Karlsruhe Institute of Technology

Data Management for Meta-Analyses of Causal Models and Measurements in Survey Research

Master's thesis at the Chair of Information Systems and Electronic Markets
Prof. Dr. Andreas Geyer-Schulz
Institute of Information Systems and Marketing
Department of Economics and Business Engineering
Karlsruhe Institute of Technology

of

Thomas Hummel

Matriculation Number: 1417443

Supervisor:
Prof. Dr. Andreas Geyer-Schulz

Date of Registration: February 16th, 2015
Date of Submission: November 16th, 2015

Declaration of Academic Integrity

I hereby declare that this thesis has been written only by the undersigned and without any assistance from third parties. Furthermore, I confirm that no sources have been used in the preparation of this thesis other than those indicated in the thesis itself.

I agree to the publication of this work on the website of the Chair of Information Services and Electronic Markets at the Institute Information Systems and Marketing of the Karlsruhe Institute of Technology (KIT).

Karlsruhe, November 16th, 2015

Abstract

Crucial to social research is the sustainable design and repeated usage of measurement instruments. In addition, it is important that researchers are able to find and compare relevant models and scales in order to make appropriate decisions.

Therefore we propose a database that allows to create a joint collection of (causal) models and (multi-item) scales in order to enable a more extensive data usage for automation of scientific workflows, for the generation of recommendations and for meta-analyses. This includes finding appropriate constructs and scales, comparisons of items and quality measures as well as the detection of undocumented links between different topics and disciplines.

In order to illustrate the potential, we refer to several approaches of measuring Technology Acceptance.

Kurzreferat

Entscheidend für die Sozialforschung ist die nachhaltige Gestaltung und wiederholte Verwendung von Messinstrumenten. Außerdem ist es wichtig, dass Wissenschaftler in der Lage sind relevante Modelle und Skalen sowohl zu finden als auch zu vergleichen, um damit passende Entscheidungen treffen zu können.

Daher schlagen wir eine Datenbank zur gemeinsamen Sammlung von (kausalen) Modellen und (Multi-Item-) Skalen vor, die Automatisierung von wissenschaftlichen Prozessen, Generierung von Empfehlungen und Meta-Analysen ermöglicht. Dazu gehört die Suche nach geeigneten Konstrukten und Skalen, der Vergleich von Items und Qualitätsmaßen, sowie die Entdeckung von undokumentierten Verbindungen zwischen verschiedenen Themen und Disziplinen.

Um dieses Potenzial zu aufzuzeigen, verwenden wir verschiedene Ansätze zur Messung von Technologieakzeptanz.

Contents

1. Motivation

According to Link [Lin11], survey researchers are not good enough at leveraging new technologies. He states that the measurement of behavior or preferences is largely stuck in using interviews or surveys to gather that data. Furthermore, there is less research on those traditional techniques and many expensive research labs at universities have been suspended.

A typical research process for a social study consists of many steps that are semantically linked but are not supported by tools. In the beginning a researcher has to identify the scope of the field at hand. This includes literature research to identify important keywords and contributors, that allow to create an overview. Using the work of influential researchers, articles of journals in that field and related papers or standards, allows to identify the state of the art. Subsequently, the researcher has to use or combine the most adequate theories or measurements in order to conduct his or her study. These studies are often based on questionnaires with the selected indicators. During analysis, the underlying models can be used to validate the models in the new context or in the new form. Additionally, they can be compared with historical findings in order to prove or improve the existing theories.

Many researchers follow these steps on their own and repeat this process for every new study. Especially young scientists or students, who are less experienced, need more time to evaluate a field. Additionally, there are research groups that divide the labor and exchange knowledge. Given the amount of articles on a specific topic, this is the only feasible option in the long run. As an example, Google Scholar provides 2.320.000 results when searching for technology acceptance, without using synonyms that may be used in other disciplines.[1]

Hence, we propose the development of an infrastructure for supporting the scientists in getting an overview of surveys and the state-of-the art in their domain on the one hand and to conserve the knowledge for future research on the other hand. Every user should be able to add information or link information and should get a better overview in return.

Selected uses for a knowledge base as proposed in this thesis are:

[1] Technology Acceptance on Google Scholar: `https://scholar.google.de/scholar?q=technology+acceptance`, number of results as of November 2015

Data Extraction and Collection A standardized entry process, supports researchers with the identification of models, concepts and measurements described in a paper. They can extract this information in a structured way.

Comparison of Theories With the help of the structured data, the comparison of articles becomes easier and more focused on general attributes. Details can still be found in the respective articles. This is possible, regardless if the goal is to compare a fixed set of articles, to search for useful scales in order to create a questionnaire or to collect information for survey articles.

Validating Theories An important aspect of science is to repeat experiments and measurements in other setups and contexts in order to prove the existing hypotheses. The easier it is to reuse measurement instruments, the more likely it is that someone will do that. Additionally, it is important keep track of the ongoing development of certain theories, they may have been adjusted due to other assumptions or findings.

Division of Labor Collaborative work needs a certain form of standardization and data exchange. With the help of the proposed system, groups of researchers can share the effort of data entry and benefit from each other. This is not necessarily limited to single teams but might be used across projects or even in the public domain. People with different levels of expertise can provide information.

Support of Subsequent Processes Structured data can be used as a guidance for a researcher's own work: It might be possible to generate parts of (LaTeX-) documents. Questionnaires could be generated with the help of existing scales and indicator variables. Analyses could be enhanced with providing the semantic relationships of variables as used in a theoretical model.

The proposed approach can be helpful in many stages of an academic career, from students to professionals. Some examples are the following:

Students Students are guided during information extraction from articles and benefit from the structured comparisons. When focusing on literature research, they can be told to read papers and extract the important aspects. Finally, they can use the database to compare the entered information, for example in order to write a term paper. Students that are conducting a larger study, for example in their thesis, benefit from data and information that is easier to access and to reuse. This includes theory building, questionnaire design and analysis of the results. That is why this leads to studies that actually might be of a higher quality, thus of a higher value for the scientific community.

Young Academics Young academics often read hundreds of papers in order to gain knowledge in their field. A structured data store allows them to focus on the content and not on the way how they organize their knowledge for common tasks. With the help of a database, they can use existing data entered by colleagues or students and can select important theories with less effort. Writing a PhD thesis, for example, takes a long time and details of the literature work might be forgotten when the results of a study have to be analyzed in the end. This could be reduced with a good infrastructure.

Researchers in General Working on different projects in different fields requires a good organization of knowledge. The possibility to get a quick overview of the latest developments, without relying on colleagues to write a comprehensive survey paper is important. New ideas can be introduced quickly and can lead to other points of view. Collaborating across disciplines or fields allows to detect unknown analogies and can point out different approaches.

Although research in the social sciences, psychology or marketing is sometimes seen as an easy observational task, it is rather complicated to explore or explain human behavior in general or to derive ubiquitous laws, that hold for everybody, everywhere, and every time, regardless of any demographic, cultural or physical attitudes.

Existing workflows do not always cover the latest processes, methods and technologies, that improve efficiency, quality and scientific value. The sole publication of written papers without data is an obvious hassle in evaluating the quality of many studies. It is impossible to use new data analysis techniques on these results or compare results of different studies with each other. In the same manner, the use of scales handbooks, as written books or in electronic versions, does not provide the necessary structure and semantic for new techniques. Furthermore, documenting new studies and results includes many steps that could easily be automated, when the underlying theories and concepts were documented properly.

Throughout the work, we will use the field of technology acceptance as an example use case. Seiffer [Sei13] has examined several models to measure technology acceptance of consumers. She compared scales from marketing, computer science and business informatics and concluded that researchers might not be able to include all related studies. The information overload and the rapid generation of knowledge leads to the fact that the different fields seem to use and create different scales for the same concepts and constructs. But even within some research fields some researchers do not consider publications of others, they do not even mention them. This is a major problem for quality of new work in particular and for the social sciences in general. (See also section 1.2).

1.1 Goals, Limits and Scope

We propose a core database that stores models and concepts with measurements, primarily based on multi-item scales, as well as relations to link them manually with each other. As an outcome, the different publications in which a model or measurement was mentioned can be shown. Furthermore, the various updates or extensions of certain scales can be tracked and a comparison of similar scales can be achieved.

We concentrate on operational definitions that contain indicator items measured with rating scales. These most often form a Likert-scale, because it is currently the predominant scale type in social research. Moreover, we handle only unidimensional scales in this work, even though we try to allow possible extensions in the future.

These future implementations shall cover more scale types, multidimensional scales, improved editing, recommender systems, the storage of survey results or automated analyses.

Nonetheless the proposed database of models and measurements can not provide fully-automated analyses. It should provide customizable interfaces for data export as well. Knowledge is only generated when people think, so technology has to support human thinking and can not create knowledge on its own [see McD00].

1.2 Use Case: Measuring Technology Acceptance

There are many different approaches on how to measure consumers' attitudes towards new technology. Users can use many different devices, applications and services and every day new technologies are being developed and presented. For producers, manufacturers and developers it is important to know whether the consumers accept the new technologies and finally will be using them.

Some of these approaches have been evaluated by Seiffer [Sei13] in order to propose a frame for a systematic comparison. Besides that, she outlines differences in the (re-)use of models in different research fields.

Seiffer [Sei13, pp. 69–72] states that "there are huge differences in both research and model development processes if one considers the different scientific disciplines". Actually, the TAM model by Davis [Dav89] and several redefinitions and extensions seem to be pretty popular amongst information systems researchers. Marketers published "more and more different models" and references were "almost usually limited to the TAM. [...] Models are not based on each other and there is no continued attempt to adapt and improve established models" [Sei13, p. 71]. Additionally, "Computer science research is more or less ignored in both fields. The ISO standards were proposed later than the TAM and other technology adoptions, still, there is a long history of quality standards in computer science research, such are for example proposed by Cavano and McCall [CM78]." [Sei13, p.71].

These results provide a view on research in the particular fields that is dissatisfying with regards to good scientific practice [see for example Deu13]. This might not be sloppy work or even intended fraud. It rather shows the necessity for supportive tools in order to keep track of all the other publications in the field and across disciplines.

Seiffer [Sei13, p. 1] did the comparison of the models in a manual way with the help of spreadsheets and proposes to set up an infrastructure for that: "Ideally, there is a database containing all this information, however, actually creating this database would go beyond the scope of this thesis. Therefore, this research proposes a comparison which then could be used to create an approach for automating model comparison."

In this work we adapt her idea to design a database in order to conduct comparisons in a more efficient and less error-prone way. Since more people can work at the same time, it is also possible for research groups to maintain a common base of information that increases over time.

1.3 Structure of the Document

In order to emphasize the relevance of a supportive tool for researchers, we commented on the different models used for technology acceptance in section 1.2. This illustrates some researchers' missing knowledge of related work. This should be prevented. Within the subsequent chapters we will use this field as example whenever possible.

In the foundations (2) we give a short theoretical overview of the philosophy of the social sciences (2.1) and its special characteristics in comparison to natural sciences as well as the idea of causal explanations and causal models (2.1.4). Then we report on scaling and measurement theory (2.2) and its important background for social research.

We comment on related work in chapter 3 with regard on how collections of measurements (3.1) and models (3.2) are presented and discuss their assets and drawbacks as well as their ease of use.

System Design details are covered in chapter 4: At first we are having a look at the system requirements in section 4.1. Next, section 4.2 describes the relevant domain and data models. This is followed by some comments and user interface design skesketches in section 4.3. Finally, we describe the concrete implementation in section 4.4.

This is followed by an evaluation (Chapter 5) where we point out strengths and weaknesses of the implementation at hand and section 5.2 proposes several possible extensions in order to improve the existing features or to extend the functionality in various ways.

Finally, we provide a conclusion in chapter 6 of this work whilst we refer to the practical benefits once again.

2. Foundations

The following foundations are intended to provide an overview of the most important topics that are related to survey-related data and, therefore, important for the design of a model and measurement database.

Handling survey-related and behavioral data is relevant for every discipline that asks people for their opinions, their attitudes or their feelings. Business studies focus on marketing, organization theory, employee and customer satisfaction or ergonomics. Sociology, economics and political sciences are explaining social systems, the distribution of labor, power and wealth; and they discuss cultural characteristics. Medicine and behavior research investigate physical or mental health issues, values, emotions or happiness as well as interpersonal relationships and communication. Additional disciplines are psychology, educational research or even computer science, amongst others. In the following we will talk about the “social sciences” as a major category of all these different branches and assume that the methodology is based on a common ground. “Asking people” in this context does not necessarily mean “interviewing” or “using questionnaires”, since it also could include sensor measurements like eye tracking or observations from third parties.

Research on personality traits, thus thoughts, behaviors and feelings of humans, and their effect on decisions, actions and opinions has different origins. People differ in their physical appearance (gender, strength, size) and in their mental abilities. Both, physical appearance and mental abilities, as well as the influence of their social interactions, might influence certain traits. In order to detect possible cause-and-effect relations, scientists try to observe behavior patterns, measure physical indicators (i.e. blood pressure, brain activity) and try to identify the characteristics of the social environment. Most often a study contains at least a questionnaire or is based on a questionnaire alone. Questionnaires can be filled by an interviewer or by the interviewee himself, either on- or offline. People are asked to express their feelings or attitudes in given situations or they are asked to compare themselves with others.

In a questionnaire, participants have to answer a set of indicator questions that enable researchers to detect underlying latent concepts. These concepts, for example the self-confidence of a person, allow to predict how people react to new challenges. Many of the physiological attributes can also be asked in a questionnaire, such as age or weight and do not have to be measured separately. Therefore, we focus on questionnaire development as foundation for social research regardless of the various disciplines and cover scales for questionnaires as major commonality.

According to Schnell, Hill, and Esser [SHE05, pp. 53–57], the main task of scientific work in the social sciences is to explain social events or facts[1]. This can be divided into three abstraction levels as done by [Wv02, p. 18]: real world phenomena, concepts of the communication level and the measurement world based on different variables.

Hence, in the beginning we discuss the way scientists should address new research and aim for meaningful outcomes. In particular this includes different philosophical points of view on how one should proceed from hypotheses or theories to descriptive findings. Here we summarize some basic terms and the most popular philosophical theories of the social sciences. The idea of causality, as relation between a cause and an effect, will be introduced. Most findings (ought to) support such a relation and, therefore, it is crucial for the understanding of the theories and models. These typically connect several concepts with each other and assign measurements to them. Even the basic idea of latent classes with indicator variables can be interpreted as causal relation in many cases, even though the direction of the effect is not always obvious (Section 2.1).

Next, we are going to cover the general principles of (quantitative) measurement, regardless of the instruments or the measured objects or phenomena. This mainly includes the explanatory power of the collected data in general and the mathematical methods that can be used to compare measurements with each other. Finally, these general foundations will be combined in order to describe the methodology for measurement and scaling used in social research. Hence, some important terms in this context will be defined and a basic research workflow will be outlined. (Section 2.2)

More practical examples can be found in the following chapter on related work (3) that provides information about several collections of measurements (3.1) and models (3.2). Here, some important meta-data of scales will be compared amongst the different ways in presenting scales.

2.1 Philosophy of the Social Sciences

Beginning with an overview of the terminology, we introduce the idea of theories and the problems of universal laws. Finally, we close with a quick sight on different standpoints on how to conduct research properly.

2.1.1 Basic Terms and General Concepts of Scientific Work

As mentioned already, an important aspect of scientific work is the way of communicating theories, ideas and findings. We have to focus on some terminological foundations in order to define the necessary terms.

[1] "Da die Hauptaufgabe der empirischen Sozialforschung in der Erklärung von sozialen Ereignissen bzw. Tatsachen liegt, [. . .]" [SHE05, p. 57]

Definitions form a linguistic basis for describing things in the present context. A definition is "an explanation of the meaning of a word or phrase, especially in a dictionary; the act of stating the meanings of words and phrases" [OALD05, "Definition"]. In order to create a definition one can use other terms that already have a certain meaning and their combination leads to this new term (intensional definition). Alternatively, one can list or enumerate all the things that this term describes (extensional definition) [SHE05, p. 52] [ISO 704:2000, pp. 15–17] [2].

These definitions can then be used to describe **objects** that are "defined as anything perceived or conceived. Some objects, concrete objects such as a machine, a diamond, or a river, shall be considered material; other objects shall be considered immaterial or abstract, such as each manifestation of financial planning, gravity, flowability, or a conversion ratio; still others shall be considered purely imagined, for example, a unicorn, a philosopher's stone or a literary character." [ISO 704:2000, p. 2]

This leads to the idea of **concepts**, that are not meant to be single objects themselves but rather an abstraction for a mental class of objects, either real or imagined. Concepts "shall be viewed not only as a unit of thought but also as a unit of knowledge."[ISO 704:2000, p. 2] Most often people talk about concepts and not about a single object: "Concept formation plays a pivotal role in organizing human knowledge because it provides the means for recognizing objects and for grouping them into meaningful units in a particular field." [ISO 704:2000, p. 3]

Concepts use **characteristics** that are part of a definition to describe those mental groups or concepts. "Properties of an object or common to a set of objects are abstracted as characteristics which are combined as a set in the formation of a concept. Characteristics are constantly being combined in order to create concepts, although differently in different cultures, fields or schools of thought." [ISO 704:2000, p. 3]

Concepts can not be treated isolated. They rather have to be seen as appropriate description of objects in a certain context. **Concept Relations** connect concepts with each other to form a **concept system**. We distinguish between hierarchical relations, either generic or partitive, that describe a *is a*-relation between the concepts, and associative relations that relate concepts thematically. "A **generic relation** exists between two concepts when the intension of the subordinate concept includes the intension of the superordinate concept plus at least one additional delimiting characteristic" [ISO 704:2000, p. 6] whereas a "**partitive relation** is said to exist when the superordinate concept represents a whole, while the subordinate concepts represent parts of that whole" [ISO 704:2000, p. 9]. **Associative relations** are the most important ones for the empirical science, since they describe "a thematic connection [...] between concepts by virtue of experience [...] with respect to their proximity in space or time. [...] Some relations involve events in time such as a process dependent on time or sequence; others relate cause and effect." [ISO 704:2000, p. 12]

More details about terminological work can be found in the literature about terminology, international standards or in discipline-specific philosophies. The International Organization for Standardization provides a systematic approach to the standardization of terminological principles and methods. According to [ISO 12620:1999, p. V][3] the following international standards are relevant for terminological work in general:

[2][ISO 704:2000] has already been revised by [ISO 704:2009], equals the German [DIN 2330:1993] that has been revised by [DIN 2330:2013]

[3][ISO 12620:1999] has already been revised by [ISO 12620:2009]

- ISO 704[4] defines naming principles.
- ISO 1087[5] defines the vocabulary of terminology similar to a dictionary.
- ISO 860[6] deals with the international unification of concepts and terms.
- ISO 10241[7] addresses the introduction of standardized terminological entries into other cultural and linguistic environments.
- ISO 12620[8] itself defines data categories for computer applications in terminology in order to support interoperability.

In the standards catalog “01.020: Terminology (principles and coordination)” of the International Classification for Standards (ICS), one can find many related standards, either low- or high-level[9]

2.1.2 Hypotheses, Theories and Laws

In order to describe the assumptions made for explanations and the deductions leading to them, researchers define scientific theories.

According to [SHE05, p. 53–57] a **theory** generally is a system of statements that describes these assumptions and deductions in the form of hypotheses or laws.

Hypotheses are statements that describe a assumed or proposed relation between two variables on the operational level. In the terminological abstraction they form an associative relation between two concepts on the theoretical level. Typically a hypothesis consists of a descriptive part in *natural language* like "‘if A happens, B happens"’ and can be operationalized in any mathematical form. A *quantitative hypothesis* is defined with precise parameters and not only approximated.

Scientific Laws have a similar structure compared to hypotheses, but the stated relationships between concepts are defined to be true. Hypotheses, on the other hand, are proposed relationships. Often they are supported statistically in a single study, whereas they might be rejected in another study.

2.1.3 Covering Laws, Nomological Explanation and Falsification

With the terminology and the structure of theories we try to explain the phenomena of the real world. However, proposed theories and hypotheses that are meant to explain an existing phenomenon or predict a future state, have to be validated at first.

[4]The latest revision is *Terminology work - Principles and methods* [ISO 704:2009].

[5]The latest revision has been divided into *Terminology work - Vocabulary - Part 1: Theory and application* [ISO 1087-1:2000] and *Terminology work - Vocabulary - Part 2: Computer applications* [ISO 1087-2:2000].

[6]The latest revision is *Terminology work - Harmonization of concepts and terms* [ISO 860:2007]

[7]The latest revision has been divided into *Terminological entries in standards - Part 1: General requirements and examples of presentation* [ISO 10241-1:2011] and *Terminological entries in standards - Part 2: Adoption of standardized terminological entries* [ISO 10241-2:2012].

[8]The latest revision is *Terminology and other language and content resources - Specification of data categories and management of a Data Category Registry for language resources* [ISO 12620:2009].

[9]at the time of writing the catalogs could be accessed via the ISO website under `http://www.iso.org/iso/home/store/catalogue_ics/catalogue_ics_browse.htm?ICS1=01&ICS2=020&` (2015-10-09)

A **scientific explanation** is an explanation of a real world phenomenon with logical arguments that are supported with data, possibly retrieved by the use of empirical methods. This might be different from the commonly used term "explanation" that is derived mainly from a certain amount of observations, but is not based upon a theoretical deduction nor a standardized measurement. People tend to "explain" events according to their experienced cause-and-effect relations.

Schnell, Hill, and Esser [SHE05, pp. 57–72] differentiate between three types of explanations, that are mainly based on works by Hempel and Stegmüller: Hempel and Oppenheim [HO48] and their "Studies in the Logic of Explanation", Hempel [Hem68] *Aspects of Scientific Explanation and other Essays in the Philosophy of Science* or also Stegmüller [Ste80] *Neue Wege der Wissenschaftsphilosophie*.

- Deductive-nomological model (DN-model or Hempel's model, Hempel-Oppenheim model or Popper-Hempel model)
- Inductive-statistical model (or also probabilistic explanation or statistical model)
- Incomplete explanations

We will have a more detailed look on each of these in the following subsections.

2.1.3.1 Deductive-Nomological Model

The **deductive-nomological model (DN-model)** is an idealized version of drawing conclusions scientifically. The basic procedure to explain a phenomenon (**Explanandum**) is to find a law that allows to logically derive this phenomenon given the existing boundary conditions (**Explanans**).

This allows to reduce higher laws, to lower laws and (theoretically) to natural laws. By this process even different sciences could be connected with each other in order to explain certain phenomena in another discipline. It is easy to create new hypotheses from existing laws and this enables us to predict phenomena with the help of logical deductions. In that sense, studies are not necessary as a proof, but they can support the deducted statement and identify errors in logical deductions. They may also lead to new ideas for deductions.

However, this approach implies that the conditions can be defined properly: Phenomena can be measured properly and the causal relationships are known. These assumptions do not hold in all cases and this uncertainty makes it difficult to benefit from. The difficulties of measurement and the idea of causality will be discussed later.

Additionally, there is a huge problem with the assumption of universal laws. A scientific law should hold at any given time, in any given society and at any given location, thus ubiquitous. In the social sciences, we normally measure occasional events with a restricted set of participants in defined locations. Even if it was possible to examine all people on earth, laws could only be confirmed for the present time and not verified for the future, where culture, knowledge and behavior might change. This is due to the fact that people can be influenced by many factors that have an impact on their feelings, thoughts and opinions. Compared to natural sciences it is often impossible to repeat studies under the same conditions, because of this property of human systems.

A counterexample would be the research by Ekman and Friesen [EF71] who published their theory of human emotions. They asked members of a preliterate culture in New Guinea to relate facial expressions with given stories, such as the loss of a child. This was the first time in emotions research to find evidence for universal facial expressions [EF71]. Other approaches are large global studies, for example the World Value Survey (WVS) in order to derive universal measures. They try to cover the "full range of global variations, from very poor to very rich countries, in all of the world's major cultural zones." They asked people in "almost 100 countries which contain almost 90 percent of the world's population, using a common questionnaire." [Ins15]

Even if a generalization is not possible, and the social sciences do not have universal laws at all, using explanations based on observed laws or hypotheses can be successful and useful. They may explain the phenomena well enough for decision support in a restricted context and might provide more information than without their presence. Nevertheless they are not derived logically correct from other laws and might change in the future.

2.1.3.2 Inductive-Statistical Model

Since it does not seem to be possible to identify deterministic laws in the social sciences, statistical models take the possibility into account that there might be measurement errors or boundary conditions that can not be controlled in an experiment. Therefore, they assume that a certain law is true for a given probability in the defined context, but it does not have to be true for every object under consideration. Using the boundary conditions for that law as an *explanans*, one could derive a hypothesis as an *explanandum*, which includes a certain probability of objects for which it applies.

The main difference of statistical models, compared to deductive-nomological models, is that statistical laws do not predict the behavior of an individual object. An individual object might be random on the micro-level, whereas the aggregated behavior of several individuals on the macro-level can be expressed and predicted with statistical laws. This makes it more difficult to retrieve such laws and to evaluate hypotheses. Nonetheless, it is used in many different disciplines, such as in thermodynamics, percolation or random-graph-theory [see, for example, Gib02].

However, some philosophers state, that the usage of inductive-statistical models is based on bad measurement tools or a lack of knowledge. Imagine coin flipping where we predict head or tails with equal probabilities. Even if you are unaware of the trajectory, the air resistance or the launch speed, the prediction is right. But what you are looking for is the correct prediction based on the existing parameters and not a chance. This is possible, as we know, but only with the underlying physical knowledge and the right measurements. [Ste74][10]

[10] cited via [SHE05, p. 67]

2.1.3.3 Falsification and Explanatory Power

As it does not seem to be possible in social research to find universal laws[11] , there is another option to create knowledge: Eliminating wrong statements to narrow down correct statements.

For example, it is not possible to empirically prove the statement “it will rain tomorrow or it will not”; therefore we should better use “it will rain tomorrow”, because we can prove that right or wrong. [Pop35, p. 13]

In theory, only one conflicting outcome is sufficient to prove any deductive-nomological hypothesis wrong. At least if it is assumed that the measurement process itself is correct and exactly replicates the real world. As we know, this is an ideal conception.

However, given the possibility of falsification, hypotheses can be judged in a more informative way: The more possible options for falsification they provide, the higher is their informative value, if they have not been disproved. In other words: The more general an hypothesis is formulated, the more valuable it is, as long as it has not been refuted, but can be refuted.

This leads to a set of criteria for hypotheses in order to compare them (according to [SHE05, p. 62]). We will point out the most important ones here:

Scope of application Describes the boundary conditions of a study (ideally without a specific reference to time or location).

Scope of objects Describes objects or individuals with respect to demographic data like *males under 30*.

Universal quantifier States that this statement should be true for all elements in the scope of objects.

Attributes or variables of comparison They are used to compare the individuals with each other.

Inductive-statistical models, on the other hand, are not that easy to reject, because one needs more than one contradicting outcome. Even if the probability of the contradiction speaks against the original hypothesis, it is possible that the sample selection was bad and it does not reflect the scope of objects or even the scope of application. So with inductive-statistical models we can not validate laws and can not falsify them either, we can just argue with probabilities that might be useful in everyday life as reference point. This might not be worthless, but one can not assume that it holds in every context and, therefore, they have to be used with care.

[11] similar to what Popper [Pop35, p. 12] says in German: “Der Schluß von den durch “Erfahrung” verifizierten besonderen Aussagen auf die Theorie ist logisch unzulässig, Theorien sind somit niemals empirisch verifizierbar. Wollen wir den positivistischen Fehler, die naturwissenschaftlich-theoretischen Systeme durch das Abgrenzungskriterium auszuschließen, vermeiden, so müssen wir dieses so wählen, daß auch Sätze, die nicht verifizierbar sind, als empirisch anerkannt werden können.”

2.1.3.4 Incomplete Explanations

In practice many explanations are not deducted from existing laws as described before or are based on any theory. These deviations from the deductive-nomological model are called incomplete explanations according to Hempel [Hem77, p. 124–142][12] and Stegmüller [Ste74, pp. 105–116][13]:

Ad-hoc explanations or hypothetical explanations Events are deducted from laws that have not been approved (yet). However, it is possible to create many "suitable" explanations for any event. These new theories can not be used as an explanans for another event, since they have not been proved themselves (yet).

Partial explanations Partial explanations happen when the explanandum is not precise enough. Thus, most often they are formulated very general. In further deductions, one can deduct many different explanations that are out of scope of the previous explanandum.

Explanation with implicit laws According to Schnell, Hill, and Esser [SHE05, p. 70] this is the most widespread incomplete explanation. "Women vote for christian parties more often than men."[14] This is seen as a causal relation, but there is no proof that sex as a biological attribute is crucial for the voting behavior. Instead, this might relate to ideals or moral concepts. These in turn do not necessarily have to be based on other laws (e.g. women are more likely to be conservative and conservative people tend to vote for christian parties).

2.1.4 Causal Explanations and Causal Models

Besides the covering-law models of explanation (section 2.1.3), causal explanations are very popular in (social) research nowadays. The main idea of causality is that there is some set of factors (causes) that are responsible for a certain phenomenon (explanandum). In short: "To say that x causes y is to say that x and y are related and the relation involves some notion of mechanism, force, production, or the like between x as cause and y as effect." [Bag80, p. 2]

According to Bagozzi [Bag80, pp. 2-3] the detailed analysis of cause goes back to - at least - Aristotle, who apparently introduced four types of causation:

definable form The "shape, pattern, nature, structure of a thing functioning as cause".

antecedent / formal cause "When the marketer notes that awareness is a necessary condition for internalizing an advertisement [...], he or she is referring to a necessary antecedent".

efficient cause The cause which started the process, for example the "influence of a salesperson or the impact of shortages on price".

final cause The purpose of the process like decision making, health, . . .

[12] German translation of extended and revised final chapter of [Hem68]
[13] [Ste74] cited via Schnell, Hill, and Esser [SHE05, p. 69–72]
[14] German: "Frauen wählen häufiger als Männer christliche Parteien."

The efficient cause is probably the most interesting cause, since this is most often the reason why something happened in comparison to a very similar situation where this didn't happen. In causal explanations many factors can lead to a certain effect, but in most cases the whole causal history of one event is not clear. "For instance, the provided explanation might be incorrect, or there might not be enough of it, or it might be stale news" [Lew86, p. 218]. Popper [Pop35, p. 27] introduces the idea of prognoses in the context of causal explanation. In his sense, we can explain an effect in advance when we know the boundary conditions. Why this might not always be accurate enough, will be discussed shortly.

"Information about what the causal history includes may range from the very specific to the very abstract." Assume that a "patient takes opium and straightway falls asleep; the doctor explains that opium has a dormitive virtue. Doubtless the doctor's statement was not as informative as we might have wished, but observe that it is not altogether devoid of explanatory information." [Lew86, p. 220] That means no definitive rule on how detailed a causal history has to be does exist.

There are several discussions about whether causal explanations are part of the DN-model or if that is a separate model. Hempel [Hem77, pp. 20–27] states that they are a part of the DN-model: Some of the causal explanations use existential quantifiers as causes, but they are not proved nor deducted from some law. Therefore, they provide less information and might deliver a wrong explanation. An example would be the "smoking of a cigarette" that might be a cause or an effect for lung cancer, but there is no evidence that it is the cigarette or some ingredient or even some completely different event that is dangerous. Advocates of the causal explanations like Lewis [Lew86, pp. 231–240] state that causal explanations might be part of the D-N-model in general, but "we do not, as least not yet, have a D-N analysis of causation". However, at the moment "often a member of the jointly sufficient set presented in a D-N argument will indeed be one of the causes of the explanandum event. But it may not be."

Counterexamples according to him are:

- Irrelevant non-causes. Think about a man taking anti-baby pills and not getting pregnant [Sal71, p. 34].
- Factors that are not events. Think about a very extrinsic person that behaves in a certain way, because of that.
- Deductions that might be logically correct but not the cause. Think about going to London for holidays, because you want to do a city trip. However, you chose London over Paris, because you have been invited by a friend to visit him.
- Several effects can be sufficient for a cause, in place of the real cause. Example: "A beer ad on my television could only have been caused by a broadcast which would also cause a beer ad to appear on your television. Then the first appearance may be a member of a jointly sufficient set for the second; still these are not cause and effect"
- A cause for that "might be nothing that could have stopped it from causing that effect without itself causing the same effect"

Another famous researcher on causality is Judea Pearl: He was awarded many times for his work on causality and causal models, amongst others with the Turing Award of the ACM

[Gol12]. Furthermore he was awarded the Lakatos Lecture Award in 2001 for achievements in the philosophy of science [Lak02] for his book *Causality: Models, Reasoning, and Inference* [Pea09]. The Lakatos Lecture Award is awarded from the department of Philosophy, Logic & Scientific Method of the London School of Economics, that originally was founded by Karl Popper in 1946. They say that "Pearl presents a unified account of the probabilistic, manipulative, counter-factual and structural approaches to causation, and devises simple mathematical tools for analyzing the relationships between causal connections, statistical associations, actions and observations."

