data, organizations have not yet effectively implemented data-related processes and are not gaining the benefits.	Organizations have built data-driven activities into their routine business processes and implement data wherever it makes sense. Generic procedures and standards on how to handle data are formalized and widespread, benefits are understood at all levels of the organization. Still benefits may not be consistently achieved in all cases.	Organizations have established a data-driven culture throughout all levels. Data is actively used to improve business processes and gain competitive advantage. Data is used to manage opportunities as well as potential negative impacts.
General Level Descriptions	**EXP1_2**	Describes a rather "analogue" phase where people have to be pushed to understand that data exists and is relevant for the future (n98). Individuals might have a certain interest in data, work digitally, but are unsure about what to do and expect from data (n101). They are unsure about the different steps that exist when working with data or do not know that they exist (n108).	Describes a state where a lot about data has already been understood theoretically, but cannot be applied in many cases and has to be further trained (n112, n113).	Describes a phase where most of the data handling steps are handled with confidence and can evaluate their competences correctly (n116).	Describes a phase where all steps are internalized and can be executed on a sophisticated level within workflows (n138).

Organizational	**Data Culture**	**Prelimi-nary**	Data is perceived as an ambiguous term which causes insecurities.	Data is perceived as an interesting concept and benefits are appreciated. Insecurities exist regarding use cases and what exactly to expect.	Data is not perceived as a source of insecurities, but rather understood as an enabler for progress and support for existing and planned activities. Higher management and leaders support data initiatives.	Psychological barriers of data have been brought down (insecurities, fear, resignation, etc.) and comfort around data is promoted. Higher management and project managers understand and support importance of dedicated resources (time, budget, human resources) for data handling and conversion.
		EXP1_2		Awareness that working with data does not require academic training in computer science (n69), Working with data does not cause fear (n71)		
		EXP3	can be completely comfortable around working with data, but if you don't have buy-in from senior management, if you don't have time and budgets and project managers who understand that you need this time ….and if not that you don't really have a data culture (n204)			
		EXP4	no employees with data experience (n248)	experts who have data knowledge join the team (n249)	embedding experts in the workflow (n250, n251)	central repository with shareable information (n252, n253)
		EXP5	No concern for and no employees with significant data science background (n313)	employees with data analytics background work in prominent positions (n314)	Project managers are fully aware of the need to collect, store and act on data. Data analysis is on the roadmap (n315)	Data analysis/insights are implemented and optimized at all levels of the workflow. Data literacy is an important skill for most employees even at a senior level (n316)
	Data Ethics & Security	**Prelimi-nary**	No guidelines are defined that ensure confidentiality, integrity and availability of data.	Rising awareness of the importance of the responsible use of data. No defined guidelines.	Awareness of the impacts of data use. Guidelines for responsible data use are defined and incorporated to activities.	Processes are in place to ensure confidentiality, integrity, and availability of data is adequately protected are clearly defined and responsibly used. Only data that is necessary is collected. Ability to guarantee that no unauthorized parties access data.
		EXP1_2	Awareness of when collecting / asking for data makes sense and when it is unnecessary (n90)	Ethical use of data, data is shared, but not all data (n95)		Written responsible data policy (e.g. Oxfam) (n94)
		EXP4	no ethics strategy at all, no awareness that it can be relevant (n257)	interest in responsible data use and data security, not defined in a clear way (n258)	data protection guidelines are defined and responsible data releases described (n259, n260)	Responsible Data Strategy for entire organization (e.g. Oxfam) (n261)

		EXP5	data are handled (if at all) in an ad hoc manner with no concern for security or ethical dilemma (n317)	some employees and projects are aware of security issues; uncoordinated attempts to store data securely (n318)	internal consistent, companywide policies for secure and ethically sound data handling (n319)	consistent, companywide policies for secure and ethically sound data handling are constantly redefined and updated. Senior employees are responsible for the enforcement of these rules (n320)
Individual	**Ask question & Define**	**Prelimi-nary**	Lacking ability to formulate questions to find meaningful answers in data. Cannot see potential of data.	Questions can be asked to data in limited number of situations and answers questions partially. See potential in data for organization / own work. Unsure what to ask for.	Questions to data are formulated precisely and target-oriented in order to find meaningful answers in most of the cases.	Questions are formulated in accordance of multidimensionality of existing data. Answers to informational needs can be consistently found in data.
		EXP1_2	No feeling for what questions can be answered by data. Overinterpretation / faulty derivation of consequences from data (n49). No knowledge on potential or further application opportunities of data (n131)	Ability to ask questions, find answers and identify patterns or tendencies through my questions/answers (n132)		Understand multidimensionality of data and ask subsequent questions. Strong feeling for what data can answer and what cannot be answered with given data (n49, n50)
		EXP4	no ability to formulate relevant questions (n263)	ability to formulate simple questions, dataset is considered to formulate questions, just for single points (n264)	ability to formulate complex question and then look for dataset that can answer the question, question reflects multidimensionality of given dataset (n265)	development of data-driven projects based on experiences, big questions as well as subquestions can be answered with given data (n266, n267)
		EXP5	Ability to ask simple questions (n321)	Ability to ask complex questions that could be answered by simple data queries (n322)	Entire projects are based on multidimensional questions that need complex data queries and multiple iterations to resolve (n323)	Entire projects are based on multidimensional questions that need complex data queries and multiple iterations to resolve. No single person can handle these inquiries (n324)
	Find	**Prelimi-nary**	Limited understanding of possible data sources. No experience and resources for identifying and selecting most relevant data sources.	Knowledge only limited to a few data sources. Growing interest in identifying and selecting most relevant data sources. Teams or individuals cannot detect when data to a given problem is not available and how to obtain new data.	Broad understanding of different data sources, most relevant ones can be chosen from a certain selection of data sources. Teams can detect when data to a given problem is not available, but don't know how to obtain necessary data from other sources.	Profound understanding of what is meant by data and awareness of the various possible types of data sources. Assessment criteria for selecting the ones most relevant to an informational need are formulated. Ability to detect when a given problem or need cannot be solved with the existing data, and are familiar with research techniques to obtain new data.
		EXP1_2	Does not collect data (n130).	Awareness of who collects data, recognize data in texts, websites and other sources (n137), Use of data portals (n81, n151)		
		EXP4	basic use of search engines (n268, n269)	advanced use of search engines complemented by asking for data sources at public authorities (IFG - freedom of information act) (n270, n271)	awareness and use of data portals for specific topics (n272)	publishing own data and matching different sources (n273)

