Tools for Visual Inquiry

A Digital Version of the Essence Configuration Table

Master Thesis

cs-23-sd-10-04

Aalborg University
Department of Computer Science

AALBORG UNIVERSITY
STUDENT REPORT

Department of Computer Science
Aalborg University
http://www.aau.dk

Title:
Tools for Visual Inquiry

Theme:
Systems Development

Project Period:
Spring Semester 2023

Project Group:
cs-23-sd-10-04

Participant(s):
Kaare Holland Hattel
Kasper Stenholt Winkler
Mads Smed Enggaard Thomassen

Supervisor(s):
Ivan Aaen

Copies: 1

Page Numbers: 74

Date of Completion:
June, 2022

Abstract:

This thesis continues upon the work of [14] by investigating usage categories related to the guiding features in Digital Essence, their influence on dangling information, and if dangling information affects inquiry. Using Principal Component Analysis and Clustering, we identified and classified four usage categories associated with the guiding features. These categories range from documentation to reflective usage. Focus group interviews revealed that the *CRUD* and *Interconnected CRUD* categories impact dangling information through an implicit and explicit approach. Focus group interviews further revealed that the groups experienced an improvement in their inquiry by identifying dangling information. Our findings contribute to digital versions of the visual inquiry tools by making the model ontology more explicitly available for the users, thus improving the internal consistency of a project.

Resume

In this thesis, we extend our previous work [14] by exploring the concept of dangling information and measuring the effects of our guiding features on reflection, inquiry, and dangling information. This thesis uses the Essence methodology and its configuration table as a basis for researching dangling information in visual inquiry tools.
This thesis investigates the research questions:

- (RQ1) What usage patterns are associated with our guiding features?
- (RQ2) Do the usage patterns influence how dangling information is reduced?
- (RQ3) Does reducing dangling information help users during inquiry?

In our research, we follow the design science research orientation[17]. Our methodology consists of five phases: *Problem awareness*, *Suggestion*, *Development*, *Evaluation*, and *Conclusion*.

To evaluate our research questions, we perform a case study and focus group interviews with students following the Software Innovation Course at Aalborg University. The course participants use Digital Essence in the case study in their innovation projects. The data collected in the case study are analyzed via a Principal Component Analysis (PCA) [18], placing each prospect in a three-dimensional space. The prospects are clustered using the K-means clustering algorithm[9], where we obtain four usable clusters. The behavior of these clusters is classified based on seven metrics, resulting in four usage categories: *Documentation*, *Interconnected Documentation*, *CRUD* and *Interconnected CRUD*.

The student groups behind two of the prospects, belonging to the *CRUD* and *Interconnected CRUD* categories, are interviewed to explore their work with dangling information. The focus group interviews revealed different approaches in which dangling information is reduced. The *CRUD* usage category reduces dangling information implicitly by using the guiding features *Guiding Documentation* and *Connections between cells* to ensure and validate the correctness of the information in their configuration table. The *Interconnected CRUD* usage category explicitly reduces dangling

information by using the guiding feature *Relationships between statements*. The focus group interviews further revealed that both groups experienced an improvement in their inquiry by identifying dangling information.

Contents

Preface **1**

1 Introduction **2**

1.1 Research Question 4

2 Guiding Features in Digital Essence **6**

2.1 Design of Guiding Features in Digital Essence 7

2.2 Design Space for the Essence Configuration Table 9

3 Research Method **11**

3.1 Choice of Methods 11

3.2 Design Options for Digital Essence 12

3.3 Remarks on the Evaluation Setup 15

3.4 Evaluation Strategy for the Case Study 15

4 Visual Inquiry Tools **19**

4.1 Tools for Inquiry 19

4.2 Dangling Information 20

4.3 Dangling Information in VITs 21

4.4 Consequence of Dangling Information 24

5 System Auditing **27**

5.1 System Auditing in Digital Essence 27

5.2 Implementation of System Auditing 29

5.3 Implementation of Digital Essence Metrics 30

6 Quantitative Study **33**

6.1 Principal Component Analysis 33

6.2 Performing the PCA 34

6.3 Results of the Quantitative Study 37

6.4 Interpretation of the Results 38

7 Qualitative Study **39**
7.1 Choice of Groups 39
7.2 Interview Guide 39
7.3 Results of the Qualitative Study 40
7.4 Interpretation of the Results 42

8 Discussion **43**
8.1 Research Method 43
8.2 Metrics 46
8.3 Essence 47

9 Conclusion **48**

10 Future Work **50**
10.1 Extending the Guiding Features 50
10.2 Outdated Information 50
10.3 Further Focus Group Interviews 51