2.1.4.1 Structural Equation Models (SEMs)

Pearl [Pea09, p. 46] uses the idea of structural equation models that is one of the main tools used in social sciences. This model is used to mathematically describe causal relationships. He defines the general form of a functional causal model as shown in formula (2.1). Here pa_i are the *parents* or causes for the effects x_i. u_i represents the error due to omitted factors.

$$x_i = f_i(pa_i, u_i), i = 1, \dots, n \tag{2.1}$$

A (nonlinear and nonparametric) generalization as used in the social sciences is shown in (2.1). In commonly used linear generalizations as shown in (2.2), the pa_is correspond to the $a_{ik}x_k$s that have nonzero coefficients.

$$x_i = \sum_{k \neq i} a_{ik}x_k + u_i, i = 1, \dots, n \tag{2.2}$$

Pearl states that "a set of equations in the form of (2.1) [here] and in which each equation represents an autonomous mechanism is called *structural model*; if each mechanism determines the value of just one distinct variable (called the dependent variable), then the model is called a *structural causal model* or a *causal model* for short."

2.2 Scaling and Measurement Theory

"A major difference between a 'well-developed' science such as physics and some of the less 'well-developed' sciences such as psychology or sociology is the degree to which things are measured." [Rob85, p. 1] However, as there are "well-developed" sciences, there are several fundamentals of proper measurement, that we will discuss in this chapter, regardless of disciplines.

"Conceptually measurement can be defined as a way of assigning symbols to represent the properties of persons, objects, events, or states, in which the symbols have the same relevant relationship to each other as do the things represented." [SA05, p. 352] This assignment can be seen as an "interface" for the real world that allows to conduct calculations on the data. Depending on the characteristics of the data, it is more or less a question of mathematical operations how this data can be altered, compared or combined.

As there is a long history of measurement and, therefore, a huge amount of literature, the interested reader is advised to take a look into the three volumes on "Foundations of

Measurement” by Krantz et al. [Kra+71; Luc+90; Sup+89] that not only defines terminology and axioms for measurement, but also covers the usage of different scale types and the measurement of latent constructs like in psychology. Smith and Albaum [SA05] and Schnell, Hill, and Esser [SHE05] are good references for measurement and scaling in marketing and social research. Additionally, Roberts [Rob85, p. 5] provides a good overview of the literature on measurement in the natural sciences and some predecessors of measurement in the social sciences. Philosophical aspects of measurement are addressed by Torgerson [Tor58] or again by Hempel [Hem52].

2.2.1 Concepts and Definitions of Measurement

The basis of measurement in general are **variables** that contain the data that describes the characteristics and attributes of phenomena, objects or persons. Variables can contain text, numbers, symbols or any representation a measurement requires. Typically the type of the attribute defines the type of a variable: Dichotomous (or binary) variables, for example, represent two states like *attribute present* or *attribute missing*. Non-dichotomous variables can represent an attribute with several possible categories like hair color (black, brown, blond, red) or a continuous value like age (24 years). Furthermore, variables are often divided into **manifest variables**, that can be measured directly and variables that are **latent** (Latin for hidden). A latent variable might for example be the intelligence of a person, that is not directly measurable with existing instruments, but is measured with the help of cognitive tests according to a given operational definition. In social, psychological or marketing research latent variables play an important role, because they represent propositional attitudes, objects of belief and are the primary bearers of truth-value.

The assignment of attribute characteristics to a variable is called operationalization. Measuring the length of an object with a ruler and storing the appropriate length in a variable “length” would be an example. Depending on the measurement instrument, the variables can be assigned automatically or have to be assigned manually. Measurement instruments are tools or utilities that allow to measure characteristics or attributes. A thermometer, for example, measures the temperature. An instrument does not have to be a physical device like a counter, it can also be a questionnaire or even a transcript that can be converted to variables in a predefined manner. The measurement process and variable assignment is defined in a scale, even though it does not have to be the technical coding of the variable itself. Instruments like a questionnaire can consist of several scales, whereas other instruments might only be able to measure a single indicator item.

A **scale** is probably the most important term in measurements and, therefore, it is used in many different ways. However, the basic idea of a scale is the description of a (standardized) measurement process in order to be able to compare the measurement objects or respondents. It is, in fact, an instruction on how to gather the data and how to use this data to compare the measured objects or respondents. First of all, it is important to define what underlying concept should be measured (see section 2.1 for the definition of the theoretical terms). This is called **theoretical definition** and is important to distinguish the scale from others using the same plurivalent name and to give others an understanding of what exactly is to be measured. The process of measuring attributes of a certain object of investigation is called **operationalization**. The **operational definition** describes *the act of measurement* and should be repeatable in other studies and by other researchers. There is typically not only one way to measure a certain concept, so there might be also several scales and several measurements that may be appropriate for a certain task.

We refer to the famous *barometer question* by Calandra [Cal68]. Intended as a bad example for exam questions, it does fit as an example for scale adequacy as well: "Show how it is possible to determine the height of a tall building with the aid of a barometer", was the question for the student. Instead of revealing the conventional answer - measuring the differences in barometer readings on top and bottom of the building - he proposed several alternative ways: use a rope to lower down the barometer to the bottom and measure the length of the rope what is the height; stop the time until the barometer hits the bottom, when thrown from the top, and calculate the height based on time and gravitational pull; measure the length of the barometer and the length of its shadow on a sunny day, then measure the length of the building's shadow and use the proportions to calculate the height. It seems to be clear what the desired way of measurement should be, but all the others are correct, too. The student even proposes to measure the height of the building in "barometer units" where one unit is supposed to be the height of the given barometer, so he doesn't even refer to the International System of Units anymore.

Following that example, it is not only important to precisely describe the way to measure something properly, but also to give instructions on how to calculate the final scores. **Scoring** is the approach on reducing the results of measurement to a single piece of information. Most often this is represented as a number, but it can also be in the form of a symbol, like a smiley. In the given example, the scoring procedure would be to calculate the height out of the difference in barometer readings in order to get a height in the desired unit, e.g. meters.

Typically, a scale is intended to measure different variables that lead to a single score when combined accordingly. Like in an exam, that tests a student on multiple questions and combines all the results in a final grade. This process can include weights for points or a conversion of points to text, like *A+* or *good*. This is especially important in the context of latent variables, as they can not be measured directly. We use other variables, **indicator variables**, in order to retrieve a score for the hidden variable. This approach is called **construct operationalization**. This means as much as the composition (*construction*) of a score with the help of manifest variables. Constructs are not restricted to latent variables, but they can also be concepts that are hard to measure because of response biases (like sexual preferences or drug consumption), high cognitive involvement (how much did you spend on food last month) or the possibility of misunderstandings (does customer satisfaction include product quality, service quality and/or staff friendliness?). An indicator variable can also be a construct itself, even though most scales refrain from relations like this, because it can lead to rather complicated nesting. If necessary, some scales adopt items from another scale, instead.

Constructs often represent partitive concept relations in a theoretical sense. Customer satisfaction might include the satisfaction with the product, with the service and with the shipping. Associative concept relations are often represented in the form of a **model**. The *Technology Acceptance Model*, for example, relates scales like *Perceived Usefulness* and *Perceived Ease of Use* (and others) in order to retrieve the *Actual System Use* of a certain technology. A model is not the result of a scale development process and does not necessarily contain rules that lead to a final score. Actually, it is the representation of a theory on the measurement level. Scales are used as selected measures for the concepts of the theory and they are used to verify the stated hypotheses and laws, usually represented in a graph as directed edges. As models represent a scientific theory, they are intended to provide some explanatory or predictive information that is included in these relations. It may happen, however, that other studies claim to use the same model but identify other concept relations. Models can represent dynamic systems where the causal history might not always be obvious, but it can also be a lack of accuracy in evaluating results and postulating the theory (assuming, of course, that the theory should be valid for the used scopes).

2.2.2 Scaling and Scale Development

Now that we know, what scales are in general, we have a look on the way, how scales are created: “Scaling is the generation of a broadly defined continuum on which measured objects are located” [Pet00, p. 62]. The term scaling or scale development represents the procedure to create a scale, here the “continuum on which measured objects are located”.

According to [Tor58] scaling procedures have to answer three questions of the final scale: measurement properties of the final scale, the task, the subject is asked to perform, when answering the questions, and finally the focus on measuring the subject, the stimuli or both.

Measurement properties are meant to be the primary types or levels of the scale, such as nominal, ordinal, interval or ratio scales. They mainly differ in their mathematical expressiveness. The choice of a particular level limits the possible actions a subject can conduct, namely it limits the number of data collection methods such as ranking or rating answers. Additionally it is important to specify whether the scale measures stimuli, such as the quality of products, or respondents in the sense of comparing their attitudes. Sometimes a scale is even intended to measure both at the same time. In the following every property will be explained in detail.

It is important to know, that the term *scale* is often used as a synonym in each of these contexts. The *primary types of scales* or *levels of scales* are often referred to as *scales* when just talking about their nature. *Ranking scales* or *rating scales* are often used without background information on scale development or scoring, because they seem to apply certain standard methods, like calculating means for scoring. The terms *Likert scale* or *summated scale* are often used instead of *rating scale*, because they are used in the scale development quite often and make use of the standard scoring methods. However, it is not always the case that a *rating scale* has been developed according to the rules for a *Likert scale*, even though the response categories look like a Likert scale. It might also be better to refer to scales developed according to Likert as Likert-style scales in order to prevent misunderstandings.

2.2.2.1 Primary Types of Scales / Levels of Scales

According to Smith and Albaum [SA05, p. 352] there are three important characteristics of real number series, that reflect the real world relations between objects with the help of numbers:

1. Order: Numbers are ordered.

2. Distance: Differences exist between the ordered numbers.

3. Origin: The series has a unique origin indicated by the number zero.

Scale	Information			
	class	order	distance	origin
Nominal	✓	–	–	–
Ordinal	✓	✓	–	–
Interval	✓	✓	✓	–
Ratio	✓	✓	✓	✓

Table 2.1: Primary Scales and their Information Content

Scale	Mathematical Group Structure	Permissible Statistics	Typical Examples
Nominal	Permutation group $y = f(x)$ [$f(x)$ means any one-to-one correspondence]	Mode Contingency Coefficient	Numbering of football players Assignment of type or model numbers to classes
Ordinal	Isotonic group $y = f(x)$ [$f(x)$ means any strictly increasing function]	Median Percentile Sign test; run test	Harness of minerals Quality of leather, lumber, wool, etc. Pleasantness of odors
Interval	General linear group $y = a + bx$ $b > 0$	Mean Average deviation Standard deviation Product-moment correlation t-test F-test	Temperature (Fahrenheit and centigrade) Energy Calendar dates
Ratio	Similarity group $y = cx$ $c > 0$	Geometric Mean Harmonic mean Coefficient of variation	Length, weight, density, resistance Pitch scale Loudness scale

Table 2.2: Primary Scales and their Characteristics, analog to [SA05, p. 358, Table 9.2]

These characteristics can be used to distinguish between the four *primary types of scales*, often also referred as *levels of scales* or *sort of scales*: nominal, ordinal, interval and ratio scales. In the following we will introduce each of these scales with regard to the level of information (table 2.1) and the set of (mathematical) operations that are allowed (table 2.2). Smith and Albaum [SA05] follow among most others the following categorization of scales published by Stevens [Ste46] in Science:

Nominal Scale A nominal scale represents a variable that uses numbers as labels or tags for the properties to be measured. It establishes a one-to-one correspondence, which, in fact, can be seen as a classification. Telephone numbers can be one example of such a correspondence, where one can “reach” a certain person with the help of this number. Another example would be the “classification” of hair colors that assigns “1 - black hair” to people with black hair, “2 - blonde hair” to people with blonde hair and “3 - brown hair” to brown-haired people.

These classifications basically allow to distinguish between objects of several classes and counting of elements in a class is one of the most rudimentary mathematical operations. One could use that for calculating the mode (modal score) as the value that appears most often. Furthermore, it allows cross-tabulation, e.g. in form of a contingency table, that provides a basic picture of the interrelation between two given variables. Based on that, one could also use contingency tests in order to calculate the likelihood that a member of one category is also a member of another category or one could try to predict a missing value based on the (conditional) probabilities of other attributes. [See discrete multivariate analysis, for example, BFH07]

However, this does not allow the usage of the usual statistical operations like calculations of mean or standard deviation.

Ordinal Scale An ordinal scale adds the characteristic or a given order to classes of a nominal scale. If one assigns numbers to people of a competition in an increasing order, where number one is the winner, we can distinguish between several participants. But we are also able to identify who's performance was better than someone else's. We can not, however, get information about the difference in performance or the distance between the ranks. Clothes can be categorized in sizes like S, M or L, where we know that L is larger than S and M, but we usually do not know the exact difference.

These assignments are unique up to a strictly positive transformation, that means we can use a function that preserves the order to modify the values.

For analysis purposes, we can use positional measures like the median, quartiles, percentiles and other summary statistics that deal with order. We can not, however, use the arithmetic means or any form of a weighted index ranking, even though this is sometimes done in a subjective way.

Interval Scale Interval scales are amongst the most used scales, especially in questionnaires. They add to ordinal scales the possibility to specify distances between values. This allows to make statements like "the difference between A and B is twice the difference between B and C" but as there is no common zero, one can not state "A is twice as large as B". A famous example are the scales for temperature, that use an artificial zero point that is not the same in different scales (and also different from the absolute zero on the Kelvin scale). As an example, 10 °C and 20 °C have the same difference in temperature as 20 °C and 30 °C but one can not state that 20 °C is twice as hot (in form of heat energy) as 10 °C.

If you want to transform one scale into another without loosing the order or distance information, you can do that with $y = a + bx; b > 0$, for instance $[°C] = -273,15 + [K]$.

With these scales, most ordinary statistical measures work, this includes arithmetic mean, standard deviations and correlation coefficients. Other statistical tests are often robust enough that assumptions do not have to be strictly met; using those is common practice even though it is not recommended without understanding.

Ratio Scale The most useful scale in terms of possible operations and analyses is the ratio scale that adds the unique zero point. Scales like that are most often found in physical sciences (length, weight). Age or income would be an example of a ratio scale, where all multiplications with positive constants lead to correct transformations.

All types of statistical analyses can be performed and the information is as high as possible.

It is important to keep in mind, that some scales look like ratio scales (as for example the Celsius scale does), but they are not. This especially applies to rating scales in questionnaires, where people often tend to see them as interval scales (how do answers differ in the number of points) and not as ratio scales (point 4 is twice as important as point 2) [Sem96].

2.2.2.2 Data Collection Methods for Questionnaires

In order to be able to collect data about a subject's opinion or attitude, a questionnaire is a typical instrument to be used. Of course, there are many more possible ways to gather data with devices more similar to instruments of the natural sciences, that might include eye-tracking in experiments or sensor data of smart phones that reveal information about activity or locations. Nonetheless, interviews and questionnaires are still heavily used to get explicit knowledge from people. Most existing scale handbooks in the social sciences contain textual questions or items as indicator variables used in questionnaires and so we will cover the most important data collection methods for questionnaires here.

We focus on two primary types of scales, namely with ordinal and ratio measurements. Checkboxes can be used for nominal variables, of course, but they usually provide too little information for multi-item scales. Interval scales use collection methods from ordinal or ratio data collections but there is no real difference in data collection. Think about the Celsius scale again, that could be measured with a number field and looks like a ratio scale, but is indeed an interval scale. For rating scales, respectively, we often assume equal distances between the ordinal categories and, therefore, can transform the answers to interval data with given mean and standard deviation.

Pairwise Comparison A pairwise or paired comparison lets the respondent choose one stimulus over a second one regarding to his preferences. Typically this should be done regarding to one property of interest of which the chosen stimulus “wins against”, “dominates” or “has more of than” the other option.

Variations of this approach can be made in the degree of enforcement to choose a winner, whereas it might also be possible to provide a form of a tie or an option not to answer. Additionally, respondents might be asked to state the intensity of their decision, for example with options to choose from like “totally preferring A over B”, “slightly preferring A over B”.

If there are more than two stimuli, say n stimuli, the total number of comparisons is $n*(n-1)/2$ if the order of the presentation is not important, otherwise it doubles. Hence, the effort for a respondent is growing fast. It would be less effort to let the respondents rank all the stimuli at once, but it is also less informative.

The main difficulty in pairwise comparisons is to detect inconsistencies in the choices, be it because of contradictions or because of other reasons. Typically one would expect a transitive order of the stimuli (A > B > C > D) or groups of similar important stimuli (A > {B,C} > D). However, it could also happen that data is not transitive anymore and it contains circles (A > B > C > A). Answers are likely to change between trials because of underlying

	A	B	C
A	0	1	1
B	0	0	1
C	0	0	0

Table 2.3: Tabular representation of pairwise comparison with a transitive order A > B > C

preference probability distributions and it can be complicated to find out the reasons and severity for the violations of transitivity. There are many approaches to analyze this data, the most famous are Spearman's rho (ρ) rank correlation [Spe04] and Kendall's tau (τ) [Ken38]: Kendall focuses on differences between rankings and not on distances as Spearman does. Thurstone's "A Law of Comparative Judgment." [Thu27][15] is famous for analyzing pairwise comparisons in the psychometric field.

For further processing, results of a pairwise comparison are often represented as a pairwise comparison table, where the value in a cell states how often the stimulus in the row has dominated the stimulus in the column (see table 2.3).

Ranking Procedures Less effort for participants are ranking procedures, where respondents "order stimuli with respect to some designated property of interest" [SA05, p. 375]. It is more practical than pairwise comparison, because it can be done with more stimuli and it takes less time and effort, even though the number of stimuli should not be too high. However, it provides less information than the pairwise comparison and might lead to false conclusions, if there is no real transitive order of the stimuli. It is also difficult to evaluate whether the respondents compared each item with another or mainly focused on the most important ones, for example.

Variations of the ranking include for example partial rankings, where users can choose k out of n stimuli (k/n). Here $1/2$ corresponds to a pairwise comparison and $(n-1)/n$ represents a full ranking. Another option is to let respondents sort items into categories of similar items and rank these categories in a second step, as it reduces the risk of forced differences.

Analysis is possible in the same way as for the pairwise comparisons, either as full rankings or partial rankings.

Rating Techniques Ratings with a defined set of answer options that have a certain order should naturally be treated as ordinal measurements. However, they are often used in interval- or ratio-measurements, when the distance between the answer options is assumed to be equal and there is possibly a real zero.

Respondents "register a degree or an amount of a characteristic or attribute directly on a scale" [SA05, p. 376]. This is used in a variety of scaling approaches like the semantic differential or the Likert summated scale.

Problems may occur when respondents do not understand the answer categories in the same way. Positive/negative worded attributes can lead to end-piling of ratings to the positive/negative end. This might lead to a lack of differentiation. Furthermore, there are

[15] [Thu27] cited via its reprint [Thu94]

many discussions about providing a middle option or not, because that might influence responses as well.

Ratings can be presented with pure numeric response categories (may be viewed as metric), graphical symbols (such as smileys or stars) or verbal descriptions (like “agree strongly - disagree strongly”) or as a mixture. It is important to keep the reference for the ratings in mind, as it may be monadic scaling, where the object is rated independently, or comparative scaling, where it is rated in relation to other objects. There are also many other characteristics to be considered for data collection as the following questions show: >Do you work with negative numbers, is the number of answer categories balanced, do you force answers, do you change positive/negative directions between questions, . . . ?

Direct Judgment A respondent provides a numerical rating to a stimulus with respect to some attribute. That is clearly an interval or ratio measurement, depending on the existence of a real zero point.

Typically one can distinguish between unlimited response categories, where the respondent might insert his age as a numeric response (e.g. for age or income), tick a mark on a line or use a slider. Limited response categories provide a set of fixed values, similar to a rating scale with the particular labels. This is based on a partition of the domain.

Fractionation/Constant Sum Fractionation is an approach to receive a numerical estimate of the ratio between two stimuli with respect to some attribute. If one would set the value of stimulus A to 1, the respondent had to insert the value for stimulus B in relation to stimulus A.

The value for stimulus A can also be empty, so that the respondent has to give a value for both. If one divides both values next, this leads to a normalization of the relation regardless of the actual values. This leads to the idea of the constant sum approach: A respondent distributes a number of points over a number of alternatives in a way that it reflects their relative magnitude of some attitudinal characteristic. The sum of the points can then be standardized across participants to make proportions comparable.

2.2.2.3 Indexes and Scales

The previously described data collection methods produce data that has to be prepared for further usage and transformation. If you take rating data as an example, multiple measurements on different respondents are often averaged in order to be able to compare different groups of people better. Ratings of similar stimuli are averaged often as well, so that one can handle a reduced number of stimuli, for example.

In order to be able to combine results in a certain way, the design of the items and scales is important. Typically items are selected in a scaling process that has a certain focus on specific outcomes, typically reducing the number of actual items without losing their ability to distinguish between specific groups.

These can be divided into techniques for measuring stimuli, subjects or both. Other classifications divide them into comparative and non-comparative scaling techniques. As there are many different approaches, we are going to focus on two that are used quite often in social research: A Semantic Differential Index and Likert Scaling, both non-comparative techniques. Comparative scaling techniques would be the pairwise comparison scale by Krus and Kennedy (1977), the Thurstone scale based on Thurstone's law of comparative judgment or the Guttman Scale.

The terms index and scale are not clearly distinguished according to [SHE05, p. 166]. Both procedures are analysis techniques and they try to combine several indicators to a new variable. They state that scaling methods might be indices themselves, because they add criteria about whether an item belongs to a scale or not. Scales Handbooks, however, often list indexes and scales in the same way.

Semantic Differential The Semantic Differential technique is an approach for measuring and comparing stimuli. It measures the direction and intensity of respondents' attitudes and is often used for further analysis with techniques such as factor analysis. It is often used to determine semantics and meanings of terms, especially adjectives, that describe concepts. According to Schnell, Hill, and Esser [SHE05, pp. 175-177] a semantic differential is an index and is not a scaling method in a narrower sense, but others like Smith and Albaum [SA05, p. 391–395] list it as a "technique for scaling stimuli", even if they state that in some cases an intervening model like this "may not be more elaborate than averaging the raw data across respondents and/or response occasions."

Typically a semantic differential uses bipolar items and a(n) (interval) rating scale and the respondent has to answer where his position lies. For example one could evaluate the test of chocolate where the respondent has to state the intensity of sweetness and bitterness. In general there are several polar adjectives that are used to describe the given concept and they are often divided into three dimensions of meaning: *strength*, *value* and *activity*.

Charles Osgood is a pioneer in this field: In his work "The Nature and Measurement of Meaning." [Osg52] he describes the development of a semantic differential [p. 222ff] with a subsequent factor analysis.

Likert Scaling The goal of Likert scaling is to use several statements to measure one underlying phenomenon. Probands have to state their degree of agreement or disagreement with several statements. Typically this approach uses rating scales and treats the answer options as equidistant, thus interval scaled. The scale should provide a good accuracy of discrimination of subjects with different phenomena or attributes.

Typically the relevant items are extracted from a huge set of "raw" items (typically about 100) that are tested in pre-studies. The selected items should be a small number (typically around eight) with a high discriminatory power [SHE05, pp. 187–191]. The procedure of this item analysis is important for the quality of the scale in the end. Again we want to mention that in practice many rating scales with 5 or 7 items are named Likert scales even if they are not based on an item analysis. They sometimes just represent a simple additive index.

Likert type scales have been proposed by Rensis Likert in 1932 [Lik32] and are the most often used scaling method in the social sciences at the moment.

2.2.3 Quality of Measurement

Even if there are good scales, good instruments or good tools, it is always possible to make mistakes or to forget an important aspect for the study. In the following we will shortly discuss some of the most important sources of variation and three typical quality measures. To which extent they are tested in a qualitative or quantitative manner is up to the people doing that. Nonetheless it is an important aspect in the comparison of scales or studies to be able to test their quality and be able to compare quality characteristics.

In fact we have to answer questions like these:

- Are we able to measure what we want?
- Do our results stay stable over time?
- Do other studies/scales/instruments, that claim to measure the same, lead to similar results? Can we explain the variations?

Objectivity, reliability and validity are the three quality attributes that we describe in the following.

2.2.3.1 Objectivity

Objectivity might be the quality measure that is the most simple one to understand, but probably the hardest to measure. It states the independence of the measurement results from the people who execute the survey. The results should be the same for two different people who carry out the study, they should not imply a bias on respondents or influence them in any way.

This, of course, is not limited to survey conduction, but refers to data analysis, interpretation and presentation. Typically difficulties with objectivity occur, when involved persons are not trained well enough. These could be, for example interviewers who are not able to ask questions neutrally or take responses while staying unconcerned. Another difficulty occurs when results are pushed in a certain direction, willingly or by matter of fact, for example when there is funding from organizations that are interested in certain outcomes. A current case of questionable objectivity would be the disclosure of Coca-Cola that they spent "$119m on health research over five years [...] including funding for a group criticized for downplaying role of sugary drinks in obesity". [Ass15]

2.2.3.2 Reliability

Reliability, saying that you want to rely on the data you gathered, is not only based on given objectivity, but also on the correct functionality of the measurement instruments. Results of a scale or instrument must be consistent for individuals at different times and over groups of individuals. We require formal accuracy in recording characteristics and reproducibility of the measurement results under equal conditions.

Possible sources of error might be changed conditions or a lack of precision of the measurement instrument. However, one has to consider statistical noise as sources of small variations that are not necessarily related to unreliable measurements.

A common measure for reliability is the internal consistency. It evaluates if the used indicator items are producing similar scores as the overall scale does.A frequently used test for internal consistency is Cronbach's alpha (α) [Cro51] that serves as a lower bound estimator for reliability, even though it has its pitfalls [Sch96; CP84].

Besides internal consistency, there are also other types of reliability, such as test-retest reliability. Test-retest reliability evaluates the variations in answers between two items in repeated measurements with the same context. [See additional measures, for example, Tro06]

2.2.3.3 Validity

Validity measures the freedom from systematic error. Results can be objective and highly reliable, but they might not be the results one wanted to measure.

A famous example as the launch of the "New Coke" by Coca-Cola in 1985, where they did a huge effort on generating a new taste for their soft drink. In a blind-test with 190.000 participants they proved that the new formula was way better than the classical one. However the company had to reintroduce the old product again, just about three months later. Customers did not want the *New Coke* to replace the *Classic Coke*. It seems that the measurement of taste was lacking the question if people wanted to replace their old coke. Additionally, it did not predict sales, as they seem not to depend only on the taste. (Validity errors: [SA05, p. 362], story: [Coc12])

Often validity is not represented as a single score, but there are different concepts of validity, that can be measured:

Content Validity Is the attribute that is measured the attribute one wants to measure? In the case of the *New Coke*: Do we want to measure taste or do we want to measure if people would like a new coke? Did we cover all reasons for replacing the old coke?

Criterion Validity Is there any external criterion we can try to explain or predict? Again, in the Coca-Cola example, we assume that we can predict purchases from taste or taste from purchases. However, it seems that taste was a bad predictor for sales. People might link other experiences or feelings with the brand or its taste. These could have an influence on purchase behavior and a stronger effect on sales than taste.

Construct Validity If one uses a construct, meaning other indicator variables predicting one (latent) variable, there might also be some additional characteristics to test:

1. Convergent Validation: When using different methods to explain the same concept - do they get similar results?
2. Discriminant Validation: If groups of participants have differences according to the measured concept, can we identify these different groups? Thus, do the participants make use of the full range of answer options.
3. Nomological Validation: Is the concept explained by theories in the same manner as it might be explained with indicator variables. Do other, casual, bystanders understand the meaning of the concept?

3. Related Work

Besides the foundations in chapter 2, we are now looking at collections of scales, measurements, models or theories. We are going to point out the form of data presentation in books and on websites. This should function as a starting point and help to reuse well-engineered structures for our new application.

Besides the collections of scales, there might be several topics that seem to be related as well. For example, it is interesting how survey tools are modeling their data and how data analysis tools are processing the data. However, we did not find any tools that cover the model perspective in addition to the representation of questions. In general there are some discussions on dba.stackexchange.com about good architectural concepts of survey tools: a database diagram of an (unknown) existing system [Sta12b], different database models with pros and cons discussed [Sta12a]. An interested reader might follow the discussions there.

After covering the different collections of measurements (3.1) and models (3.2), we are going to point out the differences of the proposed system in section 3.3.

3.1 Collections of Measurements

We report on some of the existing approaches of collecting scales, because this is the underlying function of our proposed database. Mainly researchers are using books, often titled handbooks of scale, that list many scales (3.1.1). In the web, there are approaches like wikis that should provide better access to scales for all researchers (3.1.2). Finally, there are some databases that store meta-data of scales in a structured manner (3.1.3).

3.1.1 Traditional Handbooks of Scales

The traditional way for researchers to obtain scales for their questionnaires is to use handbooks of scales. Examples of these are *Handbook of Marketing Scales. Multi-item Measures for Marketing and Consumer Behavior Research* [BNH11], *Marketing Scales Handbook*

[BH13], *Handbook of Personality Psychology* [HJB97] or *The Blackwell Companion to Sociology* [Bla04].

These topic-related collections of measurements, usually scales and indexes, allow readers to compare the scales used in a certain field and most often contain the important works of this topic. These measurements are typically extracted from journals in the period considered in the collection. Depending on the handbook several hundreds of scales in different categories can be found.

The authors try to use a standardized language and structure to describe the features of a scale and they often are experts in the respective field. Furthermore, the publishers have the rights to (re-)publish the content of specific journals or books and can provide an extensive coverage of selected topics [BHJ05, p. VII].

The update frequency of the scales depends on the publication frequency of new editions and the financial possibilities of the readers, since those books can be expensive. Furthermore, some editors decide to remove scales that are outdated or used seldom in order to replace them by more relevant ones. On the one hand, this works as a quality filter for users of these scales and allows them to rely on the most useful scales. On the other hand, a user can only retrieve scales of the licensed journals and can not compare them with older scales in order to identify the general innovations. He relies fully on the qualification of the authors and editors. Those, however, can benefit from the journal reviews, their increasing overview and enable more people to use the selected scales.

Data quality metrics in scales handbooks are often commonly used ones, such as Cronbach's alpha. They are, however, often only extracted from the original publication of the scale and there is no approach of comparing the measures in various studies that used this scale.

Relations between scales (extension of, subset of, similar to) are sometimes mentioned textually with references to other papers, but most often they are not present.

Due to the printed book format, fast changes, data reuse or data extensibility is only possible with huge efforts, that includes manual transcription of items or a deeper reading of the related papers. Therefore, automatic data analysis is only possible in a limited way.

Handbook of Marketing Scales As an example, we are having a look at the *Handbook of Marketing Scales. Multi-item Measures for Marketing and Consumer Behavior Research* that includes more than 150 scales in its third edition: It contains measures from 21 "marketing journals and marketing-related conference proceedings. In addition, the social and applied psychology, management, and organizational behavior literature, as well as several books, contributed measures to this volume." [BNH11, p. xiii]

The editions of this handbook contain carefully described the criteria for scale deletions and additions. Deletions are mainly based on the age of the scale related to the number of citations. Additions need to have "a reasonable theoretical base and/or conceptual definitions", they need "several (i.e., two or more) items or questions", should be developed "within the marketing or consumer behavior literature", have "at least some scaling procedures" applied during development and have existing "estimates of reliability and validity". [BNH11, p. 2].

Additionally, they are providing a clear structure for the presentation of the measures. Those are divided "into six general topical areas (with subtopics), and have devoted a chapter to each topical area)" [BNH11, p. 4].

Each scale presentation follows a clear structure (see [BNH11, p. 4–5]), that includes

Construct : the definition and/or theoretical base of the construct as provided by the authors of the scale

Description “the description of the measure, including the number of items, scale points, scoring procedures, and dimensionality”

Development “how the scale was developed—the general procedures used to derive the final form of the scale from the original scale development article”

Samples “the samples used in scale development and validation”

Validity “estimates of validity (i.e., reliability and convergent, discriminant and nomological validity) from development of the scale; in many cases, actual estimates are provided, however, in articles performing numerous tests of validity, a summary of the pattern of results is offered along with some example findings that provided evidence of validity.”