	EXP5	Google (n325)	Knowledge of more specialized search engines (n326)	Knowledge of public databases and more sophisticated search engines (n327)	Complex database queries (n328)
Get	**Prelimi-nary**	Data cannot be acquired without help. Does not collect data.	Data can be generated through surveys or observations. Using external data sources is challenging.	Data can be accessed using basic methods or generating new data through survey / observations. Converting input format is sometimes problematic.	Access to data through sophisticated methods or generating fresh data through surveys or observations. Ability to convert input format into a form that can be used for further processing and analysis. Loading data into a database system.
	EXP1_2	Usually get data using copy/paste or do screenshots and manually create data source (n145)	Downloads and IFG requests at public administrations, use of data portals (n146)		Ability to write automated data scrapers / scripts (n150).
	EXP4	no awareness of different data formats, information is derived from full text and used as based for further processing (n276)	use of downloads and data formats such as .csv (n277)	use of more complex data formats such as JSON and XML, use of APIs (n278)	automatic system updates that new data exists and can be integrated (n279)
	EXP5	no awareness of different data formats, information is derived from full text and used as base for further processing (n329)	use of downloads and data formats such as .csv (n330)	use of more complex data formats such as JSON and XML, use of APIs (n331)	automatic system updates that new data exists and can be integrated (n332)
Verify	**Prelimi-nary**	Critical evaluation of data does not exist. Data quality is not assessed and data evaluation criteria cannot be described.	Critical evaluation of data is vaguely defined. Data quality is not assessed consistently.	Critical evaluation of data is defined regarding certain aspects, data quality is assessed consistently.	Ability to do data quality assessment independently. Data evaluation criteria regarding authorship, method of obtaining and analyzing data, comparability and quality are precisely defined. The impact of data on science and society is internalized, so are copyright and licenses influencing data reuse. Ability of contextualizing specific information to other aspects exists throughout the organization.
	EXP1_2				Self-control, knows what to do with given data (n84), critical thinking / assessment (n85), being able to argue through and about data and engage in debates, use of statistics / values, evidence / facts to argue (n45)
	EXP3	getting a data set and being able to, like knowing what to check (n200)			
	EXP4	no awareness of how and why to critically assess data and where data comes from (n280)	data is used but there is no awareness where data comes from, different sources are not compared (n281)	data sources are checked and different sources are considered and critically assessed (n282)	Published data sources that are used transparently and provide access to the original sources so that audience can check for themselves (n283)