Bibliography **52**

A Interview Guide **55**

B Focus Group Blue **58**
B.1 Part 1: Introduction 58
B.2 Part 2: Assignment 58
B.3 Part 3: Follow-up questions 58

C Focus Group Red **60**
C.1 Part 1: Introduction 60
C.2 Part 2: Assignment 60
C.3 Part 3: Follow-up questions 60

D PCA - Jupyer Notebook **62**

E Clustering of prospects **74**

Preface

A special thanks to our supervisor Ivan Aaen for providing us with valuable insights into Essence and allowing us to experiment with Digital Essence in the Software Innovation Course.

In this thesis, we have changed the name of "Methodology Supports" to "Guiding Features."

Aalborg University, January 13, 2022

Kaare Holland Hattel
khatte18@student.aau.dk

Kasper Stenholt Winkler
kwinkl18@student.aau.dk

Mads Smed Enggaard Thomassen
mthoma18@student.aau.dk

1 Introduction

In *Design Science Research* (DSR), researchers and practitioners propose tools for visual inquiry called Visual Inquiry Tools (VITs) [25, 26, 6]. These tools support users in making 'joint inquiry' into problems, typically in a canvas-style problem space in two dimensions[5]. It helps decompose problems, typically projected into different areas of concern that can help the users' inquiry via visual elements. These tools are used in various domains, including business development, education, and innovation. The primary purpose of VITs is to facilitate joint inquiry by allowing multiple users to collaborate on and explore different aspects of a problem space. A well-known VIT is the Business Model Canvas (BMC) which cleared the path for the term and use of these tools. The BMC focuses on developing Business Models by presenting the users with nine segments, each focusing on a specific part of the Business Model. BMC helps the users reflect on the core elements of their business and can be used to communicate ideas and visions to stakeholders.

Since the creation of BMC, several canvas-based VITs have been produced (e.g., [32, 2, 25]). One such VIT is the *configuration table* used in the software innovation methodology Essence[1]. The Essence configuration table is a tool for visual inquiry into a software-based innovation project by generating and collecting statements about the problem and solution space. The table aims to provide a basis for inquiry and an overview and measure of consistency in a project. In [1], the configuration table is a canvas-based VIT providing twelve configuration cells, divided between four areas of focus, namely the Situation, Contribution, Solution, and Valuation of the innovation project. Our previous work in [14] outlines the Essence methodology and reviews the configuration table as a VIT using a design theory for VITs[4]. Here we deemed the interconnectedness of cells in the configuration table essential for inquiry.

Other works in DSR have considered the design principles of VITs, namely [23], which considers a *design taxnonomy* for VITs which describes design options to consider in different Meta Dimensions such as design-problem, process, solution, and evaluation. Specifically for innovation projects, [20] contributes a theoretical framework for understanding working mechanisms and fields of application of canvas-based VITs. The framework is represented by a *morphological box* which elaborates some of the possible design choices to guide in canvas-based VIT design.

With visual inquiry tools such as BMC and the Essence configuration table, the interconnectedness of the information inside the canvas is essential and is typically represented by a *model ontology* [4, 12, 24]. Maintaining the information detailed by the users within the tool and maintaining the relationships between the different pieces of information can be difficult if the tool does not provide guidance or an explicit model ontology. For example, in BMC, key activities are used to obtain value propositions; thus, implicit relationships exist between key activities and value propositions. By modifying either a key activity or value proposition, proper actions on the counterpart should be considered to ensure the information within the VIT depicts the correct understanding of the business. When users omit to maintain these relationships, the users can introduce information that has no inherent relationship to other parts. We call the information or data without connection to other parts in a VIT *dangling information*.

In our previous work [14], we identified issues in the 'directions of use' in the Essence configuration table. These issues originated from the maintenance of the relations between elements. Ensuring the relationships were complex as they were only implicit and not illustrated in the table. The implicit nature of the relationships can result in missing opportunities in how the information within the problem and solution space depend on each other. Similarly, in the BMC, the interconnectedness of the areas of concern is implicit in the tool rather than displayed as part of the 'directions of use'[24, p. 47]. As the amount of information within a VIT increases, it can become more difficult for the users to maintain a valuable overview of how the information relates to each other. Maintaining these relationships is vital to avoid dangling information. Currently, no related works consider the effects of dangling information, as the relationships between elements in a VIT are often implicitly stated. Our approach to investigating the issue of dangling information is to digitalize VIT to experiment with different measures more efficiently.