Scores “mean and/or percentage scores on the scale from the original development article”

Source “the source of the scale—the authors who developed the scale and the publication in which the scale first appeared”

References “critical references from articles pertaining to the topic area other than those of the source of the scale and other sources; these references typically involved description of the construct domain or definition”

Scale items “the actual items in the scale, dimensions to which the items belong, items that require reverse scoring, and where applicable, the directions and scoring procedures for using the scale”

Especially, the construct definition and domain requires a “solid theoretical definition with the construct’s domain thoroughly delineated and outlined” which should also include an “*a priori* dimensionality of the constructs domain”. [BNH11, p. 5] The scale dimensionality should “reflect the hypothesized dimensionality” as “unidimensionality is considered prerequisite to reliability and validity” [BNH11, p. 6].

Bearden, Netemeyer, and Haws [BNH11] focus as well on the validity of the scale with regard to tests, samples and response set biases. Their work has one of the best documented structures.

Marketing Scales Handbook As a second example, we are looking at the *Marketing Scales Handbook* by Bruner and Hensel [BH13] that are also covering marketing measurements. This book is designed to represent a collection of measurements for a certain period of time. Thus a new volume does not replace an old one necessarily.

Bruner and Hensel [BH13] do also presume that scales consist of at least three items, provide a level of empirical evidence and they should be reflective measures rather than formative. However, they try to review rather new scales instead of scales that have lots of uses reported over time.

Here, “the reviews are organized by scale not just by construct” [BH13, p. xiv]. This implies that the authors have defined scale names that should allow readers to identify similar scales or differentiate from scales that contain similar titles but measure different constructs.

The layout of the entries is as follows (see [BH13, p. xv]) and is similar to “the last few volumes”:

Scale Name “A short, descriptive title is assigned to each scale. Several issues are taken into account when assigning a title and the name may not be the one used by the authors. See discussion above for more details.”

Scale Description “A few sentences are used to succinctly describe the construct apparently being assessed and the structure of the measure. The number of items, the number of points on the rating scale, and the response format (e.g., Likert, semantic differential) are typically specified. If significantly different names were used by authors for the measure, they may be noted in this field.”

Scale Origin “Information about the creation of a scale is provided here, if known. Describing the source is complicated by the fact that in a substantial number of cases, the authors who use a scale do not explicitly identify the source. In many if not most of those cases, the scales are original to those authors and they do not provide details about the scale’s development. Another issue that is tactfully addressed in this field is when the source information given by authors is misleading or incorrect, something that occurs far too frequently.”

Reliability “For the most part, reliability is described in terms of internal consistency, most typically with Cronbach’s alpha or construct reliability. In the few cases where it is known, scale stability (test-retest correlation) is reported as well.”

Validity “There are several types of validity and no one study is expected to fully “validate” a scale. While it is hoped that authors of each study would provide at least some evidence of a scale’s validity, the reality has been the opposite. Most studies have not reported much if any information about their respective scales’ validities. At the other extreme, some authors have provided so much information in their articles that it is merely summarized in this field. In those cases, readers are urged to consult the cited articles for more details.”

Comments “This field is used occasionally when something significant was observed and was deemed important to point out that did not “fit” well in the other sections. For example, if something about a scale was judged to be deficient then readers have been urged to exercise caution in using the scale.”

References “Every source cited in a review is referenced in this section. The six journals that were closely examined for articles with scales are Journal of Advertising, Journal of the Academy of Marketing Science, Journal of Consumer Research, Journal of Marketing, Journal of Marketing Research, and Journal of Retailing. Citation of additional journals, books, proceedings, and other sources are provided when relevant. As stated in the Acknowledgments, in many cases the scale users themselves were contacted and provided information that helped with the description. Depending upon the extent of their assistance, they were cited as well.”

Scale Items "The statements, adjectives, or questions composing a scale are listed in this field. Also, an indication of the response format is provided unless it is has been adequately specified in the Scale Description section. For example, if a measure is described as a "Likert-type" then it can be assumed that the extreme verbal anchors for the response scale were strongly agree / strongly disagree or some close variant. Where an item is followed by an (r) it means that the numerical response should be reverse coded when calculating scale scores. Other idiosyncrasies may be noted as well. For example, when slightly different versions of the same scale are discussed in the same review then an indication is given as to which items were used in particular studies."

This means that the core structure of both marketing-related scale handbooks is similar, but not exactly the same. They follow different goals, and contain different scales. At a first sight, however, they do not seem to have many differences. Bruner and Hensel [BH13] also publish their measurements online, so that one can search for them more easily and then buy access to them. For more details on that, see subsection 3.1.3

Depending on the age of the scales handbooks, the presentation might be less well-structured and may contain only some of the attributes shown in these two books.

Measures of Personality and Social Psychological Attitudes Besides from these heavily structured books, there are also publications that focus on "more knowledgeable review essays, but at the expense of less standardized evaluations of individual instruments" [Rob+91, p. 1]. Robinson, Shaver, Wrightsman, and Andrews [Rob+91] follow a different approach in psychology [Rob+91] or political attitudes [RSW99] (both books are structured in the same manner). Experts from each field are "identifying the 10 or 20 most interesting or promising measures in their area, rather than covering all available instruments" [Rob+91].

A review of a scale should follow these evaluative criteria (see [Rob+91, p. 2]):

1. "Item construction criteria (sampling of relevant content, wording of items, and item analysis);"
2. "Response set criteria (controlling the spurious effects of acquiescence and social desirability response sets); and"
3. "Psychometric criteria [representative sampling, presentation of normative data, reliability (both test–retest reliability and internal consistency), and validity (both convergent and discriminant)]."

Most of the actual entries follow a semi-structured form that includes

- a description of the variable,
- a general description of the scale,
- information on the sample,
- information on reliability and validity,
- the primary source, as well as

- some results and comments.

Typically they always include the set of items with the given response domain and information on reverse scoring.

One must mention, that all of these books are generally not containing descriptions of causal models, even though some scales are derived from specific models. They rather focus on measurements of single concepts or constructs, what they call scales, typically. Even though they state to measure scales, the term is chosen as a more general selection of measurements, as they do not exclude indices, for example. Most of the books require a certain scale development approach and they often require unidimensionality of a scale.

3.1.2 Collaborative Semi-Structured Collections of Measures

With the help of the world wide web, the collaborative and distributed collection of scales has become a reality. An example is the Handbook of Management Scales in the form of a wikibook [Wik15a].

Handbook of Management Scales This wikibook mimics the idea of a scale handbook with structured entries in the form of wiki pages, that contain similar information. The huge difference is that every interested person is free to add or edit scales or information on scales. Furthermore, the scales are publicly accessible, even if the related journal articles are not.

The quality of entries is assured in the sense that selected contributors are able to review new content to avoid vandalism or spam. The scales themselves are extracted from journals and should therefore have a certain quality. Nonetheless, there might be transcription errors on a wiki page and depending on the community, they will or will not be detected and fixed. An advantage of the wiki system is that it is possible to hyperlink scales and, therefore, provide a (limited) way of relating scales with each other.

The data is structured in a form that new scales have to be added with the help of a template that defines the basic composition of a wiki page. This allows crawlers to extract a limited form of structured content, for example. However, the amount of structured data is limited, for example when there are different quality measures used. In general one might be able to extract the items of a scale and the full-text of different sections such as the scale description.

The structure of the entries should follow this example [1]:

Scale Name and Validity Measures Example (alpha[1] = 0.679; composite reliability = 0.701; AVE = 0.532)

Description 24 items were developed by a team of three researchers. … Face validity was tested in an interview with four business consultants. … The scale was pretested ….

Definition An example is something that is representative of all such things in a group.

Items

[1] see `https://en.wikibooks.org/wiki/Handbook_of_Management_Scales/Example`

- This is the text for the first item. (0.81)
- This is the text for the second item. (0.75)
- This is the text for the third item. (0.69)

Source `http://dx.doi.org/10.1000/182` Smith/Johnson (2011): This is a title. Journal of Examples, Vol. 1, No. 2, pp. 123-124.

Comments The items might measure "..." rather than "example". New items could be supplemented in order to measure the missing aspect .... The items five and six are formative rather than reflective and may therefore be omitted.

Related Scales

- Performance
- Benevolence

They do also give some additional advice for contributors, e.g.: Scales should have at least three items and should be published in high-ranked journals, DOIs should be used, definitions should be transcribed from the article, ...

3.1.3 Measurement Databases

Measurement Databases are a source of structured information about measurements. They usually provide a preset set of attributes and let users enter the information accordingly. There are many different facets of services, some intended to be open and collaborative, some closed and commercial.

Registry of Scales and Measures The registry of scales and measures [RegSc] does not provide complete scales itself, but they focus on important meta-data to find scales for specific domains, especially in psychology. They include a short overview of purpose and description and provide some scale characteristics such as target groups or number of items. Additionally, they inform about the availability in the public domain and then also provide the associated pdf. The registry has been developed as a follow-up of Santor's [San06] recommendation: "The American Psychological Association (APA) has estimated recently that some 20,000 new measures appear every year, many of which may only have been used once (APA, 1993). Maximizing the benefit obtained from those measures currently in existence and minimizing the possibility of creating additional measures unnecessarily will depend on the efficient exchange of knowledge among all stakeholders. To this end, the existing database of depression measures will be maintained and updated online at www.scalesandmeasures.net."

The form for adding new scales (see `www.scalesandmeasures.net/registerscales.php`) is divided into 6 parts (Title, Purpose and Description, Scales Characteristics, Item Domains, Additional Secondary Sources and Comments) that contain the following fields for attributes:

Title Title, Year, Primary Author, Primary Source, Primary Domain (Functioning, Intelligence, Mental Health and Illness or Personality), Sub Domains (depends on primary domain, e.g. for Personality: Attachment, Interpersonal Sensitivity, Loneliness, Mood, Multifactorial, Social Desirability), Key words, Original Language (ca. 120 languages), Availability (Public Domain, Proprietary or Permission Required)

Purpose and Description Purpose, Description, Definitions, Scoring, Validity, Reliability, Historical Comments

Scales Characteristics Age Group (Child, Adolescent, Adult or Elderly), Target Group (Clinical or Community), Purpose (Severity, Diagnosis or Screening), Number of Items, Number of Options, Number of Factors (1-10), Response Anchors (Likert, Graded Response or Yes/No), Response Focus (Frequency, Duration and Severity or Amount), Cut off Scores (Yes/No), Cut off Score Value, Time Frame (Past Week, Past 2 Weeks, Past Month, Current, Remote Past or Future)

Item Domains Domains and their total Items ('Number of Items used to assess this Item Domain" with a field for entering a number for each of the previously selected subdomains)

Various Additional Secondary Sources, Comments

As mentioned above, the registry focuses on the meta data of the scales. People who add scales should hand in a reference for verification. However, on the website there is only a link to some PDF documents and the single items of the domains are not listed.

Additional Databases There are also many databases that are not open to the public or commercial in general, so that we were not able to review them in detail.

Some of them are described here in short:

- *PsycTESTS* of the American Psychological Association [PsycTests] that includes many PDFs on different tests.
- *Test Collection* at ETS (Educational Testing Service) that allows to buy tests that are linked to articles. [TestLink] They offer "more than 25,000 tests and other measurement devices" and call themselves the " largest compilation of such materials in the world."
- *Test Review Online* of Buros [Buros] provides "3,500 commercially available tests, most of which have been critically reviewed by the Buros Center for Testing." The tests are also published in the Mental Measurements Yearbook [2]
- *Marketing Scales* [Bru15] is the database related to the scales handbooks for Bruner et al. (e.g. [BH13]). Therefore we can assume that the scales look similar, but there is no example online. They state to provide more than 3,400+ scales, some of them printed in the different (e)books, assigned to 12 categories (personality, values, attitudes, perception, behavioral, knowledge, advertising, satisfaction, quality, social, task, emotions)

Those websites might provide additional data compared to the ones that are (partially) open to the public, but this might also be restricted due to copyright issues or only for (paying) members.

[2] The latest Mental Measurements Yearbook is [CGJ14]

3.2 Collections of Theories

We mentioned that most of the collections of measurements (3.1) are not presenting any theories or models. They restrict themselves to publish measures or even only scales. The documentation of theories or models might be more variable, because theories do not necessarily follow a certain structure. As one can see in the following subsections, many theories are presented in the form of reviews in full-text. Many theories are following philosophical considerations that are not necessarily easy to describe mathematically or graphically. Counterexamples would be causal models or structured equation models (SEMs). It might also be the case that older theories are not designed to be represented in causal models or SEMs, whereas today many researchers would think about that.

We categorized the collections of theories similar to the collections of measurements (3.1): Firstly, we are focusing on written books (3.2.1). Secondly, we are talking about collaborative semi-structured collections (3.2.2). Finally, we are presenting structured database that store theories and models (3.2.3).

3.2.1 Books about Theories

There are many books that cover certain research fields and present the theories of this field. However, they are often providing an historical overview and describe the development and important contributors for the different theories.

Theories of Personality For example, Hall and Lindzey [HL70] provide with *Theories of Personality* an overview of original and related research on many famous personality theories. They are not providing information on single studies or even measurements, but they try to include the most relevant results.

They are providing "general categories in terms of which the theories could be described while permitting [...] a good deal of latitude within these categories so as to present each theory in the manner that seemed most natural." [HL70, p. x–xi]:

> Typically, however, each theory is introduced with an Orientation section which recounts briefly the personal history of the theories, outlines the main lines of influence upon the theory, and provides a summary of the salient features of the theory. Next the reader will find a section on the Structure of Personality in which are included the concepts designed to represent the acquisitions or enduring portions of personality. Following this is a section on Dynamics of Personality which sets forth the motivational or dispositional concepts and principles espoused by the theories. Then comes a section on Development of Personality which deals with growth and change as represented by the theory. A section on Characteristic Research and Research Methods follows, in which representative investigations and empirical techniques are presented. There is a concluding section entitled Current Status and Evaluation which outlines briefly the present state of the theory and summarizes the major contributions of the theory as well as the chief criticisms it has elicited. At the end of each chapter is a brief list of Primary Sources which represents the most important of the original sources concerning

> the theory. All of the publications referred to in the text are brought together in a final section at the end of each chapter entitled References."

This provides a large theoretical and scientific overview of the research field. It is, however, difficult to compare several theories with each other and identify similarities or differences from the survey of Hall and Lindzey alone.

Other books do not follow a given structure as well, but they are often structured similarly. Examples would be *Handbook of Personality Psychology* [HJB97], *Positive Psychological Assessment* [LS03], *States of Consciousness* [Kok07], *International Handbook of Personal Construct Psychology* [Fra03] or *Circumplex Models of Personality and Emotions.* [PC97].

3.2.2 Collaborative Semi-Structured Collections of Theories

Similar to the usage of collaborative software to create an open collection of scales (see 3.1.2), there are attempts to collect information on theories as well.

Theories Used in IS Research Wiki For example, Larsen, Allen, Vance, and Eargle [Lar+15] are running the *Theories Used in IS Research Wiki* that covers many theories in information systems research. They focus on "details about the theory, some examples of IS papers using the theory, and links to related sites" [Lar+15].

In order to add a new theory, they provide a template as well. However, this template does only contain the main headings and the content is up to the creator. There is no detailed description of the necessary contents at hand. The template looks like this[3]:

- Theory name, acronym, alternate name(s)
- Main dependent construct(s)/factor(s), main independent construct(s)/factor(s)
- Concise description of theory, diagram/schematic of theory
- Originating author(s), seminal articles, originating area
- Level of analysis
- Links to WWW sites describing theory, links from this theory to other theories, IS articles that use the theory
- Contributor(s), date last updated

Given the context of information systems, we can also have a look at the TAM model here: `http://is.theorizeit.org/wiki/Technology_acceptance_model`. The description of the theory/model itself is rather short, since it lists the used concepts and provides the causal model diagram. It is, however, useful that the TAM model has been assigned to the originating area of *Information Systems* and *Technology Adoption.* The level of analysis is defined as *individual.* In the wiki there are links to other theories as well, so that one can easily find related work.

[3]see `http://is.theorizeit.org/wiki/Theory_template`

The strength of this wiki, however, is the amount of related literature: It does not only list a single main source, but a list of seminal sources. These seminal sources can represent original sources but might also contain important comparisons with other theories in the field. Additionally, they list publications that use the presented theory or model (e.g. 66 sources for the TAM model).

The free form of the wiki provides the possibility to describe various theories and models regardless of their structure. However, it is rather complicated to do automatic analyses with this data, since it is not always semantically structured.

3.2.3 Model Databases

Databases on the other hand, are postulating a certain structure and force contributors to add data according to this structure. This allows to reuse the data automatically.

Inter-Nomological Network Here, the Inter-Nomological Network (INN) [Lar15], is probably the most similar approach to the proposed idea. It “is a tool designed to integrate the behavioral sciences by removing barriers that currently exist within and between disciplines.“ One of the maintainers is Kai Larsen, who describes his research as follows [Lar10]:

> Dr. Larsen, Director of the federally supported Human Behavior Project, is conducting research to create a transdisciplinary “backbone” for theoretical research that predicts all aspects of human behaviors. Whereas a discipline such as chemistry is served by an integrating framework, the periodic table of elements, the behavioral sciences has no touchstone. Dr. Larsen’s research uses automatic text mining technologies to create an integrating framework for predictors of human behavior. The research has implications for our understanding of all human behaviors, including technology utilization, investor decisions, cancer prevention behaviors, and high school dropout rates

The INN contains information on (model) variables as well as on items, that are used to measure these variables. The “main database contains all articles and variables extracted whereas other databases provide discipline or topic focused subsets. Selecting a subset database may increase the accuracy of the similarity search and narrow keyword searches. Our research assistants work tirelessly to add articles to the database, and we update the database quarterly” [Lar15, see the “Database Information Page”]. They provide databases for different disciplines and journals and a single database with the whole set of entries.

There is no full description of the data structure. However, one can extract some information from search results. Details of variables (that seem to be concepts of models) and papers:

- Variable Details

  Source Contains a link to paper details with the citation of the publication as label

  Variable Name of the variable (that includes its origin), e.g. “Usefulness [Original Technology Acceptance Model]”, but links to the variable again

Definition Short definition of the variable. Here "Users' perceptions of the usefulness of the new system."

Item(s) Items of the variable. These are formulated in a general form with placeholders where one might add a certain context, e.g. "Using (system) as a (technology type) enables me to (accomplish tasks) more quickly."

Citations Additional literature that cites this variable. Not necessarily linked in the database, only as pure textual citation.

- Paper Details

Source The source citation of the paper.

Variable(s) The variables contained in the citation. All linked to detail pages in the database, sometimes named with reference to the "original" model.

Citations Citations as for the variable.

In fact, the database itself is very sparse and does not provide a rich amount of content to work with. Despite the lack of content in this database, they are describing their ideas in various articles [e.g. Lar+10; LH12; LL13] and might publish more in the future. The focus here lies on the power of the search engine and the work with natural language processing, synonym detection and providing the most relevant results to the users. However, as a user, the features are not clear enough and one does not necessarily know whether the results are enhanced with artificial intelligence or not.

3.3 Summary and Implications for the Proposed System

As we have seen in the different collections of measurements (3.1) and models (3.2), the scales handbooks and measurement databases provide some mature forms and structures. Depending on the objectives, they provide a lot of meta-information that classifies the scale or they provide information on scale development and design. Especially community-edited wikis seem to be able to refer to a lot of literature that uses or describes these scales (or models). Collections of theories and models are not that well-developed till now, however. We did not find any collection that includes an extensive information on scales and on models at the same time. It seems like the underlying idea of the Inter-Nomological Network [Lar15] is the most similar to our approach, but they do not provide a lot of information on their data handling. The information in the database connects items and variables but is rather sparse with regard to other attributes of scales and models..

The proposed system should provide ideally a common database for models and measurements at the same time. It should allow users to enter the data of both dimensions and relate them with each other. Especially concepts of causal models should be linked with each other and this should be visualized for the users as well.

In comparison to the discussed handbooks, a database provides faster access and better search capabilities. It is also easier to update information and add new information. It is possible to directly link to relevant literature and to increase the work efficiency of users.

Even without the help of algorithms from machine learning or recommender engines, we want to provide a platform that allows users to enrich the existing data with their knowledge, such as links between models or scales.

There is no existing database that provides the data of models and measurements in a structured and machine-readable way, so that researchers are able to conduct studies on this information. We are especially focusing on the possibility to detect similar models and measurements in order to link them with each other. It is important to apply new techniques from information technology in order to gain insights into the existing models. This work does not include any of these algorithms, but it aims to provide data for future projects.

Furthermore, we are trying to figure out additional uses for structured data in the context of different disciplines. The export of measurements for use in a survey tool, would increase productivity and reduce errors.

In this sense, we are proposing a data hub for survey-related data, that goes beyond the existing technologies. It is important to provide a user-friendly interface and a good basic structure that can be used to build upon.

At this time, however, we do not focus on storing survey results as such. Therefore, we also decided not to support validity measures, evidence or samples, in a structured form. It should be possible to add these metrics in a free text form similar to the text blocks of survey handbooks, but not in a semantically annotated system. The support of survey results is another difficult undertaking and is out of range for this thesis. Nonetheless, it would be more than useful to add this in the future, so that quality metrics could be calculated automatically and recurring studies could be compared to previous ones. There will be no way around that to improve research quality and to conduct meta-analyses in a more efficient way.

4. System Design

After the foundations on the philosophy of the social sciences (2.1) and the foundations of scaling and measurement (2.2), we had a look at related work on storing models, scales and surveys (3).

In this chapter we are going to identify the requirements for the proposed system (4.1) with usage scenarios, use cases and restrictions of the system. The software architecture and data models are then described in section 4.2. This is followed by the proposed user interface (4.3) that allows the scientists to interact with the application. Finally, we are going to provide an insight into the actual technical realization of the system (section 4.4).

4.1 Requirements

The proposed system is an application that serves the needs of scientists in a market research process. Its main purpose is the storage of causal models and the conceptual and operational definitions of the included concepts. The measurement instruments are assumed to be questionnaires where the questions represent the indicator variables of an operational definition. However, it might also be applicable to enter models or concepts that use other indicator variables even if they can not be represented completely.

The scientists benefit from the structured representation of these models in various stages of their workflow:

- They can extract and store information from publications with the guidance of the provided structure.
- They can have a look at existing information without the need of reading a full publication.
- They can select models or measurements for use in their own studies and experiments - for the creation of new questionnaires and for the analysis of results.
- The scientists might also use this system for meta-analysis in order to get an overview of a certain field or to develop new models and measurements.

4.1.1 Scenarios

In the motivation (Chapter 1) we mentioned several possible uses and users of the proposed system. In this section we are going to present two scenarios in greater detail and extract the typical actions a user is going to execute.

These scenarios are described more detailed in the following subsections:

Meta-Analysis of Different Models A scientist compares the models of several articles (for simplicity here TAM and TAM2) with each other. At first he is guided through the data entry process and then he can use the data to create his comparison. (See 4.1.1.1)

Creating a Questionnaire from a new Model Researchers are going to prepare a questionnaire for measuring the needs of teachers regarding new computers in education. They look for adequate models and indicator items that have been proposed and tested by others and want to adapt one of these models. (See 4.1.1.2)

4.1.1.1 Meta-Analysis of Different Models

The first user scenario assumes that a scientist uses the database as a supportive tool in order to extract relevant information from different papers. In the end he compares the information of the given papers and writes a review about them. For simplicity, we assume that it is a student who compares the TAM model with the TAM 2 model for a term paper. In fact, this could lead to a field survey similar to the work of Seiffer [Sei13].

We assume that he enters information on the Technology Acceptance Model (TAM) by Davis [Dav89] in the data entry process:

Step 1: Entering Publication Information Since all information in the database is identified by its source, he chooses to add a "publication"-source at first. With the help of BibTeX/BibLaTeX document types (See [Leh+15, p.7–13]), he enters the bibliographic information of his article "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models" by Davis et al., published 1989. The system checks if the article has been added before, but here this is not the case. The student adds additional information, such as the the journal with volume and page numbers.

He might have also uploaded a BibTeX-file that already contains the data or he might have done a lookup in an external bibliographic database with a unique identifier like the doi. Depending on the source, he might upload the source file in the end or he might add references to online resources.

He also adds bibliographic references of the article as links to other publications, so that he is able to identify the foundations of the paper, such as other models or similar research, later.

Step 2: Extracting Theories and Models In the next step, the student is being asked which information he is going extract from the paper. He is able to choose amongst "data about a study", "theoretical models" and "single measurements".

He chooses to add the two models of the article at hand, the *Theory of Reasoned Action (TRA)* and the proposed TAM-model.

As there are several models in his source, he provides an additional qualifier (e.g. page number, name, ...) to refer to their location in the source. In this case, he only adds the name of the models and their abbreviations. He includes a free text description of the model (e.g. objectives, context, ...) as well.

He also states that the TAM-model is defined in his source, whereas the source is only mentioning the TRA-model as a reference. The TRA-model has been originally published by Fishbein and Ajzen [FA75] and Ajzen and Fishbein [AF80].

Step 2.1: Adding Concepts For every model, the student adds the concepts of the model. Here, the TAM consists of *Perceived Usefulness*, *Perceived Ease of Use*, *Attitude Toward Using*, *Behavioral Intention to Use)* and *Actual System Use*. They also include *External Variables*.

Each concept is described by a concept definition: Behavioral intention is "the degree to which a person has formulated conscious plans to perform or not perform some specified future behavior".

Step 2.2: Linking Concepts In the TAM model, the concept *Behavioral Intention* is causally dependent from *Perceived Usefulness* and *Attitude Toward Using*. The student adds the cause-and-effect relations of all the given concepts as directed edges.

He is able to use a regression formula with the names or abbreviations for the concepts, like $BI = A + U$. Alternatively he adds all the relations in tabular manner, where the first column represents the node of the cause and the second column represents the node of the effect. Thus, every row represents a partial causal relation. In other models he would add (empirically derived) weights on the edges, as well.

Step 2.3: Adding Modifiers Some concepts might have an influence on a causal relation but are not necessarily a cause for another concept. They are called modifiers and are often demographic variables. In this paper, there are no modifiers, so the student can skip this step.

Step 3: Describing Measurements Some concepts are measured directly or with indicator items. Thus, the student describes the measurement details, adds the variables with their item text and the measurement domain.

In this case, the paper does not include items or scale information. They only state that they used four items for *Attitude Toward Using*, for example. However, the PhD thesis of [Dav85] contains the measurement descriptions and so we use these in the context of this example. In reality, the student skips the rest of this step.

With the items of [Dav85], the student chooses *Semantic Differential Rating Scale* as a type. ("Attitude toward using was measured using standard 7-point semantic differential rating scales as suggested by Ajzen & Fishbein (1980)" [Dav85, p. 93]). He selects that this measurement is also representing an uni-modal construct. Multi-modal constructs would allow to add multiple layers of domains or facets of the indicator items.

Step 3.1: Adding Item Group Data Now, the student adds the details of the indicator domain. Here, the system proposes a standard differential rating scale and he changes that to 7-point and adds labels. In this case, the central answer option is labeled “Neutral”. As this is a semantic differential scale, the outer answer options are typically labeled with the adjectives defined as indicator items.

He also adds an introductory text, like a scenario description. Davis, for example, lets the user rate the statement “All things considered, my using electronic mail in my job is good/bad.” with different adjectives replacing good/bad. He chooses “All things considered, my using electronic mail in my job is ...” as a generic question text. [Dav85, p. 93]

Step 3.2: Adding Indicator Variables The student adds the item text of the measurements. For example, *Attitude Toward Using* is measured by the following items:

- All things considered, my using electronic mail in my job is good/bad.
- All things considered, my using electronic mail in my job is wise/foolish.
- All things considered, my using electronic mail in my job is favorable/unfavorable.
- All things considered, my using electronic mail in my job is beneficial/harmful.
- All things considered, my using electronic mail in my job is positive/negative.

The student only needs to add the adjectives as answer labels here. The question text is only added for clarification. However, the student could indicate reverse scored items or he could even change the response domain for a single item.

The item text is not standardized yet, because it targets the “electronic mail” system that Davis is evaluating. “The system” may be used as a placeholder for the use case “electronic mail”.

Step 4: Adding Relations to Existing Data After the student has finished the entry process of the model data, he adds relations to other scales.

In this case, he points out that the TRA-model is a predecessor of the TAM-model. This is regardless of the fact whether he added it in the previous steps or it has been in the database before by someone else.

Nonetheless, he already added implicitly a relation of the TRA-model in the beginning: The paper by Davis describes the model but only as secondary source. Thus, it should be identical to the one in the original publication by Fishbein and Ajzen. However, there may be differences in reality, for example due to unintended transcription-mistakes or due to small changes to use it in another context. The generated relation states that both models “should be identical”.

After entering the information of all the papers, the student wants to use the structured data of the database to compare the different models. Here, we assume that he compares the TAM model [Dav89] with the TAM 2 model [VD00].

Step 1: Viewing the Models side-by-side With the help of generated model diagrams, the student compares the two models graphically. He realizes that the *External Variables* of the first TAM model have been replaced with five other concepts. Two modifiers have been added in TAM 2 and the *Attitude Toward Using* is missing in TAM 2.

Step 2: Comparing Measurements Next, the student compares the indicator items with each other and identifies differences. He evaluates the domains of the different measurements, the assignment of variables to constructs or the labels of the indicators.

Step 3: Looking at Related Work The student benefits from existing data in the database. He navigates to the predecessors of the models or he looks for models that could be related. For example, he finds out that Venkatesh and Davis also published the UTAUT model in “User Acceptance of Information Technology: Toward a Unified View” [Ven+03], that has a huge influence in the field of technology acceptance as well.

Step 4: Exporting for Publication Now, the student exports the generated model diagrams as image files and adds them to his LaTeX-document. Additionally, he exports the indicator items as lists or tables to add them to his appendix for reference. Finally, he uses the meta-data comparison to describe the foundations and predecessors of the publications.

Step 5: Citing Original Papers Of course, there might be additional interesting content in the paper, that did not fit in the database. The student adds some examples from the TAM papers to his work. He also refers to the evaluation of the authors themselves. Sometimes it is also interesting in what context the scientists have worked, how their social relations have been to their main “competitors” or which journal published their work. This information might not be stored in the database, so the literature work is still important.

Step 6: Retaining the Findings A written paper sums up the findings and ideas of the student. He not only documents his results in this paper (in prose), but also adds some (structured) links to the database for future use. For example, he states that TAM 2 is the successor of TAM 1, but that UTAUT is a similar model to both.

In the end, the student can hand in a high quality survey, of which other researchers benefit as well. He did have to look for the specific information in the paper, but he was guided through the single steps by the system. Furthermore, he benefited from the reuse of existing information for his work and extended the database with his data and his findings. Future research will benefit from this additional information thanks to his effort.

4.1.1.2 Creating a Questionnaire from a new Model

The second user scenario focuses on a more practical task that is supported by the proposed database. In “Determinants of Computer Usage Among Educators: A Comparison Between the UTAUT and TAM Models” Ling, Downe, Ahmad, and Lai [Lin+11] want to to find out the “actual determinants of computer usage among educators” in their country Malaysia. The country wants to introduce information technology in schools and they are trying to figure out the professional needs of the educators.

Since the paper has been published in the proceedings of a postgraduate conference, it shows the level of knowledge that a young scientist or a scientist from a different field might have. In the following we are pointing out where proper infrastructure can assist. Nonetheless, the general workflow is similar to the way experienced researchers work as well, so that they benefit, too. We are assuming that the researchers from our example did not have to enter data to the database, as this has been described in the first user scenario 4.1.1.1. They can rely on the existing knowledge in the database.

The research problem and objectives have been described as following: "In this study, the researcher aims at: [1] reviewing user acceptance of technology literature and comparing the TAM and UTAUT models and [2] presenting an extension to the UTAUT model to encompass '*Perceived Needs*' as another determinant of usage intention" [Lin+11, p. 2].

Thus, first a literature review in the field of technology acceptance is necessary:

Step 1: Searching for Technology Acceptance Models The researchers start with a traditional literature overview in order to find journals, handbooks or articles regarding technology acceptance. Given the fact that they are aiming at using (or possibly extending) a measurement model, it is helpful to look for models in the proposed database. Looking for search terms like "technology acceptance", "technology needs" or "computer requirements in education", leads them to several models or concepts that could be interesting.

Step 2: Selection of Matching Models With the help of existing meta-data, they refine and filter their results according to certain disciplines (e.g. education or information systems). They also have a look at the indicator items of the models and whether they might fit to their educational context or not.

Step 3: Comparing Selected Models Now, they compare the selected models and find out their strengths and weaknesses. They read the original literature of the found models and they look for studies that have been conducted with these models in their field of interest. UTAUT, for example, has been used in educational contexts amongst others. On the other hand, TAM and UTAUT have been developed by the same researchers, so they might include the same underlying ideas.