	EXP5	data are taken at face value (n333)	Critical check of simple data quality measures (n334)	Multiple layers of data checking are implemented in standard procedures across the company (n335)	Best practices exist, employees are trained in data verification procedures and priorities (n336)
Clean	**Prelimi-nary**	Invalid records cannot be detected and normalizations strategies are not applied.	Invalid entries can be identified. Basic methods to remove invalid records are missing.	Invalid records can be detected and are removed using cleaning methods.	Independent ability to remove invalid records and translating all the columns to use a sane set of values. Ability to combine different datasets into a single table, remove duplicate entries or apply any number of other normalizations.
	EXP1_2	No awareness that given data might have to be checked, cleaned or normalized.	Awareness of some data quality criteria and specifics of data cleaning (n161), awareness that given data most often is not perfect and need normalization strategy (e.g. machine processability, standardization, completeness (n157)	Use of programs to support data cleaning (e.g. Open Refine) (n143) awareness for normalization to get comparability between different data sources, knowledge about different data formats and how to clean them, high awareness for data quality criteria (machine processable, empty fields, uniform and complete columns and rows (n159, n160)	
	EXP4	no understanding of cleaning data, data is used and information are tried to be derived from uncleaned data (n286)	growing awareness that data is not properly formatted/collected, manual fixing of errors in excel, basic data quality indices are defined (empty fields, duplicates, etc.) (n287)	data cleaning with open refine (n288)	automatic cleaning of data using a script (e.g. in R) (n289)
	EXP5	data are taken as is (n337)	Simple dealing with duplicates, missing values and extreme outliers (n338)	Sophisticated duplicate detection, string manipulation tools, contingency tables, conditional plots (n339)	Data cleaning is regarded as a challenging modeling tasks. Outliers are defined as residuals; the full power of data visualization is put to task (n340)
Analyze	**Prelimi-nary**	No knowledge about how to analyze data.	Ability to work with basic descriptive and inferential statistics. Only a few resources are allocated to data handling.	Ability to work with advanced statistics (e.g. time series, factor analysis, tree-based models, clustering, etc.). Data handling and conversion skills are promoted throughout the organization.	Ability to apply descriptive, predictive and prescriptive statistics when necessary. The organization is independently able to prepare data for analysis, analyze them in keeping with the results sought and knows how to use the necessary tools (data analysis tools both locally: Excel, R, SPSS, Stata, etc. and online). The organization makes ethical use of data, making sure that used methods are deployed and results interpreted transparently and honestly.
	EXP1_2				
	EXP4	simple use of data table to read information (e.g. looking for all values of specific group) (n290)	use of pivot tables for aggregating information (n, 291, n292, n293)	use of more sophisticated statistical methods / tools (n294)	use of machine learning concepts (n295)

	EXP5	bar and pie charts, simple scatter plots, basic summaries of data (n341)	pivot tables, multidimensional tables and summaries, histograms, boxplots, still mainly descriptive (n342)	Inferential view of the data, full account of uncertainties, linear regression, decision trees (n343)	full suite of machine learning tools, clustering, forecasting, boosting, ensemble learning (n344)
Visua-lize	**Prelimi-nary**	No awareness of the multiplicity of how data can be presented. No visualizations can be drawn from data.	Limited ability to find specific outputs especially complex visualizations. Focus on representing tables and a selection of index figures.	Static visualizations can be created from data.	High awareness of the various forms in which data can be presented (written, numerical or graphic). Static and interactive visualizations are commonly used.
	EXP1_2	describes practical visualization skills (n57, n61), diagrams in Excel (n82)			HTML embedded visualizations (n82)
	EXP4	no understanding of when standard visualizations are chosen, decision based on what looks best (trial and error) (n297)	different visualizations are created in Excel in accordance with information that want to be represented (e.g. time related data requests line diagram) (n298, n300)	Dashboards are created by programs such as Tableau or CartoDB (n301)	own, sophisticated visualizations are programmed / created that exceed standard diagram visualizations (e.g. timeline with pictures of Friedrich Wilhelm IV) (n299, n300, n302)
	EXP5	see above (n345)	see above (n346)	interactive charts, uncertainties are always visualized along with the data (n347)	Linked dynamic dashboards, anticipating user requests (n348)
Commu-nicate	**Prelimi-nary**	No stories can be drawn from data.	Limited ability to find specific outputs especially complex stories. Audience is not clearly defined.	Basic stories can be created from data and communicated to targeted audience.	Ability to synthesize, represent and communicate the results of data analysis in ways suited to the nature of the data, their purpose and the audience. Data storytelling is commonly used.
	EXP1_2			Storytelling needs more sophisticated skills, data has to be visualized context-based, connect them through narrative (n47, n60)	
	EXP4	insights from data are not communicated (n303)	static visualizations support narrative (e.g. in blogposts) (n304)	own projects are supported by interactive visualizations (n305)	data-driven campaigning (n306)
	EXP5	no need recognized to communicate beyond the employee (n349)	Simple narratives, no attempt at serious legacy code (n350)	conferences, talks, monthly updates, blogs (n351)	workshops, conferences, talks, monthly updates, blogs, reproducible research is taken seriously (n352)

Appendix G: Data Literacy Maturity Questionnaire Concept– Alpha Version

n	Scorecard Category Question	Answer
	Organization	
1	What organization are you working for?	TEXT
2	What is your position / job title?	TEXT
3	Which topic is your organization dedicated to?	TEXT
4	How many people are working in your organization?	VALUE
	Problem Definition	
4	What is the problem you want to solve using data?	TEXT
6	What do you expect from working with data?	TEXT