Digitalizing tools such as VITs requires translating the current methodology or 'directions of use' to software-based system requirements. Many works in Computer-Aided Software Engineering (CASE) tools focus on software tools used to design and implement applications [35]. In our previous work [14], we proposed Digital Essence as a methodology companion and utilized CASE tool theory to propose guiding features for a digital version of the configuration table.

In this thesis, we extend our previous work and focus on measuring the effects of our guided features on reflection and inquiry.

1.1 Research Question

In previous work [14], we proposed the guiding features: *connections between cells*, *connections between statements*, and *guiding documentation*. These guiding features aim to promote visual inquiry in a digital version of the Essence configuration table, namely the Digital Essence tool. The research in this thesis will use the Essence configuration table as a stepping stone to generalize the guiding features. We aim to investigate how these guiding features are used in Digital Essence and if they impact dangling information. Lastly, we aim to measure the effect of reducing dangling information on inquiry.

In this thesis, we investigate the following research questions:

- (RQ1) *What usage patterns are associated with our guiding features?*
- (RQ2) *Do the usage patterns influence how dangling information is reduced?*
- (RQ3) *Does reducing dangling information help users during inquiry?*

The first research question (RQ1) involves analyzing how the guiding features are used in Digital Essence. Here we aim to identify if different combinations of the guiding features are used and how these combinations can be classified. The second and third research questions (RQ2 and RQ3) address whether the guiding features impact how dangling information is reduced and if the reduction of dangling information improves inquiry. We hypothesize that dangling information is not inherently bad and that the pursuit of reducing it might help users reflect upon their problem and solution space.

To answer these research questions, we continue the groundwork presented in [14] by further exploring different VITs and describing the concept of dangling information. Additionally, we perform a case study in which students following the Software Innovation course at the University of Aalborg use Digital Essence as a methodology companion to Essence. The case study will investigate each group using Digital Essence and includes an analysis of data collected from the groups. This digital version implements the proposed guiding features from [14] to promote inquiry and reduce dangling information. Digital Essence is instrumented with a data tracking module that collects user actions that can be used to gain insight into how Digital Essence is used. The case study aims to identify different usage patterns of the guiding features and how they can be classified, thus answering RQ1. To address RQ2 and RQ3, we conduct focus group interviews with students participating in the case study. We do this to accommodate the shortcomings of the data analysis, as the user actions only show how Digital Essence is used, but not the outcome of reflection and inquiry into the innovation projects.

The remainder of the thesis is structured as follows: Chapter 2 outlines the proposed guiding features from our previous work and presents a new fourth methodology support. In addition, Chapter 2 outlines the impact of guiding features on the presentational format of the configuration table. Chapter 3 presents our research method, and Chapter 4 defines dangling information and investigates its presence in popular VITs. Chapter 5 outlines the data tracking module proposed in [14] and how the data is extracted for data analysis. Chapters 6 and 7 present the case study and focus group interviews, respectively. Lastly, Chapter 8 discusses the results of the experiments concerning the research questions, and Chapter 9 concludes the thesis.

2 | Guiding Features in Digital Essence

In previous work [14], three guiding features were proposed to make a digital version of the Essence configuration table coherently with the Visual Inquiry Tools Design and CASE tool theory. These include highlighting cell patterns, being able to relate statements in cells to each other, and guiding information via a documentation window. Each feature follows the guided CASE tool philosophy to minimize the occurrence of dangling information [35].

- **Relationships between cells**: Highlighting the relationships between the cells of the configuration table may improve inquiry by suggesting how information can be related. These relations can help the users understand how the information in each cell is related and minimize dangling information.

- **Relationships between statements**: relating the content within the cells makes the users aware of the importance of coherency and the interconnectedness of their project. Relationships between statements are a way to describe the interconnectedness and address the potential dangling information that can occur in the configuration table. The ability to relate information may promote further inquiry and reflection into the innovation project.

- **Guiding documentation**: Dangling information is also an issue if the users need clarification about the terminology used in the table. Understanding terminology can prevent incorrect statements from being inserted into cells, making the patterns hard to follow and making it hard to relate to other statements. Guiding information (such as term lists and cell descriptions) is available through a documentation window to address this.

- **Change-advis**: Building on top of the other guiding features from previous work [14, sec. 5.3], we propose another feature to minimize dangling information. This feature consists of a change-advis module that informs users about related statements if a statement is modified. The purpose of this feature is to assist users when modifying information within the configuration table, to evaluate all related pieces of information, ensuring a correct representation of the problem and solution space. Evaluating related information when modifying a statement is a way to reduce dangling information, as it ensures that the users have reflected upon the changes and the effects they might have on the related statements.