For "part one" in their paper, Ling et al. did a literature research and found out that the "TAM and UTAUT models [...] are the two most influential and robust models used in this millennium" [Lin+11, p. 2]. They also mentioned two additional models but did not describe them in greater detail. They explain the TAM and UTAUT model, point out the key concepts as well as the modifiers and provide a comparison table for both.

If they had used the proposed database, the search results would have proposed not only the TAM and UTAUT models that focus on the organization level (as they do according to [Lin+11, Table 2]) but presumably also models from other disciplines. For example, the former software usability model [ISO 9126-4:2001] or its latest successor [ISO 25000:2014] address usability in the context of software and system engineering. However, they are barely found when searching for "technology acceptance" on common search engines. Models from the field of marketing might be interesting as well, as the use of information technology should be "promoted" in educational institutions based on the needs of the educators.

Three of the six pages in the paper are describing the background of UTAUT and TAM and the comparison of the models. The researchers could have transcribed most of this content directly from the database.

Ling, Downe, Ahmad, and Lai [Lin+11] modify the chosen UTAUT model and add a new concept named *Perceived Needs toward using Technology.* They describe the reasons for this addition, but they do not conduct another literature review in order to find similar research on Perceived Needs. They could have evaluated if there is (a) an existing approach that combines technology acceptance and perceived needs or (b) an existing theory or model that is evaluating perceived needs in another context.

Maybe it was too much effort to conduct some further research or they had time constraints. With the help of the database they could have done such a research more easily and they might have found additional models, such as the "Task-Technology Fit (TTF)" model [GT95], for example. They might have just found this model in the "linked theories/models"-section of the TAM model. They might also have searched for "perceived needs" or a specific question and might have found other concepts that are named like this or are described accordingly. Either way they did not mention any other theory in their publication and they used some motivator factors and hygiene factors according to Herzberg's Motivation-Hygiene Theory [HMS59] without reference.

Unfortunately, they did not publish the newly created indicator items for the *Perceived Needs*. Even those could have been used to search for similar items used in other models, in order to find out more about related research. Using this system, they could have added their new model and they could have compared that to existing ones as well. Other users might use this model later on as well.

Using the model and measurement database, they would have been able to conduct surveys for several purposes more easily. See how researchers could have done this in the following steps:

Step 1: Composing Indicator Items In order to select useful indicator items for the proposed *Perceived Needs* construct, the researchers select some models or measurements that might be similar to their construct. For example, they are named "perceived needs" as well. Reusing an existing measurement instrument leads to better quality, since there usually have been some studies that evaluated this measurement instrument and additional studies lead to a broader use and more useful material. They built a questionnaire with the UTAUT model and add some additional measurements for *Perceived Needs*. They also create their new scale by hand, which typically needs more effort, in order to compare the results with existing constructs and scales.

Step 2: Analyzing Survey Results After conducting a survey with these indicator items, they analyze the results. For example, they use a form of exploratory analysis on the data or they confirm certain constructs with a confirmatory factor analysis. Therefore, it is necessary to replicate the causal model relations in the analysis. With the help of the database, they export the structure of constructs and items to their analysis framework, for example the statistical programming language R. Next, they conduct various analyses right away.

Step 3: Documenting the new Measurement Instrument With the help of their new study, they are able to document the new measurement instrument and the relevant factors. Furthermore, they can state certain quality criteria of the construct, such as reliability or validity. Other researchers need to know the context of the study in order to verify the results. They can add the new model and measurement instrument to the database in addition and add links to its predecessors or related models and measurements.

Step 4: Optionally further Studies Finally, they conduct additional studies in order to combine the newly created models with existing ones. This can be used to justify their modifications and to establish a new high-quality model.

As Ling et al. want to solve a practical problem, they should conduct additional studies with this instrument in order to derive knowledge for decision support for the Malaysian Government. Assuming that their model does not only fit for Malaysia but for other countries as well, they could easily reuse this model and conduct comparative studies on an international level.

Unfortunately Ling, Downe, Ahmad, and Lai [Lin+11, p. 4] only mention that "based on the data from the study, the UTAUT model is modified to include the new key construct: *Perceived Needs toward Using Technology*". They did not mention any survey. It is difficult, if not impossible, to use this paper for new research as it seems incomplete. The goal of the authors, however, might be interesting for others as well, but the outcome is not good enough for this purpose.

To sum up this user scenario, it would have been more efficient to use the existing information in the database to quickly prepare a simple paper like this. With basic navigational patterns (e.g. linked models) and a search engine, this work could have been of much higher quality. Maybe the researchers would have been able to conduct a proper study to underline their ideas and it could have been a good addition to research in that field.

4.1.2 Use Cases

This chapter describes the different use cases of the proposed system and the main roles a user can assume. In the form of UML use cases, the different use cases are outlined as diagrams that are described in the text. The elements for the diagrams in this section are described as a legend in figure 4.1.

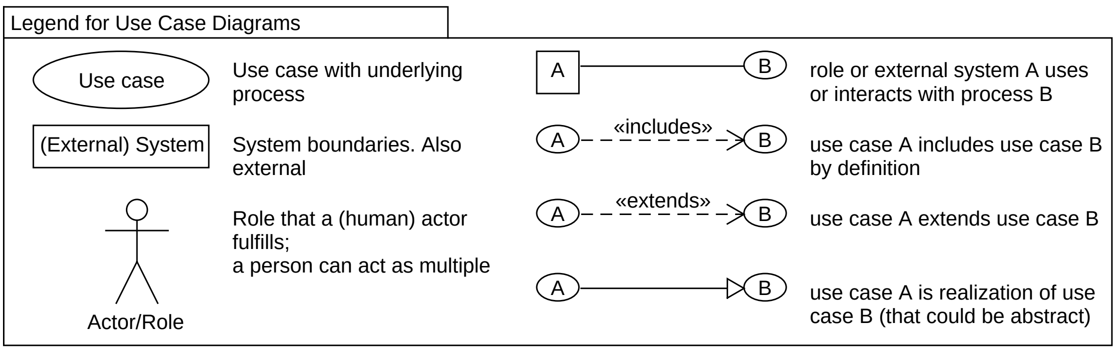


Figure 4.1: Legend for Use Case Examples

In figure 4.2 we give an overview of the use cases of this system. Every user can act in multiple roles. Data is the generalized term for details of theories/models, concepts, measurements and links between each of them. This includes meta-data such as the language, disciplines or data sources.

On the left, there are two roles that mainly add data. The **data collector** is a person, who extracts model or measurement structures from papers, books or other sources. He is only

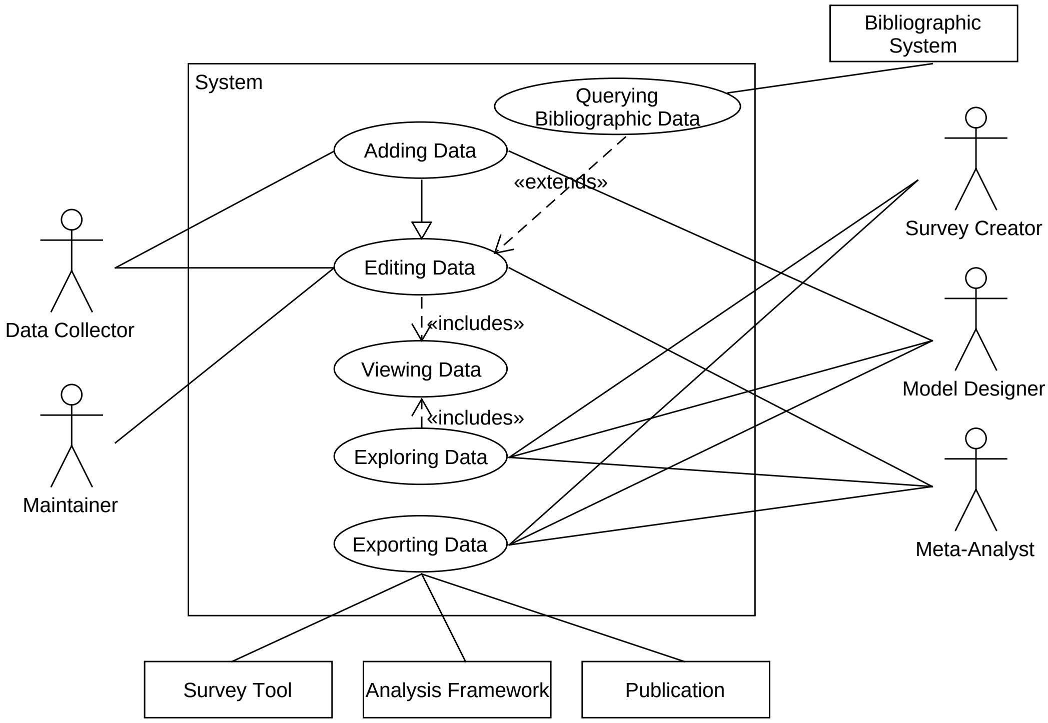


Figure 4.2: Use Case Overview

marginally interested in reusing the existing content, namely only, when it saves time during the data entry process. Typically this represents the work of a scientist, who extracts models and measurements from a paper, or it is used in combination with a role that finally uses the entered information, for example a meta-analyst.

The **maintainer** is a role that represents a person that assures the quality of the data entry (e.g. a supervisor of a student) and corrects mistakes. Alternatively, it could be a person that has a huge background of knowledge in a certain field and during his/her work can edit or add existing information. Finally, this could also be a person that adds data during his/her analysis that has not been stored before or is incomplete. For example, if the analysis differentiates between disciplines, some of them might be maintained by an expert while other disciplines might be represented sparsely.

Both of **these roles are either adding or editing data**. This can be subsumed under editing. If one edits the data, he typically also views the existing data before, so viewing data is part of the editing process. Other roles, such as the model designer or the meta-analyst, might add or alter data as well. Their additions are typically not imported from already published sources, but are based on results of their work, so they might add a technical report. The model designer might add the new model and link it to its predecessors, the meta-analyst might add some new links between models or measurements. For example, when they are very similar.

The two latter roles, however, are mainly roles that do not enter data, but use data for their tasks. If they need to add data for their analysis before, the person also has the role data collector or maintainer. It is assumed, however, that they might add data as a result of their work. All of them are targeting publications, so they want to reuse diagrams or tables from the system, for example in latex markup.

On the upper right, we start with the **survey creator**, who uses existing models and measurements to create a questionnaire for his research. In general, he/she selects a single or multiple models, that have been designed in similar contexts or for similar tasks. Typically the models are not changed, so it is just important to find appropriate models and reuse their indicator items. Therefore, the items might be exported in order to use them in a survey tool. The results of the survey are typically analyzed according to the underlying model structure as well. Therefore, he might export the structure for his analysis tool as well.

The **model designer**, in addition, is going to create a new model. Therefore, a more detailed exploration might be done. Whereas the survey creator just looks for appropriate models, the model designer might want to reuse parts of existing models, such as causal relations or operational definitions of concepts. Often, existing models are modified or expanded to fit another context. The model designer might also be a survey creator, because the models are most often designed for a new study or verified with the help of a new study. Eventually, he might add the new model to the database and link that with others.

The lower right role, the **meta-analyst**, compares different models and measurements in order to extract similar structures and modifications. This may lead to a survey of models for a paper or as a starting point for the work of a model designer. Sometimes models from different disciplines or fields are not referencing models from other disciplines or fields. Experts of one field might not know about the relevance of their model to other fields and if the terminology differs as well, they won't find them either. A meta-analyst might focus on these possible differences or he might use external tools for his analysis, such as clustering, possibly in combination with natural language processing. The findings of a meta-analyst might include additional semantic links between models or concepts, that can be added to the database. For example, he might refer to similar models from other disciplines or state relations of the models that are not mentioned in the papers in which the models have been published.

The **external systems** shown in the use case diagram, such as bibliographic services, survey tools, analysis frameworks, and publications, are generic ones and there might not be a well adapted interface. However, with the help of additional data conversion tools, it is possible to create an automated data conversion for common workflows. Some application programming interfaces (APIs) will be added in the future as well.

4.1.2.1 Adding and Editing Data

Adding and editing data deals with four main data types, namely models, concepts, measures and sources as shown in figure 4.3.

Models, concepts and measures can and should reference a source that can be used for identification and disambiguation. Therefore, it might be necessary in every case, to include the option of editing sources as well.

As models consist of concepts and concepts are often operationalized by measures, it is often useful to edit them simultaneously. There might be cases for which only the measurement is described (as in scales handbooks) or the models are not going to be operationalized.

Adding any type of data can be treated in the same manner as editing data. The only difference is, that there is no existing data. Nonetheless it is important to differentiate

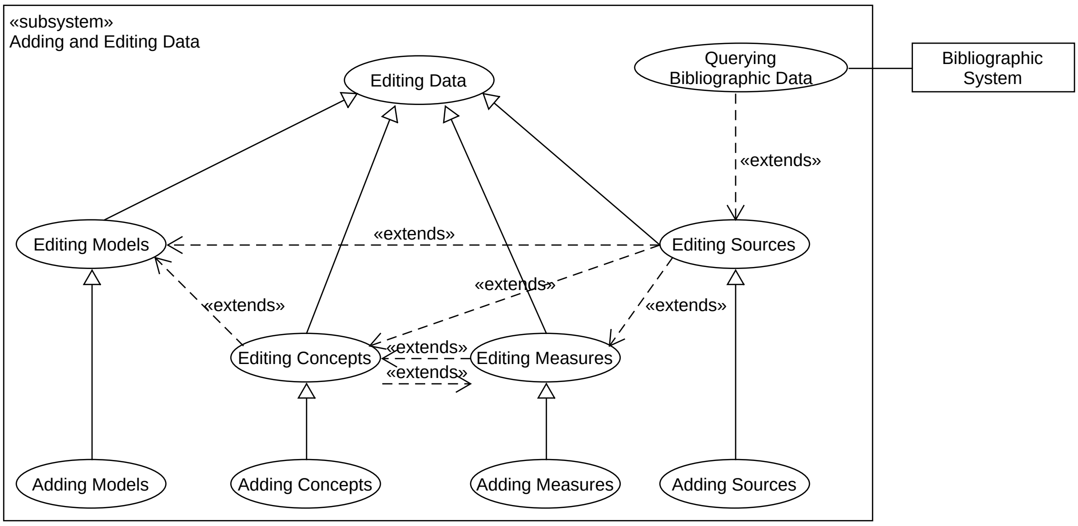


Figure 4.3: Adding and Editing Data

between those use cases, because editing can include smaller changes (e.g. correcting typos) whereas adding data normally includes a whole set of elements.

Editing data does also include the handling of duplicates. This might be supported by a separate process, that is not outlined here in greater detail.

4.1.2.2 Viewing and Exploring Data

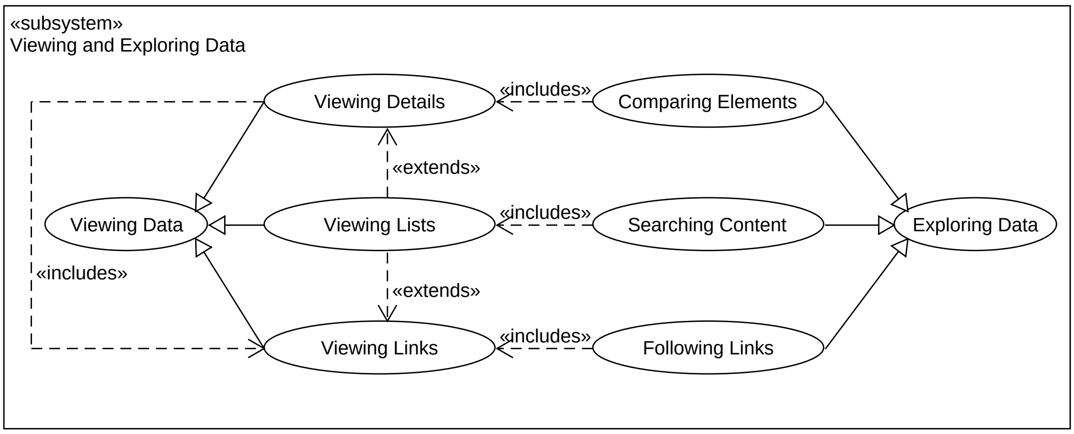


Figure 4.4: Viewing and Exploring Data

Figure 4.4 shows details of viewing and exploring data. Viewing the stored data includes models, concepts, measurements and sources as well. These are omitted here for better readability.

Typically one views the details of a certain element. A user might select the TAM model, for instance, and wants to see the existing data on this element, such as causal relations, concept names or the original authors.

Details of this element include references to details of linked information, such as the authors' details. It can also include links to other concepts that are semantically annotated, for example links to predecessors or similar models.

If the user is going to run a search, it is easier to compare search results in a summarized tabular form or as a list. These lists contain information of several elements in a compact manner. The user views lists implicitly on details pages as well, for example when listing links to other details.

The idea of the database is not only that a user can view details of a concrete element, but can also explore the data. For example, he might want to select an appropriate model for his study and, therefore, compares several models with each other.

Users that are not only working on a limited set of elements, should be able to search for elements or they are trying to follow existing links between the elements.

4.1.2.3 Exporting Data

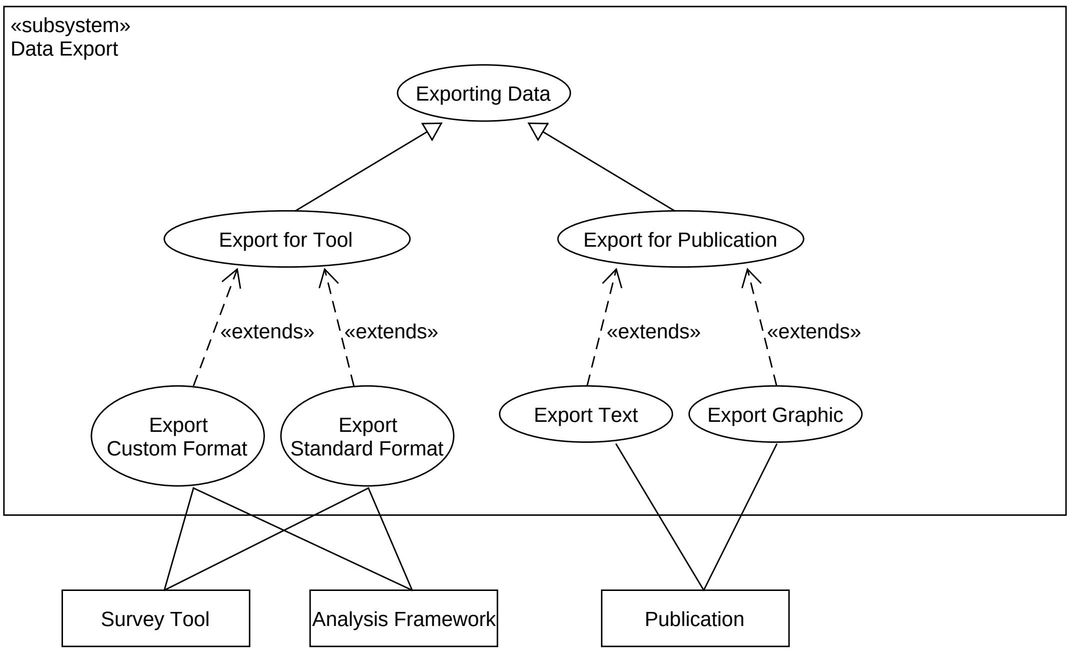


Figure 4.5: Exporting Data

Besides adding, editing and viewing data, it can be useful to reuse data in another form or format. Therefore it should be possible to export data either for publications or for tools (See figure 4.5).

Export for publications makes it easier for authors to document their workflow. They can describe the different models with the help of the existing data in the database. Especially the generation of graphics, for example a visual representation of a causal model, can be done by the proposed system.

Export for tools is a more extensive use case. Typically there are two different approaches when reusing data: On the one hand the data can be prepared for selected tools with a custom format and maybe even with the help of an interface in the form of a custom application programming interface (API). On the other hand, the data might be exported in a standard format that can be transformed to any other format with third-party tools (e.g. survey tools or analysis frameworks).

4.1.3 Restrictions and Assumptions

4.1.3.1 Application Layers

In this work we are only covering the application layer, that is directly linked with the scientific domain.

Therefore we are describing the **domain and data model** for handling survey-related data such as scientific theories and models, their description and attributes as well as their relations. Additionally, we cover the measurements for the contained concepts and constructs including the used indicator items. This information is identified by its source, typically by a scientific publication.

However, there are some other issues that have to be addressed for a **collaborative software** or **groupware** infrastructure. For example, we have not designed a rights management system with user roles, authentication or authorization. This might not be relevant for single person usage, but for many other groups. It may be interesting to allow access to a limited set of entries (probably due to rights issues), provide private working spaces for teams or to enable quality checks or reviews (for example controlling the work by a typist or a student).

The system is not yet designed as a **working environment**, in the sense that a user might want to bookmark certain models or create sets for specific tasks. Collaboration requires options to discuss content (e.g. commentary sections or discussion pages) or to find experts in an area, that are contributing actively. There are also no algorithms implemented that might add information for to the individual the user, such as recommender services or personalized searches.

We are also not talking about **version control** and **history of entries**. It might be interesting to retrieve former states or to give people credits for changes. Especially when relying on a database like this for publications, it is relevant to refer to a certain state of the data. However, as most information is based on literature, it should not change significantly over time.

All these additions can be useful and might be considered in the future. However, at this stage of the system, it is important to fulfill the core task as a data- and knowledge-base for the mentioned survey-related meta-data.

4.1.3.2 Domain-related Restrictions

The primary intention of the proposed database is to store existing theories/models, concepts and measurements. We require that every entry can by identified by a unique source, typically a written publication. It might be possible to assign technical reports or virtual identifiers as a source and to change the data over time. However, this is not the intended use and might lead to difficulties.

We allow to document the language of the models and measurements, but there is no further support available. Measurements that have been translated in several languages, for example, are treated as separate measurements in general. They can be related with a link that states the translation, and there might be several translations into the same language that are equally correct.

4.2 Data Model

This section contains the description of the data model that forms the prerequisite for the actual implementation (See section 4.4).

At first we provide an overview of the domain model (section 4.2.1) which includes elements of the market research process and uses the terms which have been defined in the section about the philosophy of the social sciences (section 2.1).

Secondly, we are proposing a collaboration tool, that allows several people to work on the same datasets. Even if it is not the main goal of this work, some fundamental collaborative design elements (subsection 4.2.2) should be implemented for a basic design. Further improvements of usability or support for other processes can be built upon that.

After the definition of the core elements and how they are related to each other, we are going to describe each of the elements and their attributes more detailed in 4.2.3.1.

4.2.1 Domain Model

In the beginning we are going to describe the schematic overview of the domain model (4.2.1.1) and cover each element more detailed in the following.

We are going to use UML class diagrams to describe the domain model. However, several elements do not need to be represented by a class in the implementation itself. Figure 4.6 shows the legend for the following diagrams that are describing the fundamental relations between the most important elements.

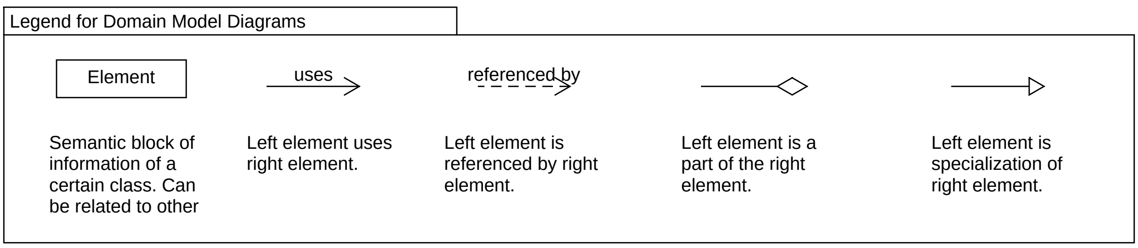
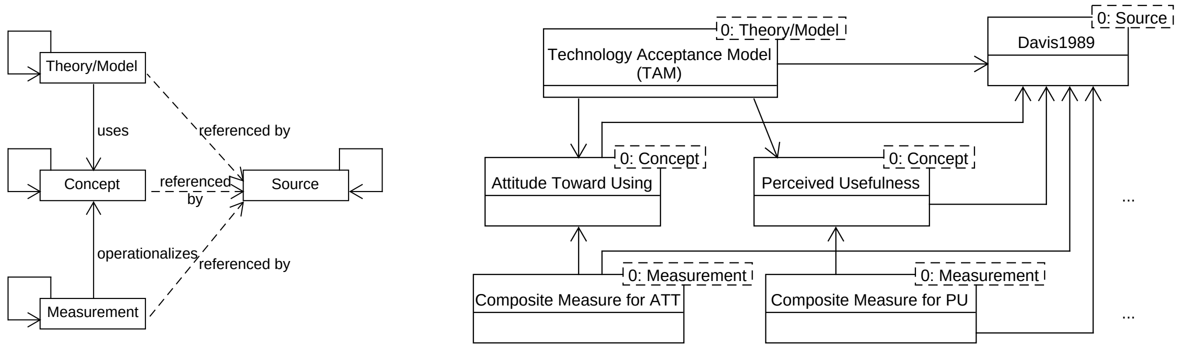


Figure 4.6: Legend for Domain Model Diagrams

4.2.1.1 Schematic Overview

The main goal of this database is to represent scientific theories or models. Figure 4.7a provides a schematic overview of the most important relations between a scientific theory, its



(a) Schematic Domain Model

(b) Example for "Technology Acceptance"

Figure 4.7: Schematic Domain Model with Example

models and the operationalizations. Figure 4.7b gives an example based on the *Technology Acceptance Model (TAM)* by Davis [Dav89].

Every model that is basically a set of concept relations, consists of several concepts. The *TAM model*, for example, consists of the concepts *Attitude Toward Using* and the *Perceived Usefulness* that are related causally with each other. There are other concepts and relations in this model as well.

Every concept may be measured in several ways, but according to the original source, there might be a defined measurement description for each concept.

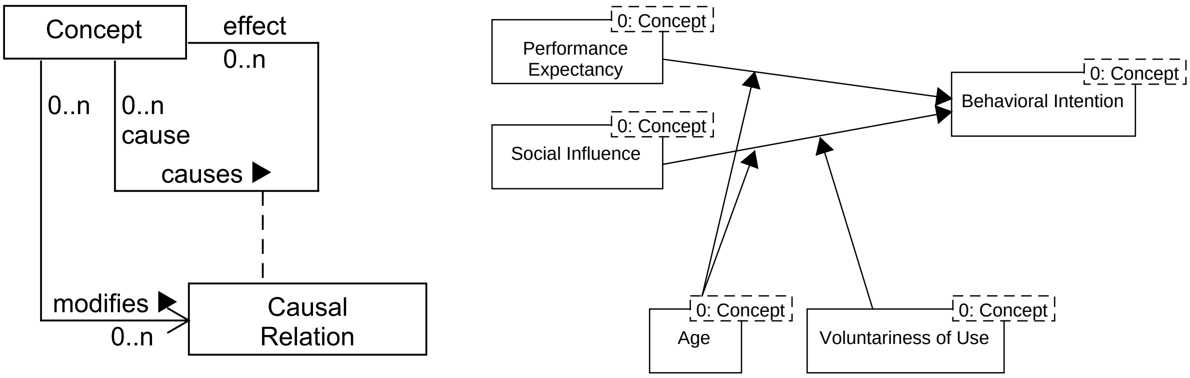
These three components are referenced by a given source, here the work by Davis [Dav89]. That means, that a source can describe multiple theories, concepts or measurements. However, a specific concept definition or measurement definition as well as a model is always related to a single primary source. They can, however, be referenced by a secondary source. Thus a given label or name for any of these elements is not necessarily a unique identifier, but has to be used in combination with the source. An additional qualifier, such as the page number, might be necessary if the label or name appears multiple times in the source.

4.2.1.2 Details of Theories and Models

As stated before, a theory or model defines the (causal) relations between concepts.

As shown in figure 4.8a, a causal relation is a reification of the concept-concept relation where one concept causes an effect that is represented by another concept. As shown in the accompanying example in figure 4.8b, *Performance Expectancy* influences *Behavioral Intention*.

The statement of the causal relation might additionally be influenced by the presence of another concept, in causal diagrams this is typically represented as a modifier. *Age* or *Voluntariness of Use* are modifiers of the outlined causal relations but are not necessarily representing a cause for *Behavioral Intention*.



(a) Details of Causal Models

(b) Example of a Causal Model

Figure 4.8: Details of Causal Models with Example

4.2.1.3 Details of Concepts

A concept can be represented by a conceptual definition or an operational definition. If the concept is contained in a scientific model, it is also implicitly described by the relations to other concepts and their conceptual definitions.

The conceptual definition is typically a description in natural language that can include a description of a certain context. In the proposed database, this includes a label and a description in prose and possibly a categorization in scientific disciplines or fields.

The operational definition is represented by its measurement definition in this model.

A philosophical concept might the same, even with different conceptual or operational definitions. It might also be different, even if the label or description sounds similar. Therefore, we qualify definitions by their source in order to be able to identify them.

Therefore, it is also possible to define concepts without the need of a model and search for possibly related concepts in natural language.

4.2.1.4 Details of Measurements

The details of a measurement, in the sense of an operational definition, are outlined in figure 4.9.

A measurement always consists of one or several indicators (with respective domains) that allow to qualify or quantify attributes of the given concept.

If one is measuring the body-mass-index (BMI) of a person, for example, a direct measure would be a question that requires to enter the BMI numerically. Here, the domain would be in the form of a numerical input given a measurement unit, if required.

Another option would be to use a composite measure, that combines the values of more than one indicator variable and processes a value for the BMI. As the BMI is calculated by weight and height of a person, these two indicator variables could be used to refer the BMI.

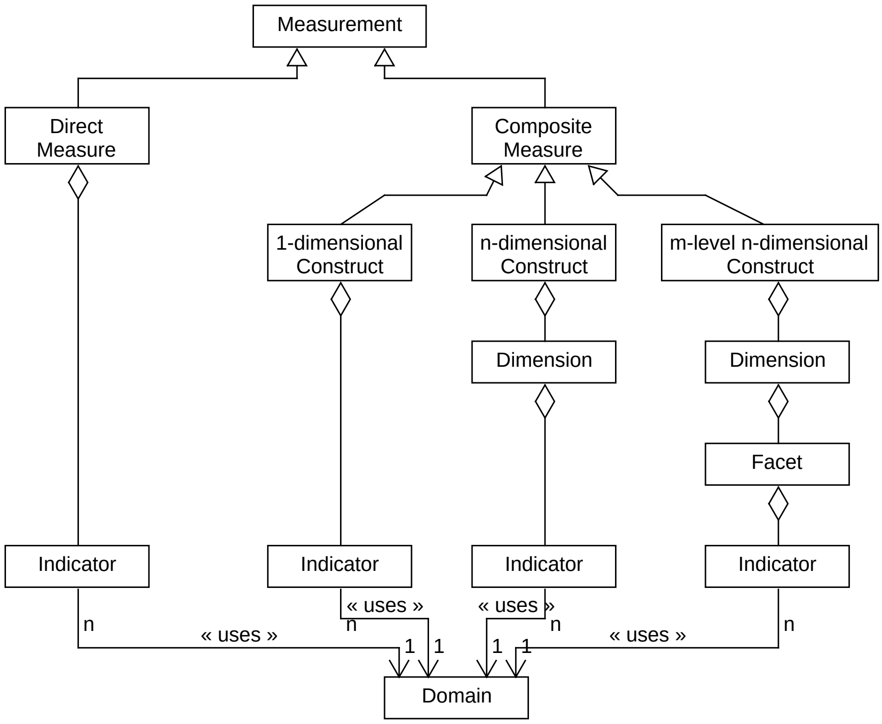


Figure 4.9: Details of Measurement

Composite measures contain indicator items that might be divided into different types of constructs. Uni-dimensional measurements, such as the example of the BMI-measurement above, combine all indicator variables to derive a value for the concept variable. Multi-dimensional constructs typically consist of more indicator variables than uni-dimensional constructs. The variables are more likely to interfere with each other so that it is difficult to assign them to a single output. Often these variables are divided into sets that represent the different dimensions. These dimensions can also be divided into subsets, we call them facets.

An example for a multi-dimensional composite measure is the personality inventory NEO-PI-3. McCrae and Costa [MC07] mention five factors (here dimensions) and 30 facets that are measured with 60 items. Another example is the Servqual measure for the quality of services (See [PZB88]). They are using 44 items that form five dimensions.