	Data Literacy Maturity					
	Organi-zational	Question	Answer Level 1	Answer Level 2	Answer Level 3	Answer Level A4
7	Data Culture	Which statement best describes the culture in your organization regarding working with data?	Data is perceived as an ambiguous term which causes insecurities. No concern for and no employees with significant data science background.	Data is perceived as an interesting concept and benefits are appreciated. Insecurities exist regarding use cases and what exactly to expect. Employees with some data analytics knowledge work in prominent positions.	Data is not perceived as a source of insecurities, but is rather understood as an enabler for progress and support for existing and planned activities. Higher management and Project managers are fully aware of the need to collect, store and act on data. Data analysis is on the roadmap.	Psychological barriers of data have been brought down and comfort around data is promoted. Data analysis/insights are implemented and optimized at all levels of the workflow. Data literacy is an important skill for most employees even at a senior level and dedicated resources (time, budget, human resources) for data handling and conversion exist.
8	Data Ethics and Data Security	Which statement best describes your organization's take on data ethics and security?	Data is handled with no concern for security or ethical dilemma. No guidelines are defined that ensure confidentiality, integrity and availability of data.	Some employees and projects are aware of security issues; uncoordinated attempts to handle data securely. Rising awareness of the importance of the responsible use of data. No defined guidelines.	Data protection guidelines are defined and responsible data releases described internally. Organization-wide policies for secure and ethically sound data handling exist.	Consistent, companywide policies for secure and ethically sound data handling are constantly redefined and updated. Senior employees are responsible for the enforcement of these rules. Processes are in place to ensure confidentiality, integrity, and availability of data.
	Individual					
9	Ask question / define	How do you evaluate your ability to formulate questions in order to find meaningful answers in data?	No feeling for what questions can be answered by data.	Ability to ask questions that can be answered by simple data queries.	Entire projects are based on multidimensional questions that need complex data queries and multiple iterations to resolve.	Entire projects are based on multidimensional questions that need complex data queries and multiple iterations to resolve. No single person can handle these inquiries.
10	Find	Which data sources / strategies do you use to find data?	Basic use of search engines.	Advanced use of search engines that might be complemented by asking for data sources at public authorities (IFG - freedom of information act).	Awareness and use of data portals / public databases for specific topics and more sophisticated search engines.	Publishing own data and matching different sources through complex database queries.

11	Get	Which strategies do you use in order to get data?	No awareness of different data formats, information is derived from full text and used as basis for further processing.	Use of downloads and data formats such as .csv.	Use of more complex data formats such as JSON and XML, use of APIs.	Automatic system updates that new data exists and can be integrated.
12	Verify	How do you critically assess the quality of your data source?	Data is taken at face value. No awareness of how and why to critically assess data.	Critical check of simple data quality measures (e.g. completeness, duplicates). Data source is not critically assessed.	Multiple layers of data checking are implemented in standard procedures across the organization. Data sources are checked, different sources are considered and critically assessed.	Best practices exist, employees are trained in data verification procedures and priorities. Data sources are used transparently and provide access to the original sources so that audience can check for themselves.
13	Clean	Which strategies do you use to clean your data?	No awareness that given data might have to be checked, cleaned or normalized. Data is taken as is.	Growing awareness that data is not properly formated, manual fixing of errors in excel, basic data quality indices are defined (e.g. simple dealing with duplicates, missing values and extreme outliers).	Use of programs to support data cleaning (e.g. Open Refine). Sophisticated duplicate detection and string manipulation. Awareness for normalization to get comparability between different data sources, high awareness for data quality criteria (e.g. machine processable, empty fields, uniform and complete columns and rows, outliers).	Automated cleaning of data using a script (e.g. in R). Data cleaning is regarded as a challenging modeling tasks.
14	Analyze	How do you analyze your data?	Simple use of data tables to read information (e.g. looking for all values of specific group, bar and pie charts, basic summaries of data)	Pivot tables, multidimensional tables and summaries, histograms, boxplots, still mainly descriptive.	Inferential view of the data, full account of uncertainties, linear regression, decision trees.	Full suite of machine learning tools, clustering, forecasting, boosting, ensemble learning.
15	Visualize	Which strategy do you use to visualize your data?	Bar and pie charts, basic summaries of data. No understanding of when standard visualizations are chosen, decision based on what looks best (trial and error).	Pivot tables, multidimensional tables and summaries, histograms, boxplots, still mainly descriptive. Different visualizations are created in Excel in accordance with information that want to be represented (e.g. time related data requests line diagram).	Interactive charts, uncertainties are always visualized along with the data. Use of visual analytics tools (e.g. Tableau or CartoDB).	Linked dynamic dashboards, anticipating user requests. Own, sophisticated visualizations are programmed / created that exceed standard diagram visualizations.
16	Commu-nicate	How do you communicate your data outputs?	Insights from data are not communicated or put into a broader context.	Static visualizations support simple narrative (e.g. in blogposts). Use of presentation tools (e.g. Powerpoint, Prezi).	Own projects are supported by interactive visualizations (e.g. conferences, talks, monthly updates, blog posts).	Data-driven campaigning. Workshops, conferences, talks, monthly updates, blog posts, reproducible research is taken seriously.
Results			Spider-web diagram, Data Literacy Maturity Grid			

Testing 1

	Very high	High	Neutral	Low	Very low
Correctness			x		
Flexibility			x		
Usability		x			
Implementability	x				
Economic Efficiency	x				

Testing 2

	Very high	High	Neutral	Low	Very low
Correctness		x			
Flexibility			x		
Usability				x	
Implementability	x				
Economic Efficiency	x				

Appendix H: Data Literacy Maturity Questionnaire Concept– Beta Version

Introduction	
Background	Within the course of the last years, data literacy turned out to be a meaningful priority to different groups. Educational programs and nonprofit organizations designed programs to familiarize children with coding, but also the emerging importance of data journalism and analytical skills on the job market, show that we have to equip ourselves with new skills for the data era. At the same time, there is a lack of systematic definitions and procedures regarding data literacy. Hence, this self-evaluation will help you to better understand, appreciate and evaluate your data handling skills.
Purpose	"This self-evaluation will help you to better understand the required skills that are needed to kick-off your data projects, identify strengths and gaps and thus will empower you to plan your future in accordance with your goals. It should more precisely help you: * To reflect on, analyze and record your existing competencies or the skills, abilities and knowledge connected to data handling skills * To better understand the type of behaviors that might be expected when working with data. * With practical advice and suggestions on how you might continue to develop your data practice. * To set out a personal development plan to identify the activities you feel are important to developing your practice."""
Who should use this model and when might you use it?	We hope that the data literacy model will be useful for individuals and organizations that work digitally no matter whether you are a newcomer to working with data or are already experienced. Of course, considering the individual and organizational context is key during the evaluation. It will be of most value when it is used as part of an ongoing reflective process building on your personal experiences. Some ideas about how and when to use the framework include: * Using it to value existing competencies and to make decisions on training and development needs. * Supporting individuals and organizations at the start of a data assignment or project to understand the competencies required, to identify their personal starting points and as a reflection and development tool. * Taking individual competencies and using them as a starting point for a data mentoring partnership. * Using it as a planning and reflection tool for individual or collaborative projects with other data practitioners as part of feedback sessions with partners and practitioners when the reflection of others can be used alongside personal reflections.

n	Scorecard Category	Question	Answer
	About you		
1		What organization are you working for?	TEXT
2		What is your position / job title?	TEXT
3		Which topic is your organization dedicated to?	TEXT
4		How many people are working in your organization?	VALUE
	Problem Definition		
4		What is the problem you	TEXT

		want to solve using data?				
6		What do you expect from working with data?	TEXT			
	Data Literacy Maturity					
	Organizational	Question	Answer	Answer	Answer	Answer
Points			2	1	4	3
7	Data Culture	Which statement best describes the culture in your organization regarding working with data?	Data is perceived as an interesting concept and benefits are appreciated. Insecurities exist regarding use cases and what exactly to expect. Employees with some data analytics knowledge work in prominent positions.	Data is perceived as an ambiguous term which causes insecurities. No concern for and no employees with significant data science background.	Psychological barriers of data have been brought down and comfort around data is promoted. Data analysis/insights are implemented and optimized at all levels of the workflow. Data literacy is an important skill for most employees even at a senior level and dedicated resources (time, budget, human resources) for data handling and conversion exist.	Data is not perceived as a source of insecurities, but is rather understood as an enabler for progress and support for existing and planned activities. Higher management and Project managers are fully aware of the need to collect, store and act on data. Data analysis is on the roadmap.
Points			3	4	2	1
8	Data Ethics and Data Security	Which statement best describes your organization's take on data ethics and security?	Data protection guidelines are defined and responsible data releases described internally. Organization-wide policies for secure and ethically sound data handling exist.	Consistent, companywide policies for secure and ethically sound data handling are constantly redefined and updated. Senior employees are responsible for the enforcement of these rules. Processes are in place to ensure confidentiality, integrity, and availability of data.	Some employees and projects are aware of security issues; uncoordinated attempts to handle data securely. Rising awareness of the importance of the responsible use of data. No defined guidelines.	Data is handled with no concern for security or ethical dilemma. No guidelines are defined that ensure confidentiality, integrity and availability of data.
	Individual					
Points			2	1	3	4
9	Ask question & Define	How do you evaluate your ability to formulate questions to find meaningful answers in data?	Ability to ask questions that can be answered by simple data queries.	No feeling for what questions can be answered by data.	Entire projects are based on multidimensional questions that need complex data queries and multiple iterations to resolve.	Entire projects are based on multidimensional questions that need complex data queries and answer all main questions and sub-questions. No single person can handle these inquiries.
Points			3	2	4	1
10	Find	Which data sources / strategies do you use in order to find data?	Awareness and use of data portals / public databases for specific topics and more sophisticated search engines.	Advanced use of search engines and use of internal data sources that might be complemented by asking for data sources at public authorities (FOI - freedom of information act).	Publishing own data and matching different sources through complex database queries.	Basic use of search engines.
Points			1	3	4	2
11	Get	Which strategies do you use to get data?	No awareness of different data formats, information is derived from full text mostly from PDFs and used as basis for further processing. Information is rarely processed digitally.	Use of more complex data formats such as JSON and XML, use of APIs.	Automatic system updates that new data exists and can be integrated.	Use of downloads and data formats such as .csv. Often use of internal data to access data.
Points			3	4	1	2