4.2.1.5 Details of Sources

In order to identify models, scales or measurements we assign them to a source. In most scenarios the source will be represented by a scientific publication such as an article or a book. These publications can be identified with the help of their meta-data or external identifiers such as a DOI or ISBN. Internal reports might also be identified with the help of a unique number.

We assume that a source is not going to change in the future and, therefore, the related information is not going to change either (the data in the database might be transcribed incorrect and corrected later, however).

Even though the database is not (yet) intended to support the development of new models directly, there might be reasons for storing information that is not documented in a publication.

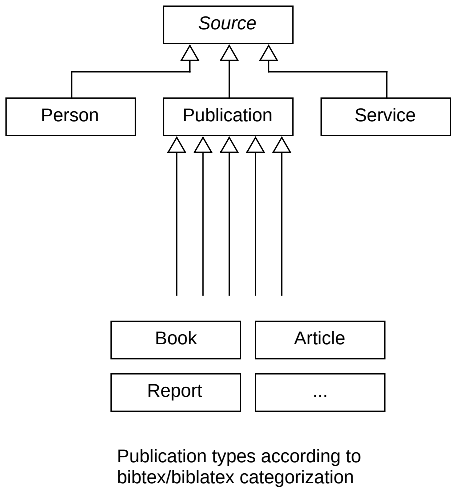


Figure 4.10: Details of Sources

For example, when developing new models it might be of interest to compare them to existing models in the database. Therefore, a person can be chosen as a source, so that others can see that the model is not yet published. There might also be services or tools that add data, such as machine learning algorithms that develop new models with artificial intelligence. Another scenario would be a survey tool that adds new measurements that are currently under test.

The different facets of a source are outlined in figure 4.10.

4.2.2 Collaborative Design

In addition to the domain model, the aspect of collaboration is an important one for the proposed infrastructure. However, there are many different use cases and scenarios that might be of interest and, therefore, the requirements for collaboration can be very different: Collaborating in a small research group is different from cross-national projects or even a platform that is open to the public. Hence, we are not going to develop a complete collaborative architecture, but focus on some important concepts that allow to extend the system later.

On the one hand we focus on the history and maintenance (4.2.2.1) of the entered data and on the other hand we look at the personalization and customization (4.2.2.2) of the application. Finally, we discuss user groups and access rights in short (4.2.2.3).

4.2.2.1 History and Maintenance

In order to be able to rely on the information in the database and to use the data for scientific publications, it is important to be able to reproduce the used data and the changes to the data in the meantime.

This means, that every entry has to contain meta-data such as the date of the last change and at least an option to retrieve a comment on the changes.

Collaboration tools such as the MediaWiki application, allow their users to view a history of changes, differences between the entries and the authors that are responsible for the changes.

Additionally, it is useful to provide an overview over the recent or total number of additions or changes in order to identify difficulties. For example, if a certain entry is edited a lot more than the average entry, it might be possible that it can not be identified properly and is, therefore, ambiguous. Sometimes it might also be of interest to list the number of total entries, activities or users.

Besides that, it is useful to provide options to flag certain entries. The users might detect a problem and should be able to inform other users or the maintainers about that. If there was a problem with disambiguation, a user could flag this entry and a warning message would be displayed. Other entries could be marked as drafts or private so that they are not proposed to others in search or comparisons. Sometimes it is also valuable to allow users to comment on entries, so that they can discuss the content or point out special features or mention additional information.

4.2.2.2 Personalization and Customization

Meta information about the users that added or changed information does improve collaboration as well as personal use. Experts on a certain field can be identified more easily, when they appear to be very active in that field. Especially in research groups, it might be easier to ask a person for clarification than to figure out some information by oneself.

People that are using the system on a regular basis want to find the entries they are working on faster. They might want to list all their edits. Additionally, it might be of interest to bookmark certain entries in order to save them for future use.

When including a search engine or recommender systems, they could improve the search results or their recommendations with regards to the single user.

4.2.2.3 Groups and Access Rights

We decided to not go into the details for access management, because that differs considerably in different usage scenarios. As we want to assign entries and changes to certain persons, it would be easy to limit access to them. One might want to introduce groups of users that can share content (e.g. different working groups) or that have different access rights (e.g. professors and students). Depending on the scenario more extensive modifications would be necessary, such as license management or access based on payments or ip-ranges.

4.2.3 Core Elements and Relations

In order to be able to store the relevant data, we define the main elements and relations that should be stored for the basic purposes of the database. The elements might be represented as multiple tables in the database design, but they form a single unit of information that is described in the following paragraph 4.2.3.1. Relations with attributes are treated as reifications and might be treated similar to elements in a database, but they have a semantically different meaning. The core relations are described in paragraph 4.2.3.2.

4.2.3.1 Core Elements and their Attributes

The core elements are representing the elements that are schematically related in the domain model. The application can be structured according to these core elements, on the user interface, in the code and in the database. Additionally, every core element should be presented to the user as a semantic view on the data without knowing the underlying object or database structure. Every core element contains additional attributes and interaction methods such as a version history or collaboration features (See 4.2.3.1).

In some cases a further differentiation of the core elements has been made in the implementation if it made sense. These additional differentiation might include elements that belong to a certain core element but are managed similar to them. For example, when describing a measurement, it might be described by a collection method. Limiting the choices for the collection method might be done directly in the source code or with the help of a configuration file. However, it might be more flexible to allow users to add new data collection methods. These methods might also be described in greater detail and they might also contain references to publication sources. As this element belongs to the measurements on the one side but can be treated separately from a concrete measurement instance, it should also include the meta-attributes as a core element does.

In the following, the most important attributes of each concept will be outlined.

Publications Publications are time-invariant documentations for research work and findings. Therefore, it is important to rely on publications as a form of reference system, where a researcher can read about the research goals, the used approach and the results. There might be a discussion on whether documents written in prose are sufficient for conserving and reusing scientific results that often include code or research data, but we won't discuss this here. Most often it is possible to store other data formats or links to databases as publication as well.

Publications are relevant for the meta-analysis of scales and models in a similar way as persons are. They are bound to a certain state of a scale or a model that can be used as a reference in future uses or modifications. Depending on the meta-data provided, publications can be classified according to publication year, journals, scientific disciplines or keywords. This information might be interesting when comparing scales/models with separate origins and development paths. For example, technology acceptance research is done in computer science (e.g. QIU-Model [ISO 9126-4:2001]), management information systems (e.g. TAM [Dav89] or UTAUT [Ven+03]) or marketing (e.g. TRI [Par00] or ABM/OAM [Dab96]) without many links in between.

Publications can be identified with uniform resource names such as ISBN (books), ISSN (serial publications) or DOI (electronic resources). Additional information can also be requested from various library services such as the Digital Public Library of America [DPLA] or the Linked Data Service of the German National Library [Hen15].

Relevant data is shown in table 4.1. It may also be possible to use external services for the general data handling, however it might be a performance issue. As the Publication is one of the core concepts of the database and people might search for it, at least the most important attributes such as author names, publication year and title should be stored. Additional information for specific analysis purposes can be fetched from other systems if needed (e.g.

different classifications of literature or tags for the publications). In order to assure that we do not drop bibliographic information that is not modeled in the database, we are storing information in the BibTeX-format as well. Particular types of publications contain additional attributes. An entry of a journal article, for example, contains the title, the volume and the number of the journal and it references the pages of the article in the journal. See the BibTeX/BibLaTeX document types (See [Leh+15, p.7–13]) for additional publication types and their attributes.

Publication
Type of Publication Authors (Order, Name, Institution, E-Mail) Title Year Abstract File External Resource (Authority, Link, ID) Status (Published, Draft, ...) Classification (Tags, Disciplines, ...) BibTeX : Text

Table 4.1: Attributes of a Publication Element

Models Models are representing a set of causal relations between concepts, as they want to test whether certain concepts influence others. The nature of models can be highly different, so that it is not easy to provide a unified structure (as shown in table 4.2). In that case, the description should include a mathematical definition of the relations.

Finding models for surveys mainly implies to find the included scales and scale items. In practice the results have to be analyzed with regard of the model definitions. Some common model types such as SEMs could be distinguished, however.

Model
Name Description Type (SEM, . . .) Concepts (nodes) with causal relations (edges)

Table 4.2: Attributes of a Model Element

Measurements A scale is the central concept of this database. The attributes of the scale are used in order to filter, compare and cluster scales.

For practical purposes, the definition of answer options and items is the most important attribute. These attributes are presented to the respondents and if they are modified, the whole scale could be different. However, as one typically needs the score as a final result, the steps to calculate that are important as well.

As there are mainly descriptions of scales in prose and no common semantic structure exists, it might be advisable to include fields for full-text descriptions (concept description, scale description, scale development description) as well as discrete characteristics of each attribute (such as data collection method is rating or scaling method is Likert scaling). We propose to define a set of required attributes in advance, so that comparison of measurements is possible. Additionally, we allow users to add additional attributes when necessary. See table 4.3 for a set of predefined attributes.

Measurements are linked to publications in a way, that there should be primary sources including the original description, and secondary sources such as handbooks of scale that often provide preprocessed information. Additionally, other references on the concept/construct, item selection or scaling/scoring methods might be important.

It is assumed that every study that uses this measurement, might include small variations of the items and so it should be treated as a separate measurement.

Measurement
Name Conceptual Definition Data Collection Method (Type of Method, Measurement Level, Response Categories (order, verbal, numerical, visual), Balanced Answers, Forced Answers, “No Answer”-Option) Dimensions/Facets/Items (Dimension Name, Facet Name, Item Text, Item Order, Reverse Item, Answer Options) Scoring Procedure (Description) Scale Development (Description, Scaling Method) Publications (Primary Source, Secondary Sources, Other References)

Table 4.3: Attributes of a Measurement Element

Meta Attributes The structure presented in table 4.4 does not represent a single concept of the database, but includes meta-data relevant for the operation of the database itself. Users can be identified as creators of an entry. This is useful creating access rights. Additionally, especially in smaller teams, experts for certain scales or topics can be identified.

The modification date is important as well, especially to indicate that there might have been errors in the data before. Of course, this does not replace a proper version control history at all.

Meta Attributes
Object Creator Creation Date
Last Modifier Last Modified Date
Language Comments on Object (Commenter, Date, Text)

Table 4.4: Additional Attributes for each of the Core Elements

4.2.3.2 Core Relations and their Attributes

As shown in the schematic domain model (4.2.1), there are several relations between the elements. We are also covering the main relations here, as database-specific relations are implementation-related but not necessarily require a global semantic.

Relations Between the Same Core Elements Every core element can link to another core element. Typically these relations are similar for all core elements. The basic relation states that there might be some semantic similarity between the entries, similar to the way publication references are referencing other publications. However, in order to have more information on the relation, we might want to type these relations.

Weak relations might be described with types as "is similar to". Different models might describe similar topics or they might contain the same idea. Sometimes, however, one could state that one model follows the idea of another model and type the relation with "based upon". It might also clearly stated that a model is a successor of another model, so one might state that it directly "succeeds" the other model. Additionally, a model can be the "specialization" of a more general model.

Regarding the content, a relation might also state that a model is a "translation" of another model. One could also state "equality" for two models, for example when they are published under separate names or in separate publications but are exactly the same.

When going a little further, one might also want to state similarity because of usage behavior or similar topics. This might be related to recommender engines that try to figure out related models because of time-related queries or user profiles.

It might also be useful to track implicit links between models when their publications are linked. For example a comparison paper might mention two different primary sources, so the models might be related somehow. Depending on the paper this relation might be typed as well (comparison, similarity, survey of the field, ...).

Relations between different Core Elements The most important relation is the relation of every core element with a certain source. As the source is used for identification, every element should be linked with a primary source, where the information came from. If the primary source is not at hand, a secondary source might provide the relevant information. However, the secondary source might contain some errors in its description, so it is necessary to document that this entry is based on a secondary source. Other related sources might be references to publications that cover the domain, that are mentioned as related examples or similar.

There might be many more facets of other sources that are referencing the model but are not referenced by the model. Papers on studies that are using a certain measurement, might not describe the measurement itself, but they might mention the original measurement. Publications that compare models might not use the models in their own research but they are evaluating or comparing specific aspects of several models and, therefore, the model is referenced as well. These publications might provide important information on the quality of models and measurements.

Another relation is the relation of a model to a concept or a concept to a measurement. Generally there is also a relation of the model to the measurement directly. Given the case that a new model is proposed, the original paper describes the model, the concepts and the operationalization of the concepts. However, another study might reuse the model structure, but they are using new operationalizations, either because they want to improve the measurement or because they are applying the model in another context. They might also adjust the concept definition as well. Therefore, a model might link to several conceptual definitions and operational definitions. Therefore, it is important to state the origin of this relation as well, as we describe in the following paragraph about the meta attributes of relations.

Additionally, a concept could be measured without an underlying model. This might be the case when a researcher tries to evaluate separate operationalizations. A concept might also be measured with the help of another model.

Abstract Relation
Core Element A Core Element B
Type of the Relation Source (e.g. Publication, Experience, Service) State of Validation (“Proposed”/”Valid”)
Creator Creation Date Comments

Table 4.5: Abstract representation of a Core Relation

Meta-Attributes of Relations Relations might be set by a user manually or by an automated script that analyzed the content. Therefore, the relation does not only need a certain type but also a differentiating factor whether the relation is an assumption or a justified fact.

This means, that a user might state that a model B is based on another model A, because the primary source of model B states this. So he would add the publication as a reference. However, he might also link two measurements that look similar in order to assist other users that might look for the appropriate measurement. An external tool that analyses the similarity of models with the help of natural language processing might add relations based on this approach. So it might add the service name and the level of similarity of these two models. The attributes also store the creator, the creation date and they might contain a natural language comment.

Therefore, meta-attributes of relations are similar to meta attributes of core elements. Every relation has to be described with these attributes. See table 4.5 for proposed attributes that every relation should contain.

4.3 User Interface

This section describes the user interface, that enables users of the system to do their work efficiently and without problems. The sketches are meant to visualize the core functionality of the system and not necessarily every form field that does not have any further implications will be shown.

We divided the description of the user interface into three stages of development. They are intended to follow a development process that includes a simple interface in the beginning and adds functionality gradually. This allows developers to test the core functionality and to fix errors more easily. Additional features can also be prioritized if they appear more important. The different stages can be seen as release plans, but they might consist of additional smaller steps during development. This procedure is based on ideas from agile development and rapid prototyping, see for example “Prototyping for Tiny Fingers” [Ret94] or the “The Agile Manifesto” [FH01].

Stage 1: Basic Table Mappings At the beginning users need domain knowledge and system knowledge in order to enter data in different forms correctly. They have to select the correct forms for data entry and have to link the data with the help of dataset identifiers. The forms are based on the underlying database structure and do not provide a separate layer for better usage. This enables the developers to identify problems in the system architecture or in the interpretation of the different data entry fields. Users can view the entered data but it might not be linked properly with related data. (Subsection: 4.3.1)

Stage 2: Support for typical Workflows and Data Types For a live system that is used by several contributors, it is important to guide them through the data entry process and provide them with the information they need. Users should not have the feeling that they are working close to the database and should find the important information at a glance. Thus, the system provides support for typical workflows and provides dashboards for the most important metrics. (Subsection 4.3.2)

Stage 3: Reusing Existing Content Existing content can be used to enhance the data collection steps in a way that similar content is shown and can be reused by the user. The system can link similar models and scales automatically. This simplifies the usage of the data. It provides different data export options for the use with external tools. (Subsection 4.3.3)

Every stage consists of interfaces for three actions with which a user is going to interact:

Data Entry Collecting information is the primary function of the proposed system. Most users will be using this functionality. It is important for data quality and efficiency, that users are able to find the correct forms and enter data correctly. Providing wizards that guide a user through different forms reduces complexity and can assist the user in the choice of the correct fields.

Data Presentation Users that are looking for specific measurements or models should be able to find them easily. They should be able to navigate from one entry to related entries directly. Comparisons of models or scales and recommendations of similar entries are providing the benefits over traditional scales handbooks.

Data Import/Export In order to get the most flexibility out of the system, it is important to provide interfaces for data import and export. Users should be able to analyze the data with other tools (e.g. spreadsheets or programming languages) or import them for further processing (e.g. survey tools). Data import options might lead to third-party additions or automated data extractions from other sources.

4.3.1 Stage 1: Basic Table Mappings

The basic functionality can be seen as a set of forms and tables that are set on top of the database abstraction layer. They do not necessarily imitate a database table alone, but they are covering the different core data types (subsection 4.2.3.1) that form semantically consistent groups regardless of their physical representation in the database. Nonetheless, the first development stage might be divided into smaller steps as well. We won't document those here, because it is up to the development team.

Users should be familiar with the different entry types and how they are related with each other. For example, a model like "Technology Acceptance" is described in a paper and relates several concepts that are described in the same paper. These concepts might be operationalized with measurements that consist of several indicator items.

4.3.1.1 Data Editing

In this first stage, the data entry process is strongly related to the underlying database model. In the first step, every database table should be editable by a single form. The user has to understand the application and database layout or should at least follow a documentation. The forms are only designed to check the entered information on a very basic level - for example that a number field contains only numbers.

Linking elements with each other might require underlying knowledge of the database identifiers in order to state foreign key relations. That is especially important, if an element can not be identified by its name or some obvious attributes.

The level of automation is zero. Forms do not support auto-completion or provide suggestions and when switching to another form, there will be no prefilled fields.

4.3.1.2 Data Presentation

Data presentation is oriented on the database layout as well. There will be no overview tables with details for all publications, models or measurements, but rather a list with all elements linked to the details page.

The details of an object do not need to look beautiful but should contain all data that is stored in the database. Causal models, for example, should be represented as a list of pairs, without graphical visualization.

Links between elements should be bullet-point lists as well without any additional information or functionality.

4.3.1.3 Data Export

There should be no application interface, because the frequency of changes is to high. Therefore, data export will not be supported by a user interface either.

However, if it is necessary to analyze the entered data with the help of external tools, one could do that with custom queries on the database. This might be useful to identify common entry types, differences in the entry types or to get a general overview. Tools that work with this data could be developed either way, but they should not rely on the interface for production use.

4.3.2 Stage 2: Support for typical Workflows and Data Types

If the previous stage has been completed, it is important to improve the workflow of the users. After the completion of stage 2, less experienced users should be able to work with the system. They should be guided through the most important processes and they should not have the feeling that they have to consider any technical aspects of the underlying system.

It is not the most comfortable system in every situation and the adaptability of the system is still low. Nonetheless, this stage is ready for everyday use.

4.3.2.1 Data Editing

When entering new data, users should be guided through one predefined process. Every workflow is supported by a wizard that presents the necessary forms and formsets.

The first step is that the core elements of the database, such as publications, models or measurements, are presented as a single unit of information, regardless of their underlying database structure. This means that adding authors for a publication should not require an additional form, but should be done in the form of the publication. Thus the new form sets are now hiding the underlying structure.

A second step is to provide common values for certain fields. It might be foreseeable that the most used measurement type will be a Likert scale. So this should be proposed to the user, possibly in the combination with other frequent measurement types. This gives the users an orientation of what is expected in the fields and leads to a better control of entered data.

Depending on the flexibility of the system, some fields might be restricted to certain values either from a list of choices or with the help of an additional table in the database. Users can add a new measurement type, but are only able to choose a previously added measurement type.

Thirdly, the form-sets should be linked with each other and save the user additional steps. When a user is adding a new model to a publication, he might do so by following a link from that publication page. Then, the form should have the referenced publication already entered, so that the user only has to enter the new data. Thinking of complete workflows, a wizard would guide the user from entering the publication data, then entering the model structure with the concept definitions and finally adding measurement details. The linking of several elements with each other in the database should be done automatically and hidden from the user, as he is following this process to achieve that.

4.3.2.2 Data Presentation

In order to support the users, the detail views of each element should contain a structure for better readability. Information that is related should be positioned close together and there should be larger free text fields for descriptions and definitions (for example transcribed from the original literature) as well as lists for single details, that can be extracted from the free form description. Some metrics should be calculated automatically from the existing information, such as the number of dimensions or items, and presented as well.

Overview lists should include more information and one should be able to filter or sort them with regards to this additional information. Publications might be sorted according to literature type, measurements might be filtered according to their number of dimensions.

If useful, generated graphics should be added. This could be the graph visualization of a causal model or the representation of a Likert scale, for example.

4.3.2.3 Data Export

A typical user needs some form of data export, for example when doing some calculations in a spreadsheet. Therefore, it should be possible to export the content of overview tables as a comma-separated value (csv) file. These can be processed with spreadsheets, statistics tools or even programming languages.

Elements with more details or a structure that can not be represented in a table, should be represented in the form of a json-file so that it can be processed by other tools as well.

Graphics can be copied normally without the need for an additional interface, but there are various reasons to support the download of specific graphic formats. For example, publications might require a higher resolution of the graphic or even a vector-graphics format. Users might want to alter images for the use in their presentations or on websites.

4.3.3 Stage 3: Reusing Existing Content

The final stage should provide a smart system. This stage might never be completed, as new additions should be added in an ongoing manner. The main goal is to reuse the existing information of the system and support the user in a way that the system proposes the information he needs.

He might get suggestions for similar entries or for entries that correlate with his user profile. The system should adjust to the needs in a semi-automated way, so that nobody has to define limited processes in advance.

4.3.3.1 Data Editing

The changes in the data editing process should not be followed by drastic layout changes anymore. Instead, the existing forms and wizards should be enhanced in details.

Forms should include auto-completion, that are no longer prefilled by humans, but change dynamically from database content. They might remember the last entries of the user and provide him personalized suggestions. It might also be possible that natural language processing algorithms propose certain categories automatically.

The system should also propose to copy content from existing entries based on similarity. If the TAM 2 model is to be entered, one should be able to copy the original TAM model and perform the necessary changes.

4.3.3.2 Data Presentation

Data should not only be accessed by a category structure or overview lists. Instead, there should be a search engine that shows results across entry types and lets the user refine search with defined criteria.

Users should be able to build their own user profile by the tracking of their own contributions. They should be able to bookmark or star interesting content for faster access.

Links to similar content based on wording, topics, user behavior or the user's profile should be presented in addition to the manually entered information by the users.

Finally, collaborative features like comment functionality or the flagging of entries for maintenance should improve the work in larger groups. With the help of statistics pages, one would be able to detect trending topics or difficulties.

4.3.3.3 Data Export and Import

Besides the data export, there might be many reasons for importing data as well. Users should be able to import data to the database from other sources or applications. However, the design of this step requires also the usage of standard formats that might have to be specified as well. Depending on the standard, any export option should export the data according to this standard as well. For a start, however, custom data formats should be supported so that the user might be able to convert external data to the appropriate format, like a csv-file or a json-file.

Users should also be able to use an application interface for the use with external tools. Once this works in both directions, third parties could develop new user interfaces or tools that use and create data.

4.3.4 Design Sketches

We are now going to look at some sketches of the user interface. These sketches are examples of how the design could look like, but especially the content is just for a better understanding and not necessarily correct or complete.

4.3.4.1 Data Editing

In figure 4.11 and figure 4.12 we show a form at stage 1 (“basic table mappings”, see 4.3.1) that allows the addition and the editing of details of a measurement. At this stage we provide a single page layout for a start, so that a user can read the source and add relevant data simultaneously without rereading the source again and again for every field. However, it might also be useful for users to fill in smaller forms, especially if they are new to the topic and they need more guidance. Note that at least two additional forms are necessary, one for the measurement items (see figure 4.13) and the other for the response categories (figure 4.14).

At the top of the measurement form there are fields for entering a construct’s primary name and a construct’s facet name. The idea is that the measurement will be most likely be identified with the help of the measured construct or the publication that uses this measurement. Constructs can also be concepts here, since the naming in scales handbooks is often similar. In the theoretical models one is usually using concepts, but in the measurement models (especially for multi-item measures) it is often the term construct, as it is constructed from multiple items. See also the section about the philosophy of the social sciences 2.1.

In the example we are using the *Ease of Use* concept or construct as described by the TAM model. This has, however, been adapted to *Technology Assisted Shopping*. This adaption can be added as a facet name, so that one can distinguish between the different facets. The construct definition closes the section about the construct details.

In the next section of the form we are going to add details of the measurement itself. The idea is that the *Measurement Description* field holds free text as described in the paper itself. It is similar to the descriptions in scales handbooks as well. After this block, several concrete attributes are asked in order to store them in a semantically unambiguous way. These attributes are better for structured and automated comparisons or for the use in additional tools but they might not cover all details mentioned in the description.

The next block of form fields covers measurement (or scale) development that should also contain a general description followed by selected attributes. These fields contain information like “Scaling method: Likert scaling”.

We are not yet adding data about studies or validity measures. They are missing here right now, even though collections of scales are normally mentioning them as well. Probably one should at least add a free form field for that as well. In the future development it is planned to add detailed attributes in comparison with survey results as well.

We are then providing a comment field, so that information can be added that might help other users, but does not necessarily belong to the previous topics.

Finally, the publication relations are missing in this example form. They should be added in this form later on as well. However in this stage, they still have to be added separately.

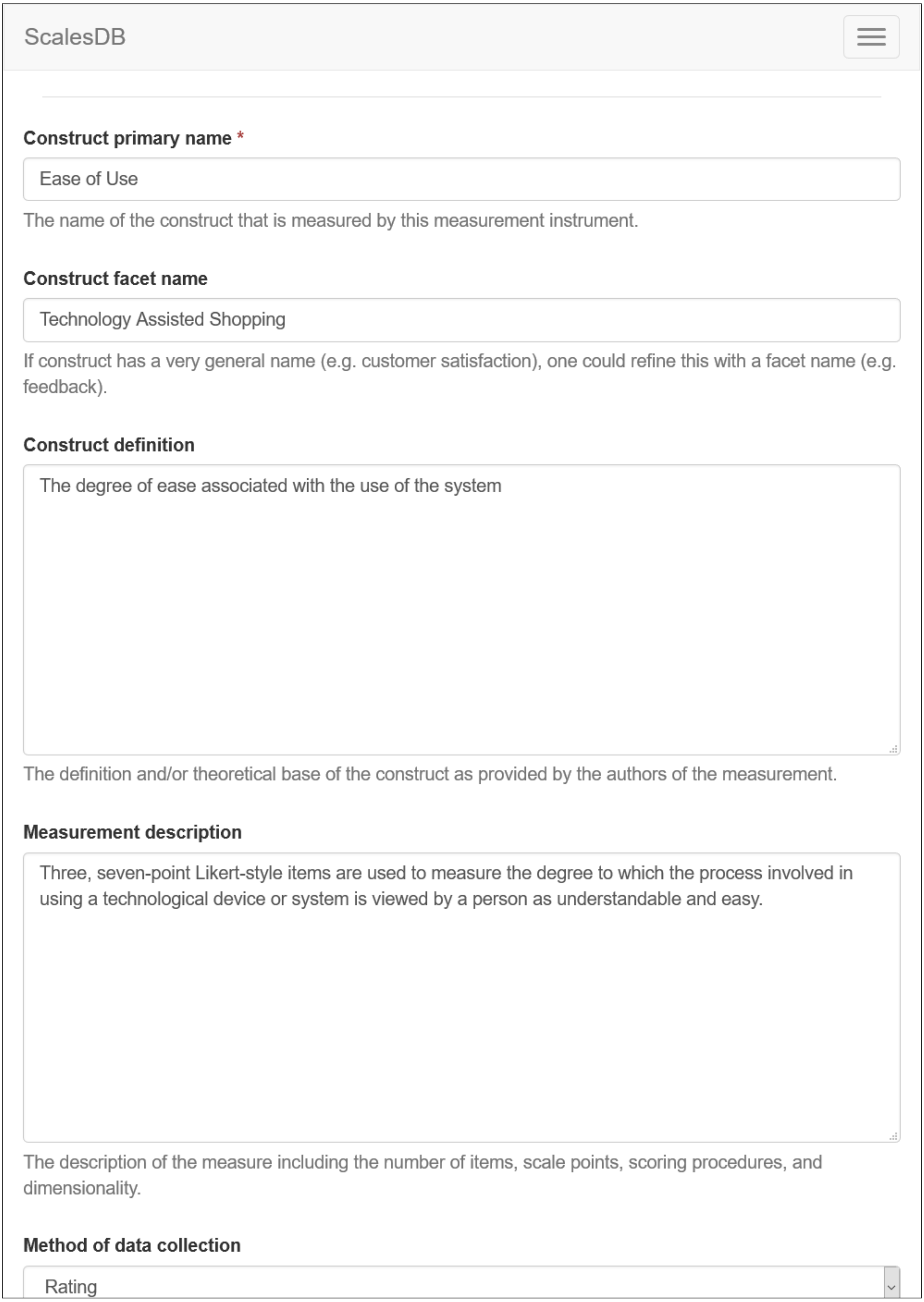
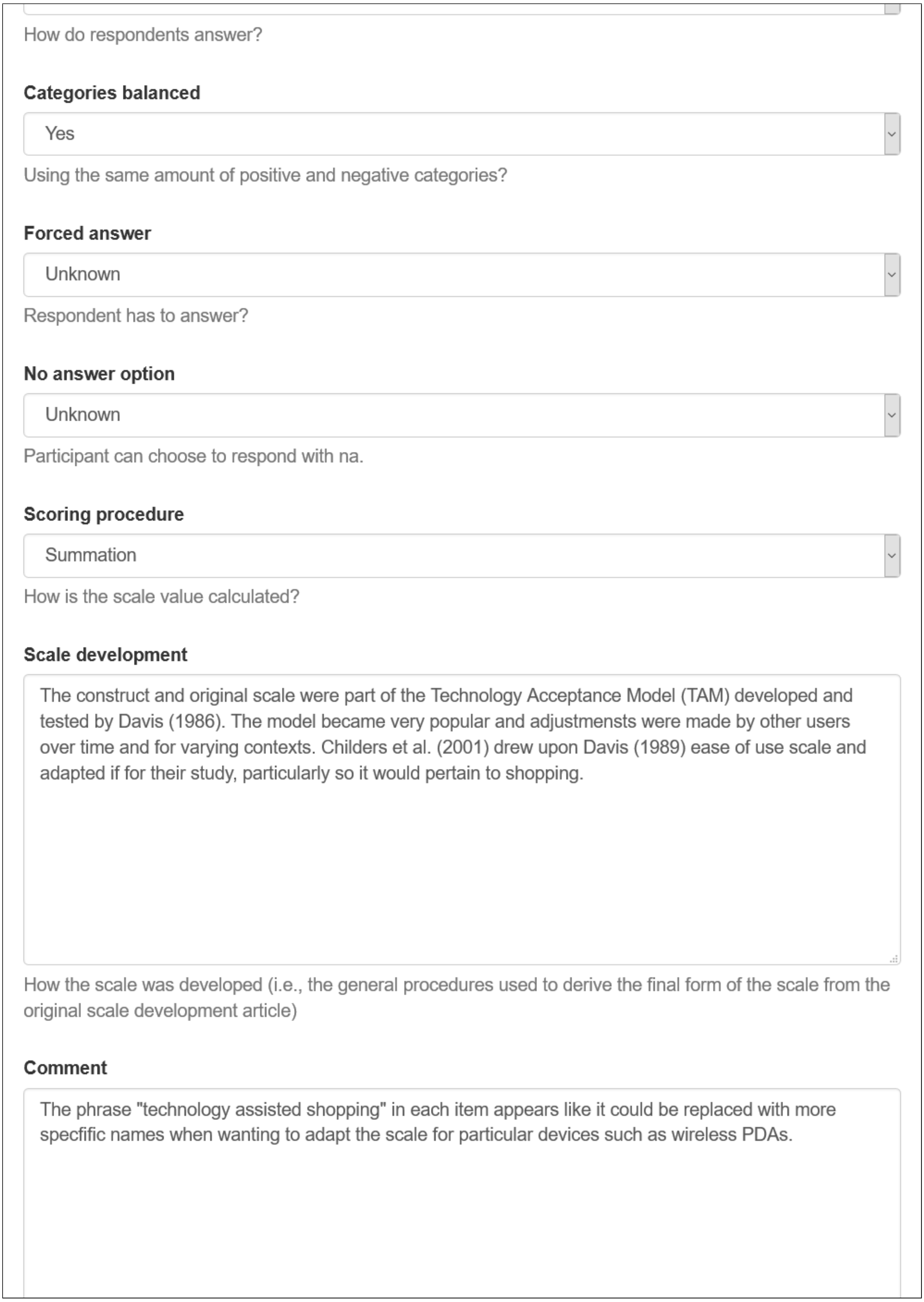


Figure 4.11: Sketch of the Measurement Form (excerpt - part 1)



How do respondents answer?

Categories balanced

Yes

Using the same amount of positive and negative categories?

Forced answer

Unknown

Respondent has to answer?

No answer option

Unknown

Participant can choose to respond with na.

Scoring procedure

Summation

How is the scale value calculated?

Scale development

The construct and original scale were part of the Technology Acceptance Model (TAM) developed and tested by Davis (1986). The model became very popular and adjustmensts were made by other users over time and for varying contexts. Childers et al. (2001) drew upon Davis (1989) ease of use scale and adapted if for their study, particularly so it would pertain to shopping.

How the scale was developed (i.e., the general procedures used to derive the final form of the scale from the original scale development article)

Comment

The phrase "technology assisted shopping" in each item appears like it could be replaced with more specfific names when wanting to adapt the scale for particular devices such as wireless PDAs.

Figure 4.12: Sketch of the Measurement Form (excerpt - part 2)

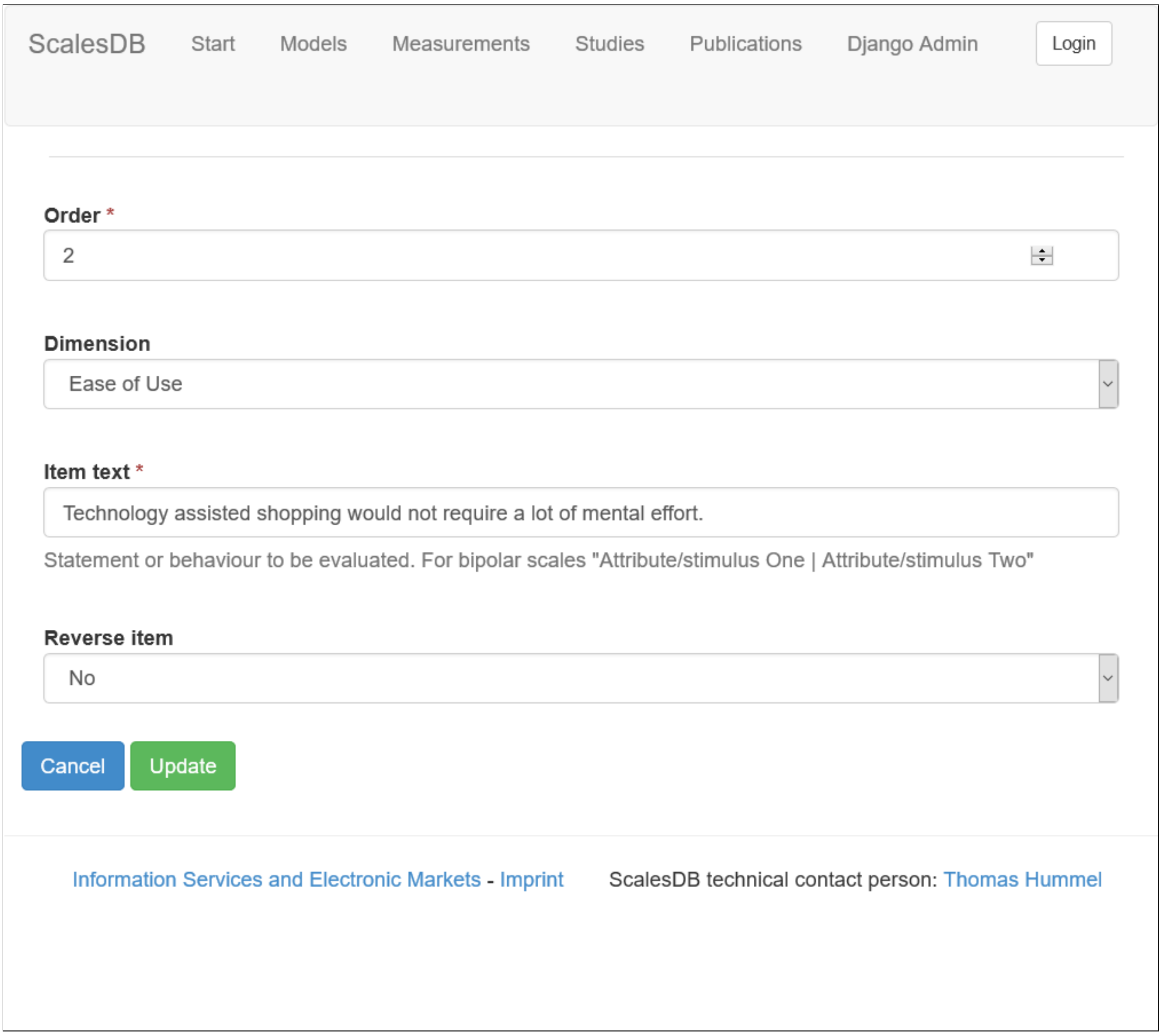


Figure 4.13: Sketch of the Item Form

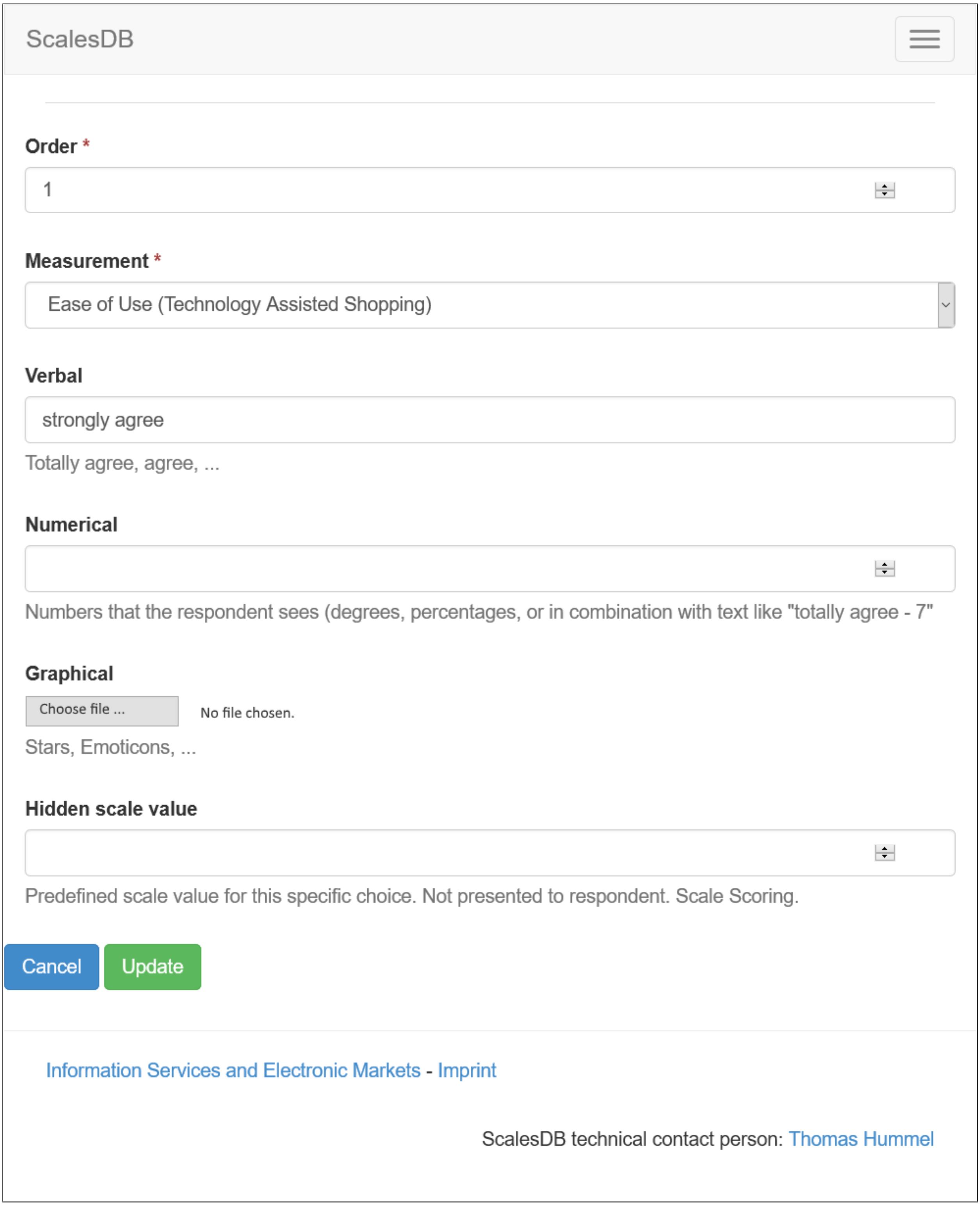


Figure 4.14: Sketch of the Response Category Form

4.3.4.2 Data Presentation

Measurement Details We are now showing a details page of a measurement in figure 4.15. This sketch shows a combined form at stage 2 ("support for typical workflows and data types", see 4.3.2). In the upper right of the screenshot, one can see a toolbox for users with several actions - at this time they lead to different forms for editing response choices, items or references. We have shown some examples of the forms in the previous section 4.3.4.1. See figure 4.16 for a magnified view of the measurement description and figure 4.17 for a magnified view of the dimensions and items section.

The details view shows the data of several database tables to provide an overview of all relevant information on this measurement. We are using *Ease of Use* as an example again. One can see the different information blocks that we have entered in the form: Construct Details, Measurement Details, Measurement Development and References.

The basic layout idea is that free-form text is shown on the left and the extracted and structured attributes are shown aside on the right. This can be seen best for the measurement description.

The response categories are representing Likert-style responses in this example and are shown in a table. They might also be visualized similar to their appearance in a questionnaire in the future. The response items have been added (similar to the response choices) with the help of an additional form, but they are shown in this combined view. This measurement in this example contains a single domain with three items.

The scale development details are presented similarly to the measurement details. Next, the view lists related sources. In this example, we have an original source where this measurement has been described. However, we used a secondary source to add the data. Even though the authors of the secondary source claim to have transcribed or described the original measurement according to the primary source, there might be differences. There is also a section called "Other sources" that references the original TAM articles as well, since this measurement has been adapted from one of the original measurements. We cut the screenshot in the publication list, so the latter part can not be found there.

This sketch does not show links to similar measurements or models that are using this measurement either. They are linked at the bottom of this page as well.

Model Details Figure 4.18 shows a model details view. A magnified excerpt of the model diagram and related measurements are shown in figure 4.19.

The theory of the model (here the TAM) is described in a free form text and includes references to other models as well. The model is presented as a graph visualization where the different concepts are connected with each other. This belongs to stage 2 ("support for typical workflows and data types", see 4.3.2), whereas the section on related measurements can already be seen as a part of stage 3 ("reusing existing content", 4.3.3).

Additional model details are rather sparse, not because of the missing information in a model, but because of the difficult standardization of model descriptions. This structure might be added in the future, but as of now most theories are described in full text and the only relevant structure is the proposed relations between the constructs.

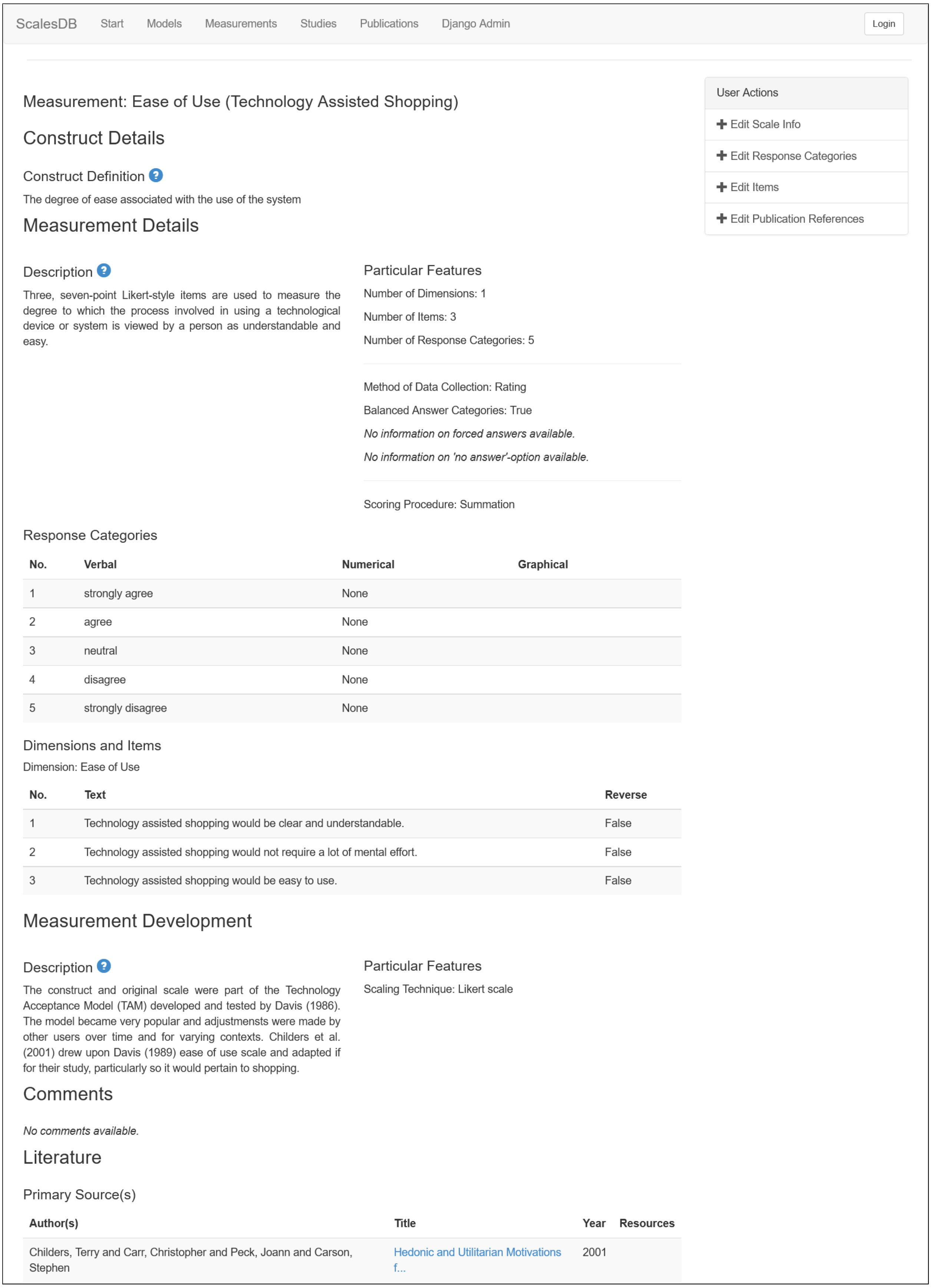


Figure 4.15: Sketch of the Measurement View (excerpt). Magnified views of the description and the items are shown in figures 4.16 and 4.17.

Description

Three, seven-point Likert-style items are used to measure the degree to which the process involved in using a technological device or system is viewed by a person as understandable and easy.

Particular Features

Number of Dimensions: 1

Number of Items: 3

Number of Response Categories: 5

Method of Data Collection: Rating

Balanced Answer Categories: True

No information on forced answers available.

No information on 'no answer'-option available.

Scoring Procedure: Summation

Figure 4.16: Description of the Measurement in a Magnified View

Dimensions and Items

Dimension: Ease of Use

No.	Text	Reverse
1	Technology assisted shopping would be clear and understandable.	False
2	Technology assisted shopping would not require a lot of mental effort.	False
3	Technology assisted shopping would be easy to use.	False

Figure 4.17: Item of the Measurement in a Magnified View

Figure 4.18: Sketch of the Model View. A magnified view of the model diagram and related measurements is shown in figure 4.19.

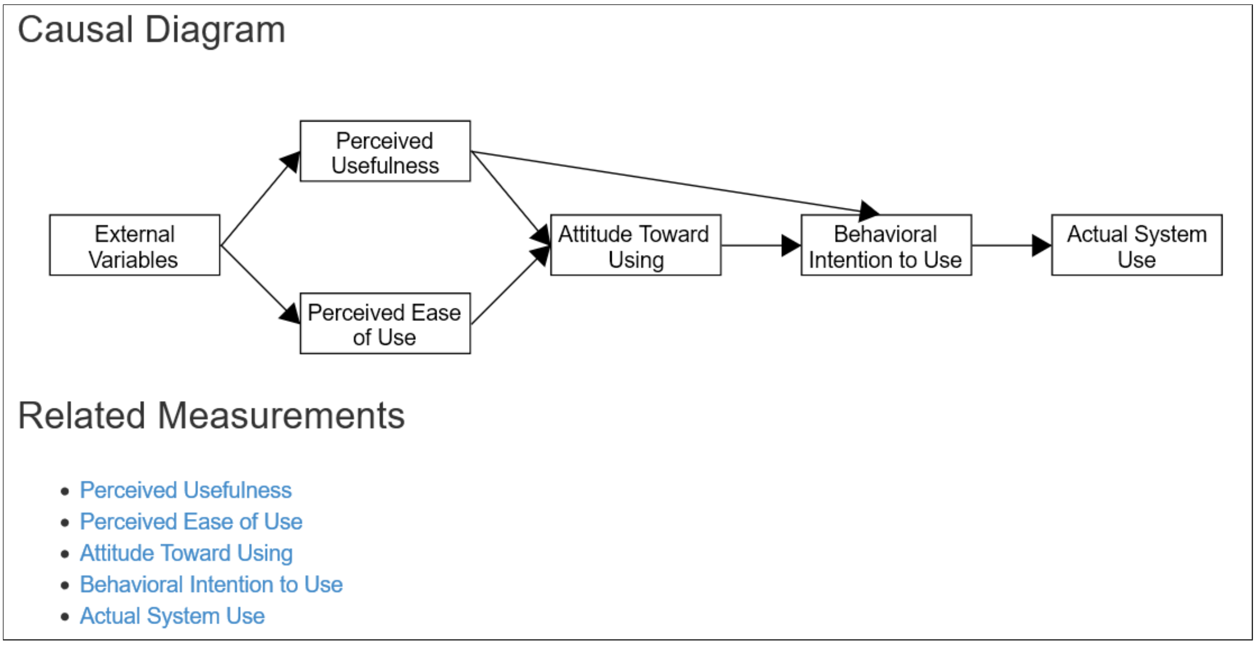


Figure 4.19: Causal Diagram and Related Measurements of the Model in a Magnified View

Publication List Finally, we are having a look at an example of a publication list (figure 4.20).

As one can see, these publications are related to the examples above and they are presented with their very basic information. The structure of this list is similar to the list of publication references on the details pages, because they provide the most basic bibliographic information.

A user toolbox on the right (similar to the one of the model or measurement details view) is currently hidden, but it could provide some additional actions as well. These actions could include adding new data, exporting data or filtering data.

We have to mention that the idea of this system is to link literature to the models and measurements, but we are not planning a full-featured publication server. It is important for basic analyses to be able to identify relevant resources. Further analyses require additional meta-data. Therefore, we allow users to add "Resources", that might link to other systems. Here, the column *Resources* contains either a file ("File"), that has been uploaded to the system, or a link to an external literature database ("LitDB"). These resources could also be DOIs or links with identifiers to concrete databases. Thus, we do not have to rebuild a complicated meta-data structure and one can also link the original publications that might not be stored in this system due to copyright issues.

Nonetheless, we might add some additional meta-data such as publication type, scientific discipline or journal as well, so that simple statistics can be calculated more easily.

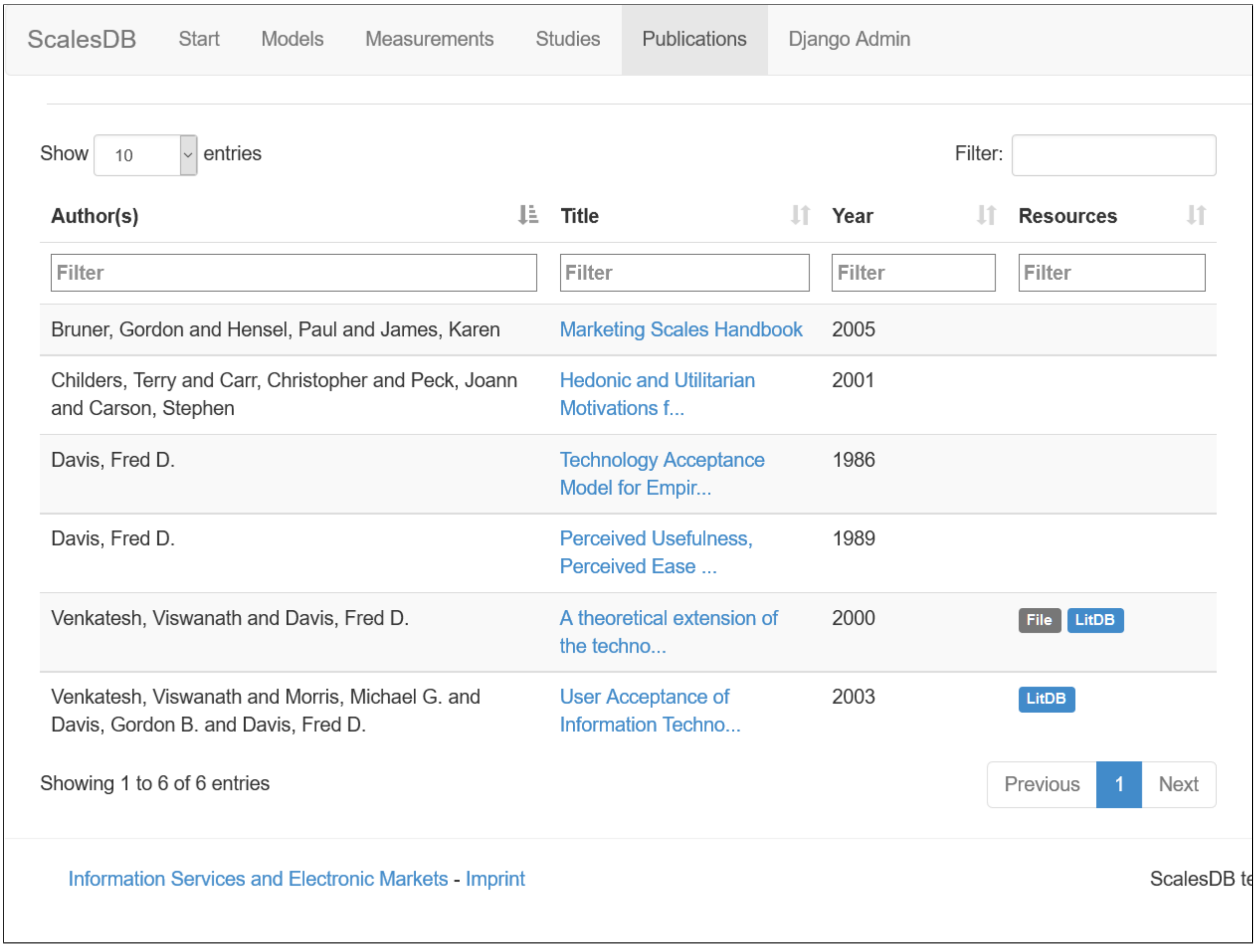


Figure 4.20: Sketch of the Publication List View

4.4 Technical Architecture

After the description of the software architecture, this section covers some of the main aspects of the technical architecture, such as the selection of tools or the general database view.

At first, we are going to have a look at our organizational constraints (4.4.1) that influenced our choice of technologies. Next, we are describing the Django Web Framework, a central component, in greater detail (4.4.2).

Finally, we are defining the domain-specific database layout in subsection 4.4.4 followed by some notes on their implementation (4.4.3) with the Django ORM.

4.4.1 Constraints of the Organizational Environment

As our abstract architecture should be realized in a concrete organizational environment, it is important to relate some of the design decisions to the additional constraints of this environment. This means that our technical infrastructure provides certain options and the technical staff advises us to use software they are able to maintain.

Additionally, the development of such a sophisticated system requires a certain flexibility in the development process in order to add new data. Rapid prototyping techniques (e.g. [Ret94]) and agile development (See [FH01]) can lead to first results that show strengths and weaknesses of the system and they can help to pursue the right adaptations. After the system reaches a certain maturity, one might change some of the components in order to improve performance or switch to specific hardware. Then, however, some restrictions might apply on changing the database or application interfaces, what should not happen on connected services either way, but could slow down the development process in the beginning.

Our service has to run on the environment of the research group and, therefore, on Linux machines. These are run by the Fedora operating system[1], currently in version 23.

In order to integrate the new application with existing ones it is proposed by our technical staff to use the Django web framework (See 4.4.2) that is based on Python. Django does also encourage rapid development as it uses an object-relational mapper (ORM) that allows to create the database with the help of Python objects and to create basic web forms out of the models as well.

Currently we are using Python 2.7 that has maintenance support until 2020, future changes to Python 3.x might be necessary. The Django framework in use has version 1.7.9. Data is stored in a PostgreSQL database of the version 9.4.5.

4.4.2 Django Web Framework

Django is a web application framework that is intended to support reusability of components and the option to add new modules. Besides the ORM it includes a template system for the user interface and a form handler that allows to generate forms from models and validate forms in general. Additionally, it provides a serializer for model instances that can produce XML or JSON and it includes support for Python's built in unit test framework.

[1] `www.fedoraproject.org`

Additionally, there is a *contrib*-package[2], that bundles applications that can be used for typical tasks, such as authentication, data administration or the generation of RSS feeds. This package can also be used to include the application into the existing infrastructure of our organization. For example, support for unified logins (e.g. shibboleth [Shibboleth]) should be added for productive use.

Django follows the DRY-Principle (Don't repeat your code)[3] and makes use of the MVC pattern (model view controller)[4] that allows to separate the data(base) model from routines and the presentation layer.

The Django REST[5] framework [djangoRest] is used for providing an API for the connection with other tools and services.

With the contrib-package, Django provides an admin interface with forms for every model, defined in the ORM. This means that there is no need to write any custom user interface for the first stage (See 4.3.1) where only overview lists and forms for each database table should be provided. The forms do even support one level of nesting, so that some database tables can already be hidden from the user, as described in stage 2 of the user interface development (4.3.2).

4.4.3 Database Realization

In this subsection we are providing a short overview of common database design decisions and their realization with the Django ORM. Based on the description of the core tables in the following section (4.4.4), we will explain their use for this work.

Depending on the database, the constraints set in the ORM are replicated in the database itself or they are validated in the framework itself. This allows to switch between different relational database management systems (DBMS).

In the future there might be reasons for selecting a specific DBMS and leveraging its special features. For rapid prototyping, however, it is convenient to use the database at hand, maybe even just a file like with sqlite.

4.4.3.1 Basic Mapping

As already mentioned, we are using an ORM for the rapid prototype of the proposed system. Thus every class that inherits from the model-base class, will be represented as a database table later on. See, for example, the model of an author in listing 4.1, that contains two fields with chars that have the maximum number of chars 48 and a verbose name that can be shown to human readers. The representation of this model is a single table in the database with three columns: an implicit primary key, the first name and the second name (See listing 4.2).

[2] https://docs.djangoproject.com/en/1.9/ref/contrib/

[3] DRY-Principle: see, for example, [HT00, p. 27]

[4] Model View Controller: see, for example, [Fow02, p. 330]

[5] The REST (Representational State Transfer) architecture has been invented by Fielding [Fie00] and is widely used in distributed web systems.

Listing 4.1: Python example of a model class for an *Author*

```
from django.db import models

class Author(models.Model):
    first_name = models.CharField(max_length=48, blank=True, null=True)
    second_name = models.CharField(max_length=48)
```

Listing 4.2: SQL statement for an *Author* table related to listing 4.1

```
CREATE TABLE "author" (
        "id" INT NOT NULL PRIMARY KEY AUTOINCREMENT,
        "first_name" VARCHAR(48) NULL,
        "second_name" VARCHAR(48) NOT NULL
        )
```

There are various possibilities to adjust the field attributes. In this example, `max_length` defines the maximum length of the string, `blank` states whether the field is allowed to be empty in a form and `null` defines whether the field may be empty in the database. Field values can be limited according to a fixed set of choices or with the help of boundaries in case of a numeric field. It is furthermore possible to define custom primary keys or add functions that validate new entries before storing the result in the database.

Every core element (See 4.2.3.1) is modeled in a separate file for clarity. Typically the core element has to be divided into different database tables, because it holds more than one entry for a certain variable. For example, a publication can have several authors. Therefore, the publication is modeled as a class and the authors are modeled as another class that is linked to the publication (as shown in listing 4.3). The representation in the database are two tables connected with a foreign key (See listing 4.4).

Listing 4.3: Python example of a model class for a *Publication* with several *Authorships*

```
from django.db import models

class Publication(models.Model):
    title = models.CharField(max_length=128)
    year = models.IntegerField(blank=True, null=True)

class Authorship(models.Model):
    first_name = models.CharField(max_length=48, blank=True, null=True)
    second_name = models.CharField(max_length=48)

    order_of_authorship = models.IntegerField(default=1)
    publication = models.ForeignKey(Publication, related_name="authors")

    class Meta:
        unique_together = ('order_of_authorship', 'publication')
```

Additionally, listing 4.3 shows the `unique_together`-constraint that allows the same combination of the order of authorship and the publication foreign key only once in the database. This can be used to assign several authors in a certain order to a publication.

Listing 4.4: SQL statement for a *Publication* table and an *Authorship* table related to listing 4.3

```
CREATE TABLE "publication" (
        "id" INT NOT NULL PRIMARY KEY AUTOINCREMENT
        ,"title" VARCHAR(128) NOT NULL
        ,"year" INT NULL
        )

CREATE TABLE "authorship" (
        "id" INT NOT NULL PRIMARY KEY AUTOINCREMENT
        "first_name" VARCHAR(48) NULL,
        "second_name" VARCHAR(48) NOT NULL,
        "order_of_authorship" INT NOT NULL,
        "publication_id" INT NOT NULL REFERENCES "publication" ("id"),
        UNIQUE (
                "order_of_authorship",
                "publication_id"
                )
        )
```

4.4.3.2 Class and Table Inheritance

Relational databases are generally not designed to support table inheritance in the way class inheritance works in object-oriented software. However, there are different relational representations that allow to represent class hierarchies: Class Table Inheritance, Concrete Table Inheritance and Single Table Inheritance. The main differences of each approach are data duplication and speed.

We are going to describe each of these representations and how we use them in our project with the Django ORM.

Class Table Inheritance With Class Table Inheritance [Fow02, p. 278] every class is represented in the database with its defined fields without inheriting the fields of the super class. This means that the super table contains data for every sub table and the sub tables only contain columns if they define additional ones.

In figure 4.21 this procedure is outlined. On the left one can see the class hierarchy of an object oriented design and on the right the respective mapping is shown. Every table contains exactly the attributes the original class has defined. The table *Domain* contains attributes such as a name or a source, that are used for each type of domain. *NumericDomain* and *VerbalDomain* contain their distinct specific attributes. A foreign key points to the related domain table with its attributes.

This allows one to query the super table and get all objects, whereas one needs joins with the particular sub table if all the details are required. Querying any sub table that includes details, requires a join with the super class. However, data is clearly structured and references to the super table as well as the sub tables are possible.

Advantages of this approach are its simplicity and the normalization of the database. Disadvantages are the necessary joins between tables that cost performance.

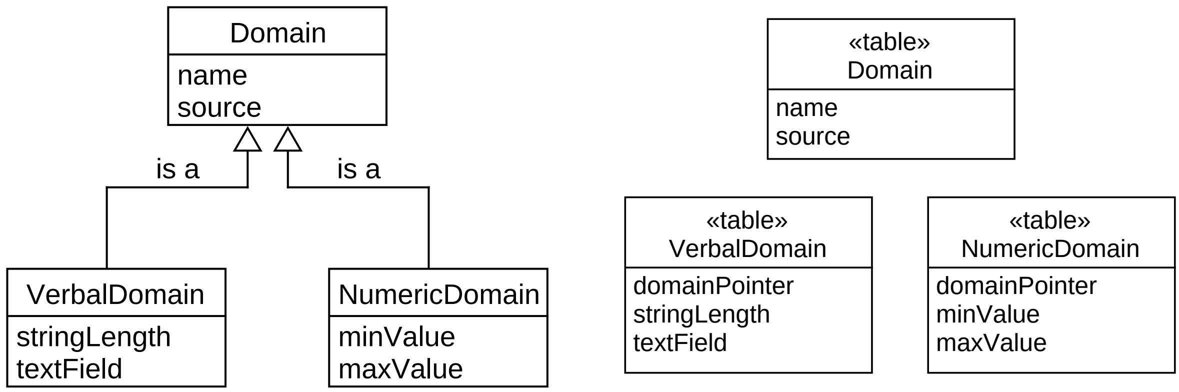


Figure 4.21: Class Table Inheritance / Table Per Class (after [Fow02, p. 278])

Listing 4.5: Python example of a concrete base class *Domain* that is inherited by the classes *NumericDomain* and *VerbalDomain*

```
from django.db import models

class Domain (models.Model):
    name = models.CharField(max_length=20)
    source = models.ForeignKey(Publication, null=True)

class NumericDomain (Domain):
    minValue = models.IntegerField()
    maxValue = models.IntegerField()

class VerbalDomain (Domain):
    stringLength = models.IntegerField()
    textField = models.BooleanField()
```

In Django this pattern is called Multi-Table-Inheritance. See listing 4.5 for the code example of our implementation and listing 4.6 for the corresponding SQL. Every subclass inherits from the superclass *Domain* and not from the generic superclass *models.Model*.