12	Verify	How do you critically assess the quality of your data source?	Multiple layers of data checking are implemented in standard procedures. Data sources are checked, different sources are considered and critically assessed.	Best practices exist, employees are trained in data verification procedures and priorities. Data sources are used transparently and provide access to the original sources so that audience can check for themselves.	Data is taken at face value. No awareness of how and why to critically assess data.	Critical check of simple data quality measures (e.g. completeness, duplicates). Data sources is not critically assessed.
Points			1	4	3	2
13	Clean	Which strategies do you use to clean your data?	No awareness that given data might have to be checked, cleaned or normalized. Data is taken as is.	Automated cleaning of data using a script (e.g. in R). Data cleaning is regarded as a challenging modeling tasks.	Use of programs to support data cleaning (e.g. Open Refine). Sophisticated duplicate detection and string manipulation. Awareness for normalization to get comparability between different data sources, high awareness for data quality criteria (e.g. machine processable, empty fields, uniform and complete columns and rows, outliers).	Growing awareness that data is not properly formatted, manual fixing of errors in excel, basic data quality indices are defined (e.g. simple dealing with duplicates, missing values and extreme outliers).
Points			4	2	1	3
14	Analyze	How do you analyze your data?	Full suite of machine learning tools, clustering, forecasting, boosting, ensemble learning.	Pivot tables, multidimensional tables and summaries, histograms, boxplots, still mainly descriptive.	Simple use of data tables to read information (e.g. looking for all values of specific group, bar and pie charts, basic summaries of data)	Exploration of relationship between variables in a dataset, full account of uncertainties, linear regression, decision trees.
Points			1	4	2	3
15	Visualize	Which strategy do you use to visualize your data?	Bar and pie charts, basic summaries of data are created in Excel. No understanding of when standard visualizations are chosen, decision based on what looks best (trial and error).	Linked dynamic dashboards, anticipating user requests. Own, sophisticated visualizations are programmed / created that exceed standard diagram visualizations.	Pivot tables, multidimensional tables and summaries, histograms, boxplots, still mainly descriptive. Different visualizations are created in Excel in accordance with information that want to be represented (e.g. time related data requests line diagram).	Interactive charts, uncertainties are always visualized along with the data. Use of visual analytics tools (e.g. Tableau or CartoDB).
Points			2	3	1	4
16	Communicate	How do you communicate your data outputs?	Static visualizations support simple narrative (e.g. in blogposts). Use of presentation tools (e.g. Powerpoint, Prezi).	Own projects are supported by interactive visualizations (e.g. conferences, talks, monthly updates, blog posts).	Insights from data are not communicated or put into a broader context.	Data-driven campaigning. Workshops, conferences, talks, monthly updates, blog posts, reproducible research is taken seriously.
Points			1	4	2	3
17	Assess & Interpret	How do you evaluate your ability to assess and interpret data outputs?	Data outputs are used at face value without questioning their correctness and message	Data outputs and results are consistently questioned and challenged, interpretation extents the obvious and information are successfully translated into actionable knowledge.	Growing awareness for critically assessing data outputs and interpreting the results. Insecurities regarding what exactly to pay attention to.	Data outputs and results are interpreted confidently and critically. Evaluation criteria are internalized.
Results Spider-web diagram, Data Literacy Maturity Grid						

List of References

Argyris, C. (1970), *Personality and Organization; the Conflict between System and the Individual.*, Harper & Row, New York.

Argyris, C. and Schön, D.A. (1978), *Organizational Learning*, Addison-Wesley Pub. Co, Reading, Mass.

Becker, J., Knackstedt, R. and Pöppelbuß, J. (2009), “Developing Maturity Models for IT Management: A Procedure Model and its Application”, *Business & Information Systems Engineering*, Vol. 1 No. 3, pp. 213–222.

Bhargava, R. and D’Ignazio, C. (2015), “Designing Tools and Activities for Data Literacy Learners”, available at: https://datatherapy.files.wordpress.com/2015/07/designing-tools-and-activities-for-data-literacy-learners.pdf (accessed 27 October 2016).

de Bruin, T., Rosemann, M., Freeze, R. and Kulkarni, U. (2005), “Understanding the Main Phases of Developing a Maturity Assessment Model”, *16th Australasian Conference on Information Systems*, available at: https://www.researchgate.net/publication/27482282_Understanding_the_Main_Phases_of_Developing_a_Maturity_Assessment_Model (accessed 20 December 2016).

Calzada Prado, J. and Marzal, M.Á. (2013), “Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents”, *Libri*, Vol. 63 No. 2, pp. 123–134.

Cambridge Dictionary. (2016), “Meaning of maturity in the English dictionary”, available at: http://dictionary.cambridge.org/dictionary/english/maturity.

Chall, J.S. (1983), *Stages of Reading Development*, McGraw-Hill, New York.

Chiesa, V., Frattini, F., Lazzarotti, V. and Manzini, R. (2008), “Designing a performance measurement system for the research activities: A reference framework and an empirical study”, *Journal of Engineering and Technology Management*, Vol. 25 No. 3, pp. 213–226.