Listing 4.6: SQL statement for table creation of the *Domain* superclass with the subclasses *NumericDomain* and *VerbalDomain* related to listing 4.5

```sql
CREATE TABLE "test_domain" (
        "id" INT NOT NULL PRIMARY KEY AUTOINCREMENT,
        "name" VARCHAR(20) NOT NULL,
        "source_id" INT NULL REFERENCES "test_publication" ("id")
        )

CREATE TABLE "numericdomain" (
        "domain_ptr_id" INT NOT NULL PRIMARY KEY REFERENCES "domain" ("id"),
        "minValue" INT NOT NULL,
        "maxValue" INT NOT NULL
        )

CREATE TABLE "verbaldomain" (
        "domain_ptr_id" INT NOT NULL PRIMARY KEY REFERENCES "domain" ("id"),
        "stringLength" INT NOT NULL,
        "textField" bool NOT NULL
        )
```

Concrete Table Inheritance With Concrete Table Inheritance [Fow02, p. 285] every table contains all related attributes, also the inherited ones. The abstract base class is not mapped to the database, a concrete subclass is required to be mapped to a database table. Subclasses/Tables should not contain semantically similar fields if one wants to avoid possible duplicates.

In figure 4.22 this approach is outlined. On the left one can see the class hierarchy again. In the mapping on the right, each concrete class is mapped to a table with all its attributes. Depending on the mapper, the superclass (Person) might be represented as an additional table as well. However, most often it is designed as an abstract class.

With this approach, queries on a certain element do not require any joins. However, when getting data of all elements, the tables must be joined. Common attributes are not clearly divided from specific attributes.

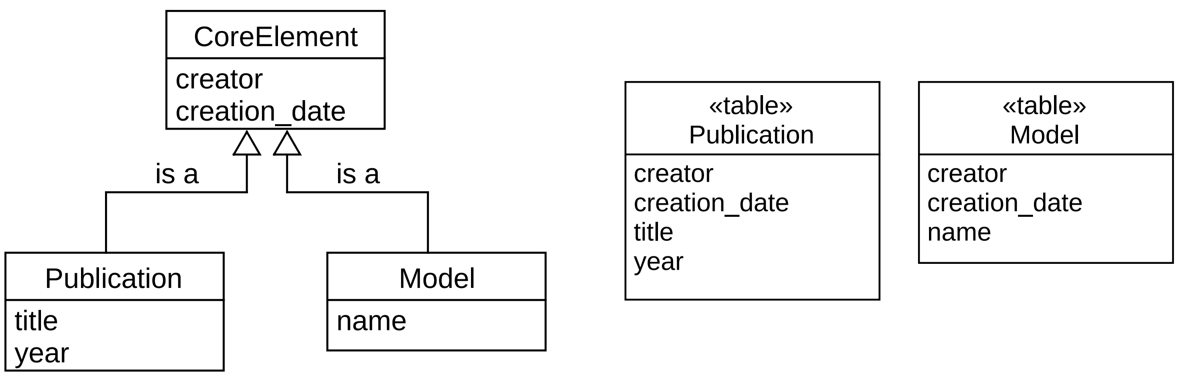


Figure 4.22: Concrete Table Inheritance / Table Per Subclass (after [Fow02, p. 285])

Listing 4.7: Python example of an abstract base class *CoreElement* that provides the attributes of the core elements to the *Publication* model and the scientific *Model* model

```
from django.db import models

class CoreElement(models.Model):
    creator = models.CharField(max_length=48)
    creation_date = models.Date()

                class Meta:
        abstract = True

class Publication(CoreElement):
    title = models.CharField(max_length=128)
    year = models.IntegerField(blank=True, null=True)

class Model(CoreElement):
    name = models.CharField(max_length=128)
```

Listing 4.8: SQL statement for table creation of the core elements *Publication* and *Model* that inherit the fields *creator* and *creation_date* from an abstract base as shown in listing 4.7

```
CREATE TABLE "publication" (
        "id" INT NOT NULL PRIMARY KEY AUTOINCREMENT,
        "creator" VARCHAR(48) NOT NULL,
        "creation_date" DATE NOT NULL,
        "title" VARCHAR(128) NOT NULL,
        "year" INT NULL
        )

        CREATE TABLE "model" (
        "id" INT NOT NULL PRIMARY KEY AUTOINCREMENT,
        "creator" VARCHAR(48) NOT NULL,
        "creation_date" DATE NOT NULL,
        "name" VARCHAR(128) NOT NULL
        )
```

The advantage of this approach is that it generally avoids many joins. It is faster to get single objects. However, disadvantages are the difficulty of key management, because of the missing super table (reference integrity), and changes to the super class or the general hierarchy affect all sub tables. This leads to more effort in maintenance.

In Django this pattern is called Abstract Class Inheritance. See listing 4.7 for a code example of our implementation: We are using this pattern for the addition of a limited set of meta-attributes to core elements, such as the creator or date. The core elements do not have common domain-related attributes so they are different. Queries over all core elements are not as frequent as for the domain-related attributes, so we are saving joins. When doing a query over all meta-attributes, joins would be necessary in every design approach, but they are supported more clearly by the ORM with this design.

Single Table Inheritance With Single Table Inheritance (STI) [Fow02, p. 278] (also called Table-Per-Class-Hierarchy or Table-Per-Hierarchy) the fields of all classes are represented in one single table. If a subclass has additional fields to the superclass, additional columns will be added and set NULL for the other subclasses. Columns should be semantically different and named meaningful in order to select the correct ones.

In figure 4.23 this approach is outlined. Again, one can see the class hierarchy on the left that is mapped to the single table on the right. Adding additional attributes would result in another column. The item type allows to identify the possible attributes for this item.

With this approach no joins are necessary anymore, since all data is stored in one table. However, getting specific elements requires knowledge of the structure or appropriate mappers. Adding database constraints is complicated, since the fields in the subclasses are required to be NULLed as they might not exist in other subclasses.

The biggest advantage of this approach is that it avoids joins for any query and it is easily extendable. Disadvantages are the missing data integrity due to the denormalized database. Column assignment can be confusing to database maintainers later on and depending on the data and database management system, there can be a lot of wasted space.

Django does not provide direct support for this pattern. However, there is a possible workaround where multiple abstract models are all inherited by a large model. Proxies are used for this access control on the attributes that are allowed for this particular class. Many users recommend the extension django-typed-models[6] that adds single table inheritance but is not officially part of the core system. Alternatively, the table can (always) be created in the database and used with custom queries of views from Django.

In our implementation it could be used to store additional attributes for the different types of response items. Unipolar items are usually using an item text and state if they are scored reversely. Bipolar items might add additional text, often two adjectives, that are used as poles. They are seldom queried separately and it can be useful to use the benefits of the faster single table. In order to keep a normalized database structure, we are using class table inheritance instead. Additional attributes can be discriminated better and extensions to the database layout can be done without the need for additional documentation.

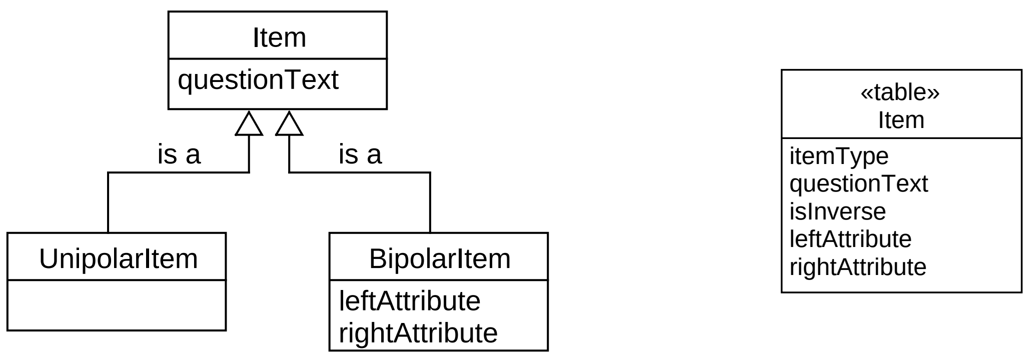


Figure 4.23: Single Table Inheritance/ Table Per Hierarchy (after [Fow02, p. 278])

[6]django-typed-models: https://github.com/craigds/django-typed-models

4.4.4 Database Tables

We are now describing the representation of the core elements and the core relations, described in 4.2.3.

Therefore, we are using entity-relationship diagrams. In figure 4.24 we list the used elements in the diagrams. In particular it is important to note that we have color-coded three different rectangles, each representing a different type of database table. Green tables with a bold and underlined heading represent the core elements as previously defined. Those tables may refer to cyan tables that are independent from any core element, but are only used in the context of this core element. They can be edited separately and are not identified by a core element. Tables in white (or without color) are hidden from the user and are used to represent the information in the relational design. They are linked to a core element or to another discrete table with a foreign key relationship that is required. Thus, they would be deleted, when the linked table is deleted as well.

The diagrams outline the relationships between the different tables. We have cut the number of attributes for each table to reduce complexity.

Arrows between the tables or elements are representing foreign key relationships in the direction of the arrow. So, any instance can be referenced by an unlimited amount of other instances, whereas any single instance can have only a single foreign key reference with the same semantics to another instance.

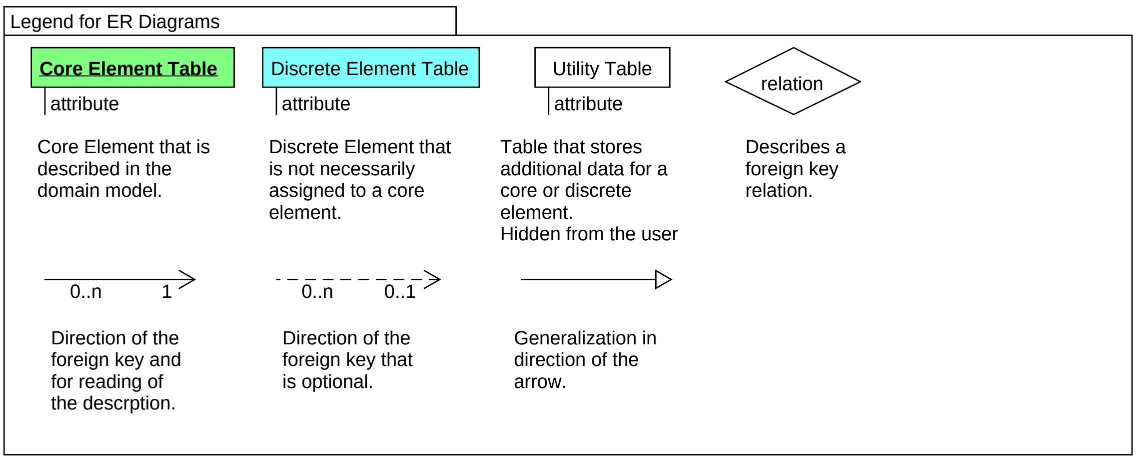


Figure 4.24: Legend for Entity-Relationship Diagrams

4.4.4.1 Publication

See figure 4.25 for an ER-diagram of the publication element. We are going to describe the database layout in the following.

On top we are starting with a table that is called *Publication Reference* and is representing a typed relation with all other elements that are referenced in a publication or referencing something in a publication. The *Publication Relation* is a many-to-many relation that includes attributes on the type of the reference (e.g. is it a primary reference or just mentioning a

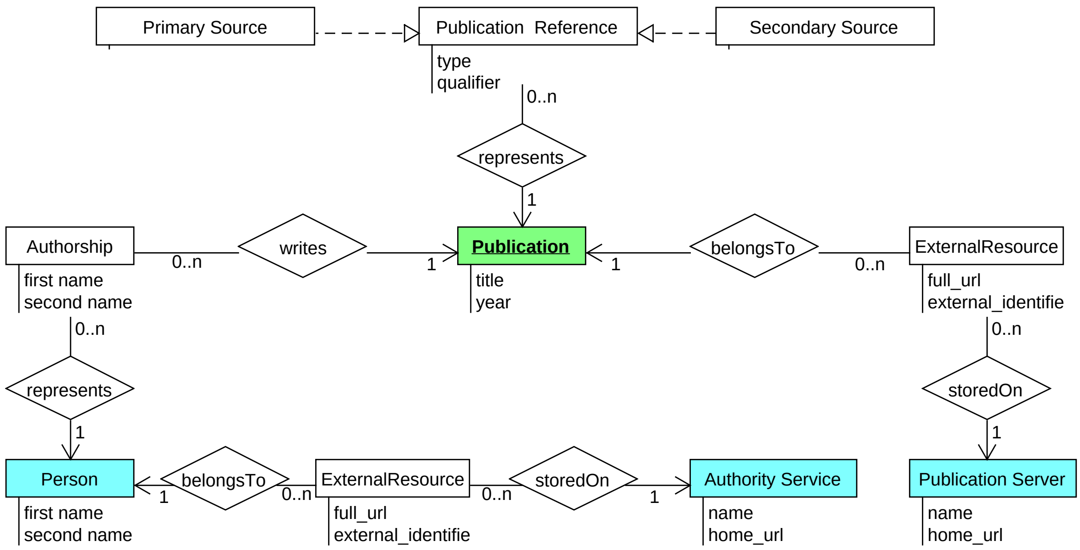


Figure 4.25: EER diagram of the publication layout (with a limited set of attributes)

model/measurement) and contains a qualifier that can be used to identify specific pages or chapters in the publication.

The publication itself can be linked with external resources that provide additional information. Typically one would set a link to the DOI service, to a library system or to the original source.

On the other hand, a publication is often defined by its authors. We are using an Authorship table that includes the information on the authors, as there can be more than one author. In order to be able do identify a unique person, every authorship can be linked with a unique Person table as well. These persons could be identified with the help of external services. An example would be the German Integrated Authority File service (GND - Gemeinsame Normdatei) [Wie15] or even the comprehensive Virtual International Authority File service (VIAF) [Mur12], that is a superset of the German GND. Additionally information about a specific person, however, can also be found on other websites in structured formats (wikidata.org, dbpedia.org) or unstructured formats (personal websites, wikipedia.org).

4.4.4.2 Model

Next, we are having a look at scientific theories and models as shown in figure 4.26. A model is a representation of a scientific theory and can be directly assigned to a primary source, so we are using this publication reference here.

As we have already described in the domain model (See 4.2.1.2), a model contains a description and idea in natural language, but is mainly describing causal relations between concepts. Therefore, we are listing all concept representations that are mentioned in a model (*Concept Representation* and *Concept Membership*). In the case that the model structure is unknown or can not be represented in the database, we can list the included concepts.

The table for causal relations is modeled as a table that represents the edges of a causal graph with the concepts as nodes. So it has some additional attributes for the propagated

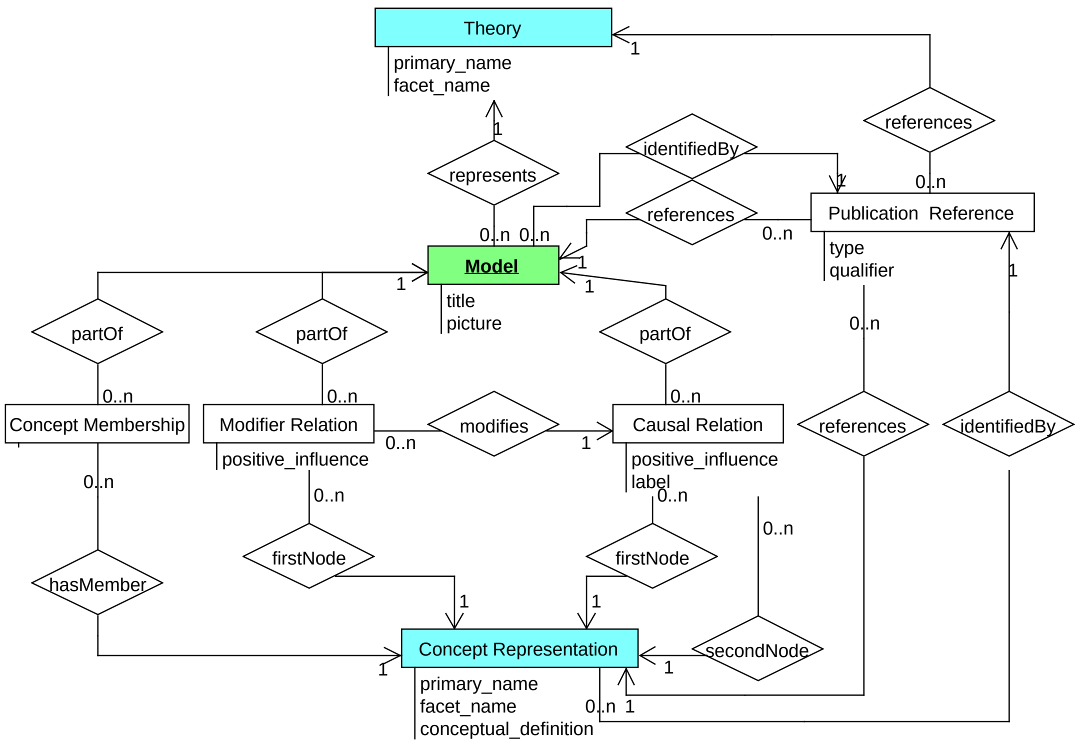


Figure 4.26: EER diagram of the model layout (with a limited set of attributes)

correlation, that can either state a positive influence or a negative influence. In addition, the correlation has optional labels. Each causal relation can be modified by another concept, so they can serve as a node in the Modifier Relations table, that is designed similarly apart from that.

4.4.4.3 Concept

Next is the description of the layout for the concepts. This is outlined in figure 4.27.

As we have already seen, we are using a Concept Representation table in order to store the names and conceptual definition as they might appear in a model description or a measurement description. As there might be different concepts with the same name but different meaning, we are using the concept table for the disambiguation. Concept Representations can then be assigned to Unique Concepts even if they are named differently in models or measurements.

We did not yet find an authority service that is maintaining unique identifiers for concepts as there are for publications or people. However, in the semantic web, concepts are often identified by referencing (disambiguated) Wikipedia articles or URLs that are describing the respective concept. Therefore, the concept can also be linked to external services. An usage example would be to include the description of the Wikipedia article or to improve the article with additional concept descriptions.

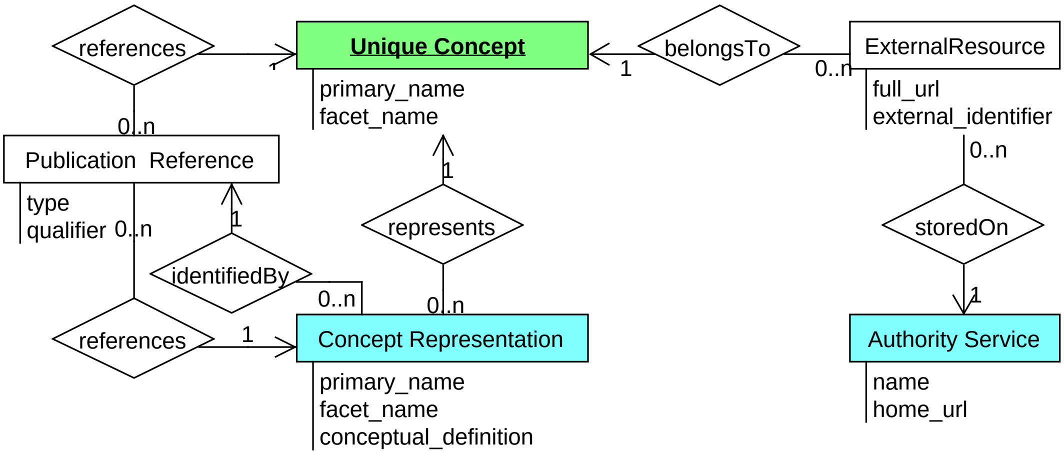


Figure 4.27: EER diagram of the concept layout (with a limited set of attributes)

4.4.4.4 Measurement

Finally, figure 4.28 outlines the attributes and relations for a measurement. A measurement is always operationalizing a certain conceptual representation as shown on the top.

The most important characteristic of a measurement are its items that are shown on the lower right. The items can be assigned to a domain or even a facet. In practice a measurement does not necessarily consist of more than one domain that would represent the measured concept itself. The domain would not need to be divided into facets. The item itself can also be unipolar (e.g. Likert scale) or bipolar (e.g. semantic differential scale).

Furthermore, a measurement is using a certain response domain for the items. The domain can be directly related to the measurement and contain specific answer options or it might be a generic domain, such as a 5-point Likert-style rating scale. The domain can require different response styles, so that there might be categorical responses, numeric responses or verbal responses. In some occasions a domain might even require more than one response style, for example when users should evaluate the characteristics of a certain product and they have to state the importance of these characteristics for themselves at the same time.

Even though the response domain is typically linked to a measurement, it might differ for each dimension, facet or even item. Therefore, we assume that the domain is inherited from measurements to domain to facet and to item. They can, however, state to use a new domain.

The two cyan tables on the left of the diagram are used to maintain a generic set of scoring procedures or scaling methods that can be used as attributes for the measurement. They are, however not predefined in the code but should be extended by the users if necessary. Therefore, they are treated as discrete tables.

To conclude we should not forget to mention the publication references in the upper left. They are used again for the identification of the core tables and discrete tables. They might also add some extra references as well.

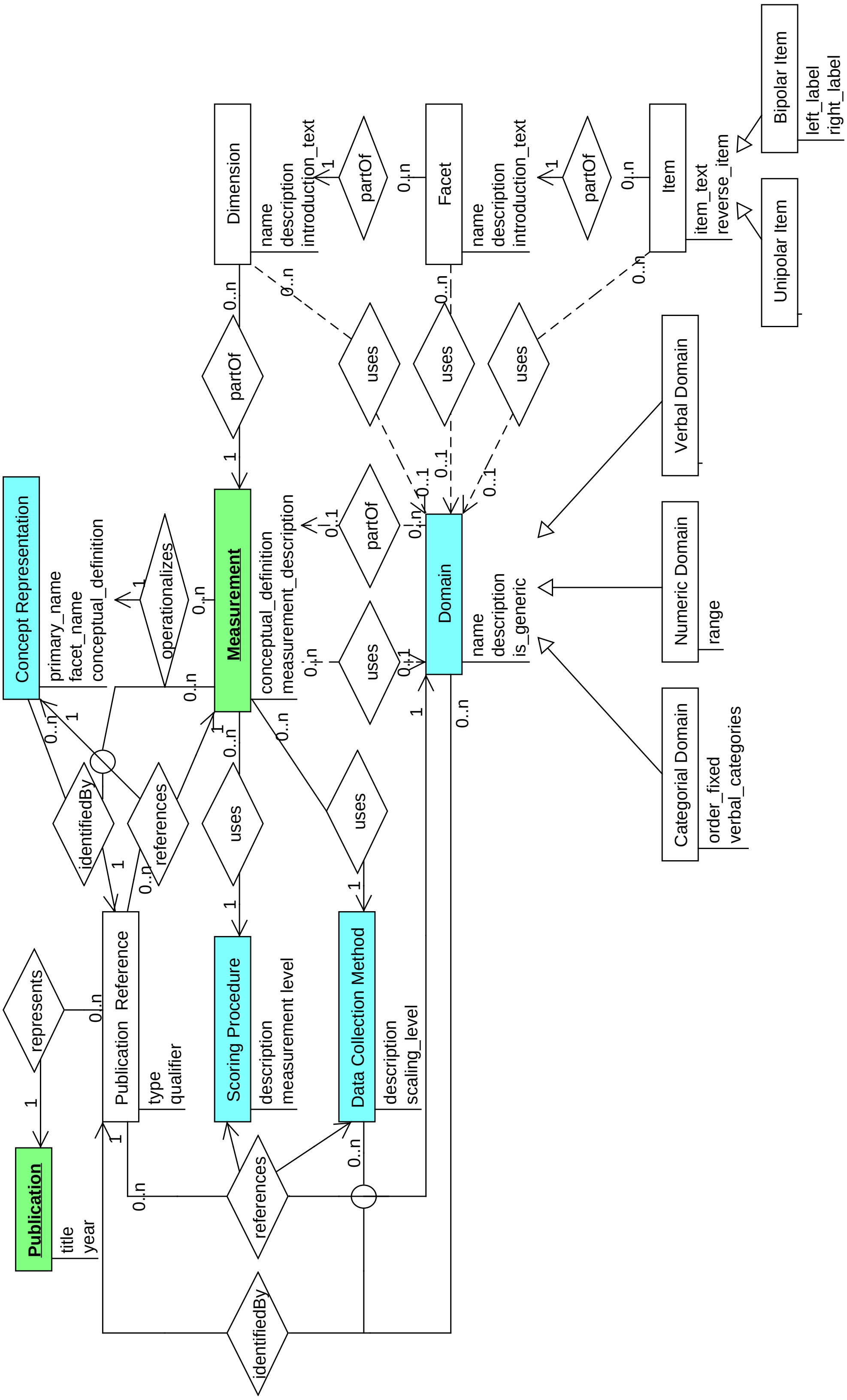


Figure 4.28: EER diagram of the measurement layout (with a limited set of attributes)

5. Limitations and Extensions

The proposed database is intended to act as a central element of a larger infrastructure for model and measurement design, development and meta-analysis.

This chapter discusses some general limitations of the system that are related to the general limits of information technology in science, legal considerations and technical limitations of the implementation. In comparison to external limitations, technical limitations can be eliminated with varying effort in the future.

Additionally, we are going to present possible extensions in the form of new features or additional services. These include improvements of data handling or presentation as well as support for possible analyses or usages of the data. Additional services do not have to be built in the database itself but can also use APIs or exported files.

5.1 Current Limitations

As mentioned before, the limitations of the proposed system can be divided into external limitations that can not be eliminated easily and limitations of the implementation that can be improved by future work or changes in code and tools.

An open question is how much computational intelligence can contribute to innovative research tasks. In subsection 5.1.1 we are pointing out that the proposed system is a tool that supports researchers but one can not (yet?) expect to retrieve scientific results without manual interference.

Legal restrictions, especially copyright laws, are an issue for many information systems and we cover their impact on the proposed database in subsection 5.1.2.

Finally, sections 5.1.3 to 5.1.5 are covering current limitations of the system that might prevent it from being as flexible as planned. Considerations contain the choice of the correct database management system (5.1.3), the proper definition of the API (5.1.4) and the generalization of questions (5.1.5).

5.1.1 Philosophy of Science

Scientific laws can not be verified with a database like this, but they can be confirmed and reassured in different environments with different conditions. The more data has been collected systematically, the more likely one can find general laws, even if it is not possible to state their universal validity.

Furthermore, research is made easier and more efficient in the sense that information can be provided in a structured manner and it is possible to use information technology to analyze this data without the need of manual transcription of papers and books. This allows researchers to focus on the underlying theoretical and practical research problems without dealing with information overload.

Nonetheless, it is up to the researcher not to rely on tools alone just because of their ease of use or the efficiency of work. They have to be aware of the strengths and weaknesses, for example poor data quality of collaborating scientists or missing data that has not been entered yet. Often tools allow to follow certain ideas very well, but often they restrict the users in their creativity for alternative solutions. Therefore, users should be able to analyze the data on their own and extend the infrastructure in the way they like. Especially for statistical analyses or machine learning approaches, it can be difficult to provide generic algorithms that find the optimal parameters on their own and are still interpretable by the researcher.

5.1.2 Copyright

The rights on intellectual property are present in scientific research as well. Therefore, one has to prove carefully, whether one is allowed to store detailed information of scientific work in a database in order to provide a new service.

In most countries there is a certain form of *fair use* law that covers, amongst others, the right to cite parts of a work for “commentary, search engines, criticism, parody, news reporting, research, teaching, library archiving and scholarship” [Wik15b]. Google Books seems to be legit (as least in the US) for archiving purposes, but it does not necessarily have to be true for other databases, especially if you provide additional services. See, for example, the article of Reichman and Okediji [RO12] that discusses the difficulties of copyright laws and their impact on automated knowledge generation.

Depending on the scope of usage, the system might have to provide proper access control or suitable copyright checks. Especially, if a service should not only provide information for a single person or a research group, but the research community or the general public, this is important.

Existing collections of scales, like scale handbooks, usually have the right to reprint the scales published in certain journals. On the one hand this is a good way to populate the database in a legal way. On the other hand it implies that not *all* relevant information can be stored and used, when it is not licensed. This leads to the absurd system that a researcher should cover all relevant literature, but is not allowed to access it or can not afford it. The replication of surveys is also hampered by copyright law if the questionnaires are protected.

5.1.3 Database Management System

We are using the Django O/R-Mapper for modeling the data and, therefore, we are able to use different relational databases. As we already mentioned, it might be good enough for a prototype of such a system, but in the long-term some changes might be necessary. With the help of the mapper it is possible to use MySQL or PostreSQL that both have their benefits. As for the scaling we did not go into details here, since this is strongly dependent on the final environment.

However, one could also argue that the choice of the database system is more crucial. In fact, as Fowler [Fow12] discusses in his article, mainly on O/R-Mappers, one should probably think about graph databases for a system like this. The database links many different models, scales, items or publications with each other and, therefore, can be seen as a single unit that is accessed and they can be easily represented in a graph format. Additionally, new characteristics of e.g. scales could be added more easily. However, one has to keep in mind that relational data often might be easier to use in analytics tools, that often can handle csv files, tables or data.frames quite well.

The database layout does also lack considerations on transaction handling that gets more important when larger modifications take place. Scalability is an issue as well, when deploying the system for the use with larger research groups or even publicly. It is, however, difficult to anticipate all future usage scenarios and their particular bottlenecks.

5.1.4 Specifications of API and Exchange Formats

Finally, we did not provide any specification for our proposed application programming interface in this work. As described in the use cases, our system should be able to export and import data (See 4.1.2.3).

We mentioned that we use the Django Rest Framework in the section on the technical implementation (See 4.4.2) but we did not specify any exchange formats yet.

After the core functionality of the database is running, effort should be put into a consistent definition of exchange formats, that might also include a selection of existing standards. Possibly, it is necessary to develop new standards from scratch.

The use of semantic technologies and ontologies would also help to interrelate the knowledge in this database with existing knowledge as in linked open data in different natural languages. An approach would be the Simple Knowledge Organization System (SKOS) as recommended by W3C [MB09]. Additionally, other existing data providers tend to provide linked open data. An example is the Bibliographic Framework Transition Initiative of the Library of Congress [Bibframe] where the German National Library takes part [Deu14].

5.1.5 Generic Wording

When using measurements in a practical context, they are sometimes referring to specific people or objects, for example in the question text. Comparing questions gets more complicated, if they have to be adapted to new context. Davis [Dav85] used the item “All things considered, my using electronic mail in my job is good/bad” in his TAM model, because he evaluated an electronic mail system. Reusing this item in another context would not alter the idea, but change the label of the indicator item. It is helpful to introduce wildcards like %the system% in order to treat the item as a generic one. This makes text comparison easier and could also be used to insert new context variables for a certain export.

5.2 Possible Extensions

There are dozens of possible extensions to the system at hand. We try to categorize these into different categories:

- Collaboration features (5.2.1) improve the cooperation in a team.
- Included meta-analyses (5.2.2) are extending the usability of the database without the need for additional tools.
- Support of questionnaire generation and the storage of results complete the management of survey-related data (5.2.3).
- Support of multiple languages for international and intercultural studies.

5.2.1 Collaboration Features

As we mentioned before in section 4.2.2, we did not yet implement many features that are required for a proper collaboration.

In particular, one has to think about the different stakeholders of the system and their access rights (5.2.1.1). Besides the rights to view or edit data, it might also be useful to keep track of changes in the system with a version control (5.2.1.2). Lastly, it is useful to provide options to discuss about or comment on specific entries in order to improve data quality (5.2.1.3).

5.2.1.1 User Management and Access Control

As soon as more than one person is working on the same data, it is important to think about user management and access control.

In smaller teams, it might be sufficient to identify who added or altered information, so that one can discuss issues with each other. In larger teams, one might want to restrict editing options or even access to certain information.