Crosby, P.B. (1980), *Quality Is Free: The Art of Making Quality Certain*, Mentor, New York.

Data-Pop Alliance. (2015), “Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data”, available at: https://datatherapy.files.wordpress.com/2015/10/beyond-data-literacy-2015.pdf (accessed 26 October 2016).

Datenschule. (2016), “OKF Datenschule”, *Datenschule*, available at: https://dev.datenschule.de/ (accessed 31 October 2016).

Deahl, E. (2014), *Better the Data You Know: Developing Youth Data Literacy in Schools and Informal Learning Environments*, Massachusetts Institute of Technology.

Debnath, S., Lee, B. and Tandon, S. (2015), “Fifty years and going strong: what makes behaviorally anchored rating scales so perennial as an appraisal method?”, *International Journal of Business and Social Science, Vol. 6, No. 2*, pp. 16–25.

Desouza, K.C. and Smith, K.L. (2014), “Big Data for Social Innovation (SSIR)”, *Stanford Social Innovation Review*, available at: https://ssir.org/articles/entry/big_data_for_social_innovation (accessed 14 November 2016).

Enkel, E., Bell, J. and Hogenkamp, H. (2011), “Open Innovation Maturity Framework”, *International Journal of Innovation Management*, Vol. 15 No. 6, pp. 1161–1189.

Felden, C. (2006), “Integrating structured and unstructured data in a b Integrating structured and unstructured data in a business usiness usiness-intelligence ntelligence ntelligence-system”, available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.467.8602&rep=rep1&type=pdf (accessed 27 October 2016).

Flyverbom, M. and Koed Madsen, A. (2015), “Sorting data out - unpacking big data value chains”, *Die Gesellschaft Der Daten: Über Die Digitale Transformation Der Sozialen Ordnung*, Transcript, Bielefeld, pp. 93–121.

Fraser, P., Moultrie, J. and Gregory, M. (2002), “The use of maturity models/grids as a tool in assessing product development capability”, Vol. 1, IEEE, pp. 244–249.

Fry, B.J. (2004), *Computational Information Design*, Massachusetts Institute of Technology, Boston.

Gottschalk, P. (2009), “Maturity levels for interoperability in digital government”, *Government Information Quarterly*, Vol. 26 No. 1, pp. 75–81.

Gray, J., Chambers, L. and Bounegrou, L. (2012), *The Data Journalism Handbook*, available at: https://openzazu.files.wordpress.com/2012/08/the-data-journalism-handbook-gnv64.pdf (accessed 27 October 2016).

Hellberg, A.-S. and Hedström, K. (2015), “The story of the sixth myth of open data and open government”, *Transforming Government: People, Process and Policy*, Vol. 9 No. 1, pp. 35–51.

Hevner, A., March, S., Park, J. and Ram, S. (2004), “Design Science in Information Systems Research”, *MIS Quarterly*, pp. 76–105.

Hillson, D.A. (1997), “Towards a Risk Maturity Model”, *The International Journal of Project and Business Risk Management, Vol. 1, No. 1,* pp. 35–45.

Hughes, H., Williamson, K. and Lloyd, A. (2007), "Critical incident technique", *Exploring Methods in Nformation Literacy Research. Topics in Australasian Library and Information Studies, Number 28, Centre for Information Studies, Charles Sturt University, Wagga Wagga, N.S.W.*, pp. 49–66.

Kanter, B., Delahaye Paine, K. and Paarlberg, W.T. (2012), *Measuring the Networked Nonprofit: Using Data to Change the World*, First edition., Jossey-Bass, a Wiley Imprint, San Francisco, California.

King, G. (2014), "In: Why 'Big Data' Is a Big Deal", *Harvard Magazine*, 18 February, available at: http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal (accessed 24 November 2016).

King, W.R. and Teo, T.S.H. (1997), "Integration Between Business Planning and Information Systems Planning: Validating a Stage Hypothesis", *Decision Sciences*, Vol. 28 No. 2, pp. 279–308.

Kütting, G. (2009), *Conventions, Treaties and Other Responses to Global Issues*, EOLSS, in partnership with UNESCO, Oxford.

Loch, C.H. and Tapper, U.A.S. (2002), "Implementing a strategy-driven performance measurement system for an applied research group", *Journal of Product Innovation Management*, Vol. 19 No. 3, pp. 185–198.

Maier, A., Moultrie, J. and Clarkson, J. (2012), "Developing Maturity Grids for Assessing Organisational Capabilities: Practitioner Guidance", *IEEE TRANSACTIONS ON ENGINEERING MANAGEMENT, VOL. 59, NO. 1*, pp. 138–159.

Mandinach, E.B. and Gummer, E.S. (2013), "A systemic view of implementing data literacy in educator preparation", *Educational Researcher, 42(1),* pp. 30–37.

Marshall, M.N. (1996), “Sampling for qualitative research”, *Oxford University Press, Vol. 13, No. 6*, pp. 522–525.

McCormack, K., Willems, J., van den Bergh, J., Deschoolmeester, D., Willaert, P., Indihar Štemberger, M., Škrinjar, R., et al. (2009), “A global investigation of key turning points in business process maturity”, *Business Process Management Journal*, Vol. 15 No. 5, pp. 792–815.

McGrath, M.E. and McGrath, M.E. (Eds.). (1996), *Setting the PACE in Product Development: A Guide to Product And Cycle-Time Excellence*, Rev. ed., Butterworth-Heinemann, Boston, MA.

Mettler, T. and Rohner, P. (2009), “Situational maturity models as instrumental artifacts for organizational design”, ACM Press, p. 1.

Meuser, M. and Nagel, U. (2009), “Das Experteninterview — konzeptionelle Grundlagen und methodische Anlage”, in Pickel, S., Pickel, G., Lauth, H.-J. and Jahn, D. (Eds.), *Methoden der vergleichenden Politik- und Sozialwissenschaft*, VS Verlag für Sozialwissenschaften, Wiesbaden, pp. 465–479.

Paulk, M., Curtis, B., Chrissis, M. and Weber, C. (1993), *Capability Maturity Model for Software*, Software Engineering Institute, Carnegie 1188, Mellon University.

Pawelke, A. and Tatevossian, A.R. (2016), “Data Philanthropy: Where Are We Now? | United Nations Global Pulse”, *United Nations Global Pulse*, 13 November, available at: http://www.unglobalpulse.org/data-philanthropy-where-are-we-now (accessed 13 November 2016).

Pöppelbuß, J. and Röglinger, M. (2011), “What makes a useful maturity model? A framework of general design principles for maturity models and its demonstration in business process management”, *European Conference on Information Systems, ECIS 2011 Proceedings*, available at:

http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1027&context=ecis2011 (accessed 13 August 2016).

Ross, J. (2003), "Creating a strategic IT architecture competency: learning in stages", *MIS Quarterly Executive Vol. 2 No. 1, University of Minnesota*, available at: http://www.umsl.edu/~lacitym/misqearch.pdf (accessed 21 October 2016).

Rummler, G.A. and Brache, A.P. (1995), *Improving Performance: How to Manage the White Space on the Organization Chart*, 2nd ed., Jossey-Bass, San Francisco, CA.

Sein, M., Henfridsson, O., Purao, S., Rossi, M. and Lindgren, R. (2011), "Action Design Research", *MIS Quarterly Vol. 35 No. 1*, pp. 37–56.

Shaw, J. (2014), "Why 'Big Data' Is a Big Deal", *Harvard Magazine*, 18 February, available at: http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal (accessed 24 November 2016).

Slater, D. (2016), "Research Results Part 1: Defining Data Literacy", *School of Data*, 8 January, available at: http://schoolofdata.org/2016/01/08/research-results-part-1-defining-data-literacy/ (accessed 27 October 2016).

Snell, S., Morris, S. and Bohlander, G.W. (2016), *Managing Human Resources*, 17e ed., Cengage Learning, Boston, MA, USA.

Stebbins, R.A. (2001), *Exploratory Research in the Social Sciences*, Sage Publications, Thousand Oaks, Calif.

Süssenguth, F. (2015), "Die Organisation des digitalen Wandels - Zur Funktion von Digitalisierungssemantiken", *Die Gesellschaft Der Daten: Über Die Digitale Transformation Der Sozialen Ordnung*, Transcript, Bielefeld, pp. 93–121.

UN Data Revolution Group. (2014), *A World That Counts - Mobilising the Data Revolution for Sustainable Development*, United Nations Secretary-General's

Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG), available at: http://en.unesco.org/un-sab/system/files/Data%20Revolution%20Report%20A-World-That-Counts_0.pdf (accessed 27 October 2016).

UNESCO. (2004), *The Plurality of Literacy and Its Implications for Policies and Programmes*, Paris, available at: http://unesdoc.unesco.org/images/0013/001362/136246e.pdf (accessed 27 October 2016).

Vahey, P., Yarnall, L., Patton, C., Zalles, D. and Swan, K. (2006), “Mathematizing middle school: Results from a cross-disciplinary study of data literacy”, *Annual Meeting of the American Educational Research Association*, presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA, available at: https://www.sri.com/sites/default/files/publications/imports/TWD_AERA_Mathematizing_Middle_School.pdf (accessed 1 December 2016).

Vallerand, J., Lapalme, J. and Moïse, A. (2015), “Analysing enterprise architecture maturity models: a learning perspective”, *Enterprise Information Systems*, pp. 1–25.

Webster, J. and Watson, R.T. (2002), “Analyzing the past to prepare for the future: Writing a literature review”, *MIS Quarterly, Vol. 26, No. 2*, pp. xiii–xxiii.

Wolff, A., Gooch, D., Cavero Montaner, J.J., Rashid, U. and Kortuem, G. (2016), “Creating an understanding of data literacy for a data-driven society”, *The Journal of Community Informatics*, pp. 9–26.

Yin, R.K. (2014), *Case Study Research: Design and Methods*, Fifth edition., SAGE, Los Angeles.