For example, there might be copyright issues that require to restrict access to certain publications (See also 5.1.2). Collaboration across institutions that do not want to share some of their unpublished resources is another example. A third example is the collaboration of researchers and students and the requirement that a student's work has to be approved by an adviser.

Open collaboration in the web requires a completely different model of user management and access control. Sites like Wikipedia[1] or MusicBrainz[2] can stand as an example for open collaboration (See also [LC01] about Wikis for Collaboration).

We have refrained from modeling and implementing a full-featured user management, because it highly depends on the usage scenario that is addressed and is not directly related to the scientific goals of this database.

5.2.1.2 Version Control

In addition to access control, it might be useful to track changes and allow roll-backs in case of vandalism or unintended deletions. In the proposed system, data is mainly extracted from papers and won't change continuously. Nonetheless, it might be corrected or extended from time to time.

It is important for research to reference a certain state of information. So they still have to cite the original source, whenever possible. When citing the entry in the proposed database, they have to refer to a given state.

A possible design pattern to implement this, might be memento. The intent is: "Without violating encapsulation, capture and externalize an object's internal state so that the object can be restored to this state later. " [Gam+95]

5.2.1.3 Communication

Finally, a huge benefit from collaborative software is not only to have access to the same data, but also to benefit from the experience of others. Users should be able to discuss entries or to add information that is not domain-knowledge as such.

It might be the case that a user detects a typo in a measurement text, but does not have access to the original sources, so he is not able to verify if this typo happened by chance. Nevertheless, he can write a comment to inform other users about his concern.

A discussion does not only have to deal with errors or problems, but can also include ideas for improvement or additional information for clarification.

5.2.2 Meta-Analysis Support

One of the main goals of the proposed system is to make model and measurement selection and comparison easier and more efficient. Regardless of the reasons for such a comparison, the researchers can be supported by functionality that saves them time and can do some automated analyses.

Even though the data can be exported and evaluated by an external tool, this always requires additional effort and knowledge of the users.

[1]Wikipedia: `https://www.wikipedia.org/`
[2]MusicBrainz: `https://musicbrainz.org/`

5.2.2.1 Manual Comparisons

A basic approach for comparison is to view the data side-by-side.

The researcher might be interested in the differences between two or three measurements, so he benefits from a view that lists the attributes of each measurement in a comparison table. The system can indicate important attributes that are similar or different from each other.

Concerning causal models, it is easier to do a visual comparison of causal graphs, than to compare the raw entries of nodes and edges. Therefore a visualization approach should be developed, where different graphs can be compared with each other. Especially when publishing meta-analyses it can be easier for others to interpret a good visualization.

Measurements or models might also be mentioned in the same publications, for example in survey papers or when they have been evaluated as alternatives. Presenting those publications to the user are helping to identify existing scientific work and can reduce unnecessary replication.

5.2.2.2 Automated Detection of Related Entries

Manual comparisons still require a lot of work by the researcher. Especially searching for related models or measurements takes a lot of time. Approaches from cluster analysis or recommender systems can propose possible relations.

With the help of distance metrics or similarity metrics, an algorithm is able to calculate similarities of entries. As an example, measurements are similar if they include similar questions and domains. This can be used to calculate distances to the other measurements in the system, even if they are not linked by an explicit relation. Models are similar if they connect the same concepts with each other and maybe also use the same measurements for them.

In addition to analyzing the content of the entries, one can also look at the user behavior. Similar to recommender systems for libraries, one can analyze click-streams and assume that entries that are viewed in the same session are related (See e.g. [GNT03]).

There are many options to enhance the automated detection of similarities. Clustering techniques can detect similar sets of entries, natural language processing can be used to enhance comparison of the texts and machine learning techniques in general can be used for specific tasks.

If possible, the proposed approaches should be integrated into the normal workflow of the user. When they are looking for a model, there should be an information box or a section that points him to related models, so that he does not have to care about the algorithms in the background. Knowledge of the algorithms or adaption of the algorithms is only necessary for larger meta-studies that require a documentation and evaluation of the learning process.

5.2.3 Survey Handling

In addition to the possibilities of meta-analyses and the digital administration of models and measurements, the main goal is their use in surveys. Functionality for the automated generation of questionnaires and their analysis brings a huge improvement into survey-research. Storing the results of surveys in combination with the meta-data of the used models and measurements provides many opportunities to compare studies, instruments or the models across time and context.

5.2.3.1 Questionnaire Creation

In the context of empirical research, most people use some kind of survey tools that generate questionnaires or run online surveys. It would be a simplification if they could just use the items of the database and load them into the survey tools. It would also reduce transcription or copy-and-paste errors and might be eminently useful for more laborious questions like pairwise comparisons, that could be easily automated.

Going into detail, questionnaire creation has many more facets that could be supported. The structure and the design of the whole questionnaire and every single question influences participants. There are many options that can be thought of.

5.2.3.2 Analysis of Survey Results

Given the case that a survey has been created with the help of existing models and measurements, there exist common analysis techniques. Examples would be to test the internal structure of a construct or to perform a regression based on the proposed causal model. Survey tools typically do not store any meta-information about the items of the questionnaire and leave it to the user to keep track of it.

Providing the meta-data as addition to the raw survey results, allows to automate data analysis. It is not necessary that the data analysis is done with the proposed system, but it should be possible to export at least a meta-data file such as a code plan to a statistical tool.

Nonetheless, it would facilitate survey analysis a lot, if the database was extended by a service that performs automated analyses.

5.2.3.3 Storage of Survey Results

Finally, the results of a survey can be stored in the database as well. With the help of automated analyses, the relevant validity measures can be calculated automatically and improve the information on models or measurements. This adds a lot of options to evaluate models or measurements compared to the current situation, in that it is hard to get access to the raw data of a study.

It is possible to reuse data and conduct meta-studies without the need for new cost- and time-expensive studies.

Storing survey results does not only require to save results according to the existing models and measurements, but has to cover demographic information as well as additional questions that might not be part of a model or construct.

6. Conclusion

In “As We May Think” [Bus45] foresees a future world in that we step towards “the massive task of making more accessible a bewildering store of knowledge.“ Vannevar Bush has intelligent machines in mind, but he does not refer to or even knows anything about the latest findings in machine learning nowadays or even computers in general.

Today the predicted bewildering store of knowledge is all around us and we have to find solutions that allow researchers to find the necessary information fast, reliably and easily. In the context of survey research in the social sciences, computer sciences or economics, we proposed a database that should allow better insights into scales and their usages. Till now, most of the resources are not machine-readable as they are contained in printed works or in a digital format like PDF without sufficient semantic annotations.

With the help of the proposed database and an analysis infrastructure built upon that, it should be easier for researchers to get a better access to this knowledge. Researchers are able to find relevant scales faster, are supported in survey creation and survey analysis. Moreover, there are even more possibilities of meta-analysis that allow to evaluate the reuse of instruments or to track the changes in the instruments over time.

This does not lead to universal laws as such, but it gives researchers the opportunity to use scales and models that have been tested on larger samples, in more environments and by more fellow scientists than ever before. It would be easier to identify hypotheses that hold in several contexts and design new quality measures to identify them.

Easier tools and better insights would also allow students and junior researchers to benefit from these tools. In the sense of blue collar science [Gey+12], students writing their bachelor’s theses would be able to profit from the existing scales and could generate results of higher quality, without the need for time-consuming literature research. Groups of students could create meta-models of scales on a certain topic, while everybody just has to read a part of the papers and pass the contents into the structured database. Therefore, the students and science in general would benefit.

After this thesis, the implementation has to be extended to a state, where it provides adequate usability and passes several pre-tests. Next, it has to reach the necessary publicity amongst other researchers, because it might not be possible to identify, access and transcribe all the necessary resources due to rights or costs. Thirdly, license problems may arise and prohibit public access to the scales, so one has to find solutions with the content providers. Lastly, but probably even more important, is the difficulty of accessing survey results data. With that, the calculation of quality measures and the comparison of the scale in different studies would benefit a lot.

As a conclusion one might say, that this approach is a follow up of defining important characteristics of scales. However, follow-up research has to be done in order to improve the data management as well as to create algorithms that actually work on the data. Following this field of research, new technologies could be introduced into survey research. This would finally prove some of the criticism of Link [Lin11] wrong.

Bibliography

[AF80] Ajzen, Icek and Fishbein, Martin. *Understanding Attitudes and Predicting Social Behavior*. Englewood Cliffs, NJ: Prentice-Hall, 1980.

[Ass15] Associated Press. *Coca-Cola discloses it spent 119m dollars on health research over five years*. Sept. 2015. URL: http://www.theguardian.com/business/2015/sep/22/coca-cola-discloses-health-research-funding (visited on 2015-11-12).

[Bag80] Bagozzi, Richard P. *Causal Models in Marketing*. Theories in Marketing Series. New York: John Wiley & Sons, Inc., 1980.

[BFH07] Bishop, Yvonne M. M., Fienberg, Stephen E., and Holland, Paul W. *Discrete Multivariate Analysis Theory and Practice*. New York: Springer, 2007.

[BH13] Bruner, Gordon C and Hensel, Paul J. *Marketing Scales Handbook*. English. Vol. 7. Carbondale, Ill.: GCBII Productions, 2013.

[BHJ05] Bruner C., Gordon, Hensel, Paul J., and James E., Karen. *Marketing Scales Handbook: A Compilation of Multi-Item Measures for Consumer Behavior & Advertising*. Vol. 4. Chicago, IL: American Marketing Association, 2005.

[Bibframe] Bibliographic Framework Initiative (Library of Congress). *BIBFRAME - Bibliographic Framework Initiative*. URL: http://www.loc.gov/bibframe/ (visited on 2015-11-15).

[Bla04] Blau, Judith R., ed. *The Blackwell Companion to Sociology*. Blackwell Companions to Sociology. Malden, MA: Blackwell Pub, 2004.

[BNH11] Bearden, William O., Netemeyer, Richard G., and Haws, Kelly L. *Handbook of Marketing Scales. Multi-item Measures for Marketing and Consumer Behavior Research*. 3rd ed. Association for Consumer Research Series. Thousand Oaks: Sage, 2011.

[Bru15] Bruner, Gordon C. *Marketing Scales*. 2015. URL: https://www.marketingscales.com/ (visited on 2015-11-05).

[Buros] Buros Center for Testing (University of Nebraska–Lincoln). *Test Reviews & Information*. URL: http://buros.org/test-reviews-information (visited on 2015-11-15).

[Bus45] Bush, Vannevar. As We May Think. *The Atlantic* (July 1945 Issue), July 1945, pp. 112–124.

[Cal68] Calandra, Alexander. Angels on a Pin. *Saturday Review* 51, 1968, p. 60.

[CGJ14] Carlson, Janet F, Geisinger, Kurt F, and Jonson, Jessica L. *The Nineteenth Mental Measurements Yearbook*. Lincoln, Neb.: The Buros Center for Testing, 2014.

[CM78] Cavano, Joseph P. and McCall, James A. A Framework for the Measurement of Software Quality. In: *Proceedings of the Software Quality Assurance Workshop on Functional and Performance Issues*. New York, NY, USA: ACM, 1978, pp. 133–139.

[Coc12] Coca-Cola Conversations Staff. *Coca-Cola fans share their stories*. Nov. 2012. URL: `http://www.coca-colacompany.com/stories/coca-cola-stories-new-coke` (visited on 2015-11-12).

[CP84] Churchill Jr., Gilbert A. and Peter, J. Paul. Research Design Effects on the Reliability of Rating Scales: A Meta-Analysis. *Journal of Marketing Research* 21 (4), Nov. 1984, pp. 360–375. DOI: `10.2307/3151463`.

[Cro51] Cronbach, Lee J. Coefficient alpha and the internal structure of tests. *Psychometrika* 16 (3), July 1951, pp. 297–334. DOI: `10.1007/BF02310555`.

[Dab96] Dabholkar, Pratibha A. Consumer Evaluations of New Technology-Based Self-Service Options: An Investigation of Alternative Models of Service Quality. *International Journal of Research in Marketing* 13 (1), 1996, pp. 29–51.

[Dav85] Davis, Fred D. A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results. Thesis (Ph. D.)–Massachusetts Institute of Technology, Sloan School of Management, 1986. PhD thesis. Massachusetts Institute of Technology, 1985.

[Dav89] Davis, Fred D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of lnformation Technology. *MIS Quarterly* 13 (3), Sept. 1989, pp. 319–340.

[Deu13] Deutsche Forschungsgemeinschaft, ed. *Sicherung guter wissenschaftlicher Praxis – Safeguarding Good Scientific Practice*. Erg. Aufl. Weinheim: Wiley-VCH, 2013.

[Deu14] Deutsche Nationalbibliothek. *Standardisierung*. Sept. 2014. URL: `http://www.dnb.de/DE/Standardisierung/standardisierung_node.html` (visited on 2015-11-15).

[DIN 2330:1993] DIN-Normenausschuss Terminologie (NAT). *Begriffe und Benennungen; Allgemeine Grundsätze*. Standard DIN 2330. DIN Deutsches Institut für Normung e. V., Dec. 1993.

[DIN 2330:2013] DIN-Normenausschuss Terminologie (NAT). *Begriffe und Benennungen - Allgemeine Grundsätze*. Standard DIN 2330. DIN Deutsches Institut für Normung e. V., July 2013.

[djangoRest] *Django REST framework*. URL: `http://www.django-rest-framework.org/` (visited on 2015-11-15).

[DPLA] Digital Public Library of America, ed. *Digital Public Library of America*. URL: `http://dp.la/info/` (visited on 2015-11-14).

[EF71] Ekman, Paul and Friesen, Wallace V. Constants Across Cultures in the Face and Emotion. *Journal of Personality and Social Psychology* 17 (2), 1971, pp. 124–129. DOI: 10.1037/h0030377.

[FA75] Fishbein, Martin and Ajzen, Icek. *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research.* Reading, Mass.: Addidon Wesley, 1975.

[FH01] Fowler, Martin and Highsmith, Jim. The Agile Manifesto. *Software Development* 9 (8), 2001, pp. 28–35.

[Fie00] Fielding, Roy Thomas. Architectural Styles and the Design of Network-based Software Architectures. PhD thesis. University of California, Irvine, 2000.

[Fow02] Fowler, Martin. *Patterns of Enterprise Application Architecture.* Amsterdam: Addison-Wesley Longman, Nov. 2002.

[Fow12] Fowler, Martin. *bliki: OrmHate.* May 2012. URL: http://martinfowler.com/bliki/OrmHate.html (visited on 2015-11-03).

[Fra03] Fransella, Fay, ed. *International Handbook of Personal Construct Psychology.* Chichester, West Sussex, England: J. Wiley & Sons, 2003.

[Gam+95] Gamma, Erich, Helm, Richard, Johnson, Ralph, and Vlissides, John. *Design Patterns: Elements of Reusable Object-Oriented Software.* Addison-Wesley Professional Computing Series. Reading: Addison-Wesley Publishing Company, 1995.

[Gey+12] Geyer-Schulz, Andreas, Meyer-Waarden, Lars, Kunze, Jonas, and Sonnenbichler, Andreas. *Blue Collar Science: The Student as a (Marketing) Scientist.* Tech. rep. em-kit-2012-12-07. Karlsruhe: Informationsdienste und elektronische Märkte, KIT, Dec. 2012.

[Gib02] Gibbs, J. W. *Elementary Principles in Statistical Mechanics.* New Haven: Yale University Press, 1902.

[GNT03] Geyer-Schulz, Andreas, Neumann, Andreas, and Thede, Anke. An Architecture for Behavior-Based Library Recommender Systems. *Information Technology and Libraries* 22 (4), 2003, pp. 165–174.

[Gol12] Gold, Virginia. *Judea Pearl Wins ACM A.M. Turing Award for Contributions that Transformed Artificial Intelligence.* Mar. 2012. URL: http://www.acm.org/press-room/news-releases/2012/turing-award-11 (visited on 2015-11-02).

[GT95] Goodhue, Dale and Thompson, Ronald. Task-Technology Fit and Individual Performance. *Management Information Systems Quarterly* 19 (2), Dec. 1995.

[Hem52] Hempel, Carl Gustav. Fundamentals of Concept Formation in Empirical Science. *International Encyclopedia of Unified Science.* Foundations of the Unity of Science 2, 1952.

[Hem68] Hempel, Carl G. *Aspects of Scientific Explanation and other Essays in the Philosophy of Science.* 2. print. New York: Free Press, 1968.

[Hem77] Hempel, Carl G. *Aspekte wissenschaftlicher Erklärung*. Ed. by Posner, Roland. Grundlagen der Kommunikation. Berlin/New York: Walter de Gruyter, 1977.

[Hen15] Hentschke, Jana. *Deutsche Nationalbibliothek - Linked Data Service*. Nov. 2015. URL: `http://www.dnb.de/DE/Service/DigitaleDienste/LinkedData/linkeddata_node.html` (visited on 2015-11-14).

[HJB97] Hogan, R., Johnson, J., and Briggs, S. *Handbook of Personality Psychology*. San Diego: Academic Press, 1997.

[HL70] Hall, Calvin S. and Lindzey, Gardner. *Theories of Personality*. 2nd ed. New York: Wiley, 1970.

[HMS59] Herzberg, Frederick, Mausner, Bernard, and Snyderman, Barbara. *The Motivation to Work (2nd ed.)* Vol. xv. Oxford, England: John Wiley, 1959.

[HO48] Hempel, Carl G. and Oppenheim, Paul. Studies in the Logic of Explanation. *Philosophy of Science* 15 (2), Apr. 1948, pp. 135–175.

[HT00] Hunt, Andrew and Thomas, David. *The Pragmatic Programmer: From Journeyman to Master*. Reading, Mass: Addison-Wesley, 2000.

[Ins15] Institute for Comparative Survey Research. *World Value Survey: What We Do*. 2015. URL: `http://www.worldvaluessurvey.org/WVSContents.jsp?CMSID=WhatWeDo` (visited on 2015-11-10).

[ISO 10241-1:2011] ISO/TC 37 Terminology and language and content resources. *Terminological entries in standards - Part 1: General requirements and examples of presentation*. Standard ISO 10241-1. ISO International Organization for Standardization, Apr. 2011.

[ISO 10241-2:2012] ISO/TC 37 Terminology and language and content resources. *Terminological entries in standards - Part 2: Adoption of standardized terminological entries*. Standard ISO 10241-2. ISO International Organization for Standardization, Aug. 2012.

[ISO 1087-1:2000] ISO/TC 37 Terminology and language and content resources. *Terminology work - Vocabulary - Part 1: Theory and application*. Standard ISO 1087-1. ISO International Organization for Standardization, Oct. 2000.

[ISO 1087-2:2000] ISO/TC 37 Terminology and language and content resources. *Terminology work - Vocabulary - Part 2: Computer applications*. Standard ISO 1087-2. ISO International Organization for Standardization, Mar. 1, 2000.

[ISO 12620:1999] ISO/TC 37 Terminology and language and content resources. *Computer applications in terminology - Data categories*. Standard ISO 12620. ISO International Organization for Standardization, Oct. 1999.

[ISO 12620:2009] ISO/TC 37 Terminology and language and content resources. *Terminology and other language and content resources - Specification of data categories and management of a Data Category Registry for language resources*. Standard ISO 12620. ISO International Organization for Standardization, Dec. 2009.

[ISO 25000:2014] ISO/IEC. *Systems and Software Engineering – Systems and Software Quality Requirements and Evaluation (SQuaRE) – Guide to SQuaRE.* Standard ISO 25000. International Standardization Organisation/International Electrotechnical Commission, 2014.

[ISO 704:2000] ISO/TC 37 Terminology and language and content resources. *Terminology work - Principles and methods.* Standard ISO 704. ISO International Organization for Standardization, Nov. 2000.

[ISO 704:2009] ISO/TC 37 Terminology and language and content resources. *Terminology work - Principles and methods.* Standard ISO 704. ISO International Organization for Standardization, Nov. 2009.

[ISO 860:2007] ISO/TC 37 Terminology and language and content resources. *Terminology work - Harmonization of concepts and terms.* Standard ISO 860. ISO International Organization for Standardization, Nov. 2007.

[ISO 9126-4:2001] ISO/IEC. *Software Engineering - Product Quality - Part 4: Quality in use metrics.* Standard ISO 9126-4. International Standardization Organisation/International Electrotechnical Commission, 2001.

[Ken38] Kendall, M. G. A New Measure of Rank Correlation. *Biometrika* 30 (1/2), 1938, pp. 81–93. DOI: `10.2307/2332226`.

[Kok07] Kokoszka, Andrzej. *States of Consciousness: Models for Psychology and Psychotherapy.* Emotions, personality, and psychotherapy. New York: Springer, 2007.

[Kra+71] Krantz, David H., Luce, R. Duncan, Suppes, Patrick, and Tversky, Amos. *Additive and Polynomial Representations.* Vol. I. Foundations of Measurement. San Diego, CA: Academic Press, 1971. 583 pp.

[Lak02] Lakatos Award Team. *2001 Lakatos Award.* 2002. URL: `http://www.lse.ac.uk/philosophy/blog/2002/04/09/2001-lakatos-award-judea-pearl/` (visited on 2015-11-02).

[Lar+10] Larsen, Kai R., Lee, Jintae, Li, Jingjing, and Bong, Chih How. A Transdisciplinary Approach to Construct Search and Integration. In: *AMCIS.* 2010, p. 524.

[Lar+15] Larsen, K. R., Allen, G., Vance, A., and Eargle, D., eds. *Theories Used in IS Research Wiki.* 2015. URL: `http://is.theorizeit.org/wiki/Main_Page` (visited on 2015-11-05).

[Lar10] Larsen, Kai R. *Kai R. Larsen: Research In Progress.* 2010. URL: `http://leeds-faculty.colorado.edu/larsenk/research.htm` (visited on 2015-11-05).

[Lar15] Larsen, Kai R., ed. *Inter-Nomological Network - Constructing Theories. Integrating Research.* 2015. URL: `http://inn.theorizeit.org/` (visited on 2015-11-14).

[LC01] Leuf, Bo and Cunningham, Ward. *The Wiki Way: Quick Collaboration on the Web.* Boston: Addison-Wesley, 2001.

[Leh+15] Lehman, Philipp, Kime, Philip, Boruvka, Audrey, and Wright, Joseph. *The Biblatex Package.* Version 3.1. Oct. 2015.

[Lew86] Lewis, David K. Causal Explanation. In: *Philosophical Papers*. Volume II. New York: Oxford University Press, 1986. Chap. 22, pp. 214–240.

[LH12] Larsen, Kai R. and Hovorka, Dirk S. Developing Interfield Nomological Nets. In: *System Science (HICSS), 2012 45th Hawaii International Conference on*. IEEE, 2012, pp. 5194–5203.

[Lik32] Likert, Rensis. A Technique for the Measurement of Attitudes. *Archives of Psychology* 22 (140), 1932, pp. 5–55.

[Lin+11] Ling, Loh Wai, Downe, A.G., Ahmad, W.F.W., and Lai, Teh Thian. Determinants of Computer Usage Among Educators: A Comparison Between the UTAUT and TAM Models. In: *National Postgraduate Conference (NPC), 2011*. 2011, pp. 1–6.

[Lin11] Link, Michael W. Discussion: Why Aren't Survey Researchers Better at Leveraging New Technologies? In: *Proceedings of the Tenth Conference on Health Survey Research Methods 2011*. Vol. 10. Hyattsville, Maryland: National Center for Health Statistics, 2011, pp. 275–282.

[LL13] Li, Jingjing and Larsen, Kai. Tracking Behavioral Construct Use through Citations: A Relation Extraction Approach. In: *Thirty Fourth International Conference on Information Systems*. Conference on Information Systems. Milan, 2013.

[LS03] Lopez, Shane J. and Snyder, C. R., eds. *Positive Psychological Assessment: A Handbook of Models and Measures*. Washington: American Psychological Association, 2003.

[Luc+90] Luce, Duncan, Krantz, David H., Suppes, Patrick, and Tversky, Amos. *Representation, Axiomatization, and Invariance*. Vol. III. Foundations of Measurement. San Diego, CA: Academic Press, 1990. 355 pp.

[MB09] Miles, Alistair and Bechhofer, Sean, eds. *SKOS Simple Knowledge Organization System Reference (W3C Recommendation)*. Aug. 2009. URL: `http://www.w3.org/TR/skos-reference/` (visited on 2015-11-04).

[MC07] McCrae, Robert R. and Costa Jr., Paul T. Brief Versions of the NEO-PI-3. *Journal of Individual Differences* 28 (3), 2007, pp. 116–128. DOI: `10.1027/1614-0001.28.3.116`.

[McD00] McDermott, Richard. Why Information Technology Inspired But Cannot Deliver Knowledge Management. *Knowledge and Communities* 41 (4), 2000, pp. 21–35.

[Mur12] Murphy, Bob. *Virtual International Authority File service transitions to OCLC; contributing institutions continue to shape direction through VIAF Council*. en. Apr. 2012. URL: `http://www.oclc.org/news/releases/2012/201224.en.html` (visited on 2015-11-14).

[OALD05] Hornby, A. S., Wehmeier, Sally, McIntosh, Colin, Turnbull, Joanna, and Ashby, Michael. *Oxford Advanced Learner's Dictionary of Current English*. eng. New ed., 7. ed. Oxford: Oxford Univ. Press, 2005.

[Osg52] Osgood, Charles E. The Nature and Measurement of Meaning. *Psychological Bulletin* 49, 1952.

[Par00] Parasuraman, A. Technology Readiness Index (TRI) : A Multiple-Item Scale to Measure Readiness to Embrace New Technologies. *Journal of Service Research* 2 (4), 2000, pp. 307–320.

[PC97] Plutchik, Robert Ed and Conte, Hope R. *Circumplex Models of Personality and Emotions.* Washington, DC: American Psychological Association, 1997. DOI: 10.1037/10261-000.

[Pea09] Pearl, Judea. *Causality: Models, Reasoning, and Inference.* 2nd ed. Cambridge: Cambridge University Press, 2009.

[Pet00] Peterson, Robert A. *Constructing Effective Questionnaires.* Thousand Oaks: Sage Publications, 2000.

[Pop35] Popper, Karl R. *Logik der Forschung.* German. Wien: J. Springer, 1935.

[PsycTests] *PsycTESTS®*. URL: http://www.apa.org/pubs/databases/psyctests/ (visited on 2014-12-05).

[PZB88] Parasuraman, A., Zeithaml, Valerie A., and Berry, Leonard L. SERVQUAL: Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality. *Journal of Retailing* 64 (1), 1988, pp. 12–40.

[RegSc] *Registry of Scales and Measures: Psychological Tests, Scales, Questionnaires and Checklists.* URL: http://www.scalesandmeasures.net/ (visited on 2014-12-05).

[Ret94] Rettig, Marc. Prototyping for Tiny Fingers. *Communications of the ACM* 37 (4), Apr. 1994, pp. 21–27.

[RO12] Reichman, Jerome and Okediji, Ruth. When Copyright Law and Science Collide: Empowering Digitally Integrated Research Methods on a Global Scale. *Minnesota Law Review*, Jan. 2012, pp. 1362–1480.

[Rob+91] Robinson, John P., Shaver, Phillip R., Wrightsman, Lawrence S., and Andrews, Frank M. *Measures of Personality and Social Psychological Attitudes.* Measures of social psychological attitudes v. 1. San Diego: Academic Press, 1991.

[Rob85] Roberts, Fred S. *Measurement Theory with Applications to Decision-making, Utility and the Social Sciences.* Ed. by Rota, Gian-Carlo. Vol. 7. Encyclopedia of Mathematics and its Applications 7. Cambridge: Cambridge Univ. Press, 1985, p. 442. 420 pp.

[RSW99] Robinson, John P., Shaver, Phillip R., and Wraghtsman, Lawrence S. *Measures of Political Attitudes.* San Diego, CA: Academic Press, 1999.

[SA05] Smith, Scott M. and Albaum, Gerald S. *Fundamentals of Marketing Research.* Vol. 1st edition. Thousand Oaks, Calif: Sage, 2005. 881 pp.

[Sal71] Salmon, Wesley C. Statistical Explanation. In: *Statistical Explanation and Statistical Relevance.* Pitt paperback 69. Pittsburgh: University of Pittsburgh Press, 1971, pp. 29–87.

[San06] Santor, Darcy A. REJOINDER: New Coke, Rosetta Stones, and Functional Data Analysis: Recommendations for Developing and Validating New Measures of Depression. *Measurement: Interdisciplinary Research and Perspectives* 4 (3), 2006, pp. 188–197.

[Sch96] Schmitt, Neal. Uses and Abuses of Coefficient Alpha. en. *Psychological Assessment* 8 (4), Jan. 1996, pp. 350–353.

[Sei13] Seiffer, Anja. Determinants of Consumer's Intention to Use and their actual Use of Technology: A Comparison of Established Models. Bachelor's thesis. Karlsruhe: KIT, 2013.

[Sem96] Semon, Thomas T. Is a rating of 10 twice as good as a rating of 5? *Marketing News* 30 (11), May 1996, pp. 16–16.

[SHE05] Schnell, Rainer, Hill, Paul B., and Esser, Elke. *Methoden der empirischen Sozialforschung.* 7. München: R. Oldenbourg Verlag, 2005.

[Shibboleth] *Shibboleth Consortium - What's Shibboleth.* URL: `http://shibboleth.net/about/` (visited on 2015-11-15).

[Spe04] Spearman, Charles. The Proof and Measurement of Association Between Two Things. *The American Journal of Psychology* 15 (1), 1904, pp. 72–101.

[Sta12a] StackExchange Community. *Survey database design: associate an answer to a user - Database Administrators Stack Exchange.* 2012. URL: `http://dba.stackexchange.com/questions/16002/survey-database-design-associate-an-answer-to-a-user` (visited on 2015-11-15).

[Sta12b] StackExchange Community. *Survey database design: first version. Are there errors? - Database Administrators Stack Exchange.* 2012. URL: `http://dba.stackexchange.com/questions/16066/survey-database-design-first-version-are-there-errors` (visited on 2015-08-20).

[Ste46] Stevens, S. S. On the Theory of Scales of Measurement. *Science* 103 (2684), July 1946, pp. 677–680. DOI: `10.1126/science.103.2684.677`.

[Ste74] Stegmüller, Wolfgang. *Wissenschaftliche Erklärung und Begründung.* Probleme und Resultste der Wissenschaftstheorie und Analytischen Philosophie. Bd. I. Berlin/Heidelberg/New York: Springer, 1974.

[Ste80] Stegmüller, Wolfgang. *Neue Wege der Wissenschaftsphilosophie.* Berlin/Heidelberg/New York: Springer, 1980.

[Sup+89] Suppes, Patrick, Krantz, David H., Luce, Duncan, and Tversky, Amos. *Geometrical, Threshold and Probabilistic Representations.* Vol. II. Foundations of Measurement. San Diego, CA: Academic Press, 1989. 493 pp.

[TestLink] *Test Link: About the Test Collection.* URL: `http://www.ets.org/test_link/about` (visited on 2014-12-05).

[Thu27] Thurstone, Louis L. A Law of Comparative Judgment. *Psychological Review* 34 (4), 1927, p. 273.

[Thu94] Thurstone, L. L. A Law of Comparative Judgment (Reprint). *Psychological Review* 101 (2), 1994, pp. 266–270.

[Tor58] Torgerson, Warren S. *Theory and Methods of Scaling.* Oxford: Wiley, 1958.

[Tro06] Trochim, William M.K. Types of Reliability. *Research Methods Knowledge Base*, 2006.

[VD00] Venkatesh, Viswanath and Davis, Fred D. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science* 46 (2), Feb. 2000, pp. 186–204. DOI: 10.1287/mnsc.46.2.186.11926.

[Ven+03] Venkatesh, V., Morris, Michael G., Davis, Gordon B., and Davis, Fred D. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly* 27 (3), 2003, pp. 425–478.

[Wie15] Wiechmann, Brigitte. *Integrated Authority File (GND).* Feb. 2015. URL: http://www.dnb.de/EN/Standardisierung/GND/gnd_node.html (visited on 2015-11-14).

[Wik15a] Wikibooks. *Handbook of Management Scales — Wikibooks, The Free Textbook Project.* 2015. URL: https://en.wikibooks.org/w/index.php?title=Handbook_of_Management_Scales&oldid=2985091 (visited on 2015-11-15).

[Wik15b] Wikipedia. *Fair Use.* Page Version ID: 688279215. Oct. 2015. URL: https://en.wikipedia.org/w/index.php?title=Fair_use%5C&oldid=688279215 (visited on 2015-11-03).

[Wv02] Watt, James H. and van den Berg, Sjef. *Research Methods for Communication Science.* 2002. URL: http://www.cios.org/readbook/rmcs/rmcs.htm (visited on 2015-10-07).