2.1 Design of Guiding Features in Digital Essence

In this section, we describe how each of the guiding features is modeled and implemented as part of Digital Essence on a conceptual level.

The configuration table comprises twelve cells, each containing statements providing information about the software innovation project. The first two proposed guiding features concern the cells and statements in the configuration table. To provide a clear overview of the relationships between the cells and statements, the diagram in Figure 2.1 illustrates the relationships between these cells and statements in a simplified manner.

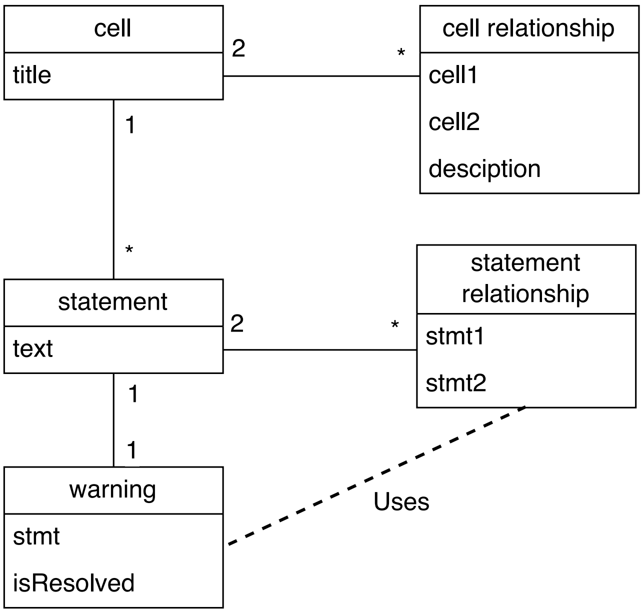


Figure 2.1: Diagram of the data model in Digital Essence

Each cell has a title corresponding to the name in Essence. In [14], we presented the patterns of the configuration table. These patterns indicate how the information within each cell is supposed to be related to information in other cells. The relationships between cells feature adherence to these patterns. The relationships indicated by the patterns are modeled via the *cell relationship* in Figure 2.1. Each cell relationship includes a reference to two related cells and a relation description.

An example of this is the relationship between Capabilities and Value where Value is "Achieved through" the Capabilities. Similarly, each statement relationship consists of a reference to two related statements. Lastly, the cells can hold statements as part of their content

By modeling the relationships of cells and statements in this way, we can develop a change-advis module to inform users about the potential impact of changes made to their configuration. The change-advis module is designed to advise about the relationships between related statements whenever a user changes a statement. If the statement relates to other statements, it generates a warning informing the user to evaluate the related statements to ensure cohesiveness.

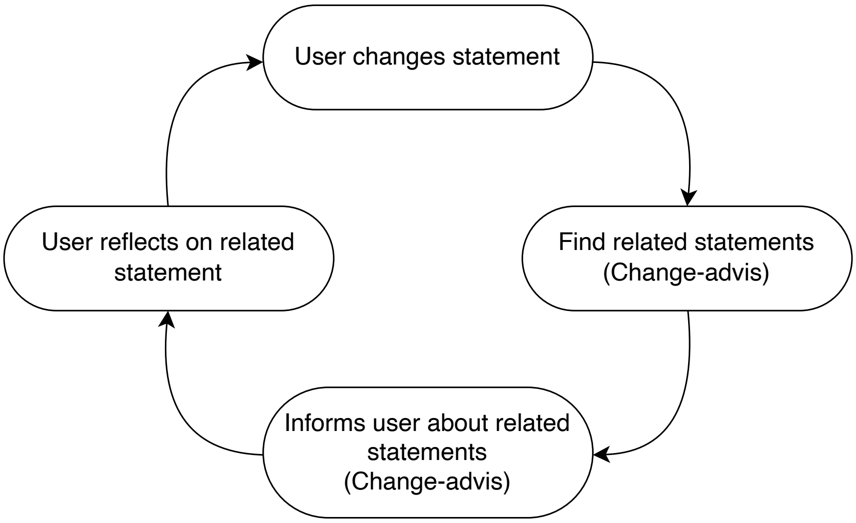


Figure 2.2: Diagram of change-advis flow in Digital Essence

In Figure 2.2 illustrates what is needed to generate warnings. Here the change-advis module needs to find related statements using the statement relationship and use them to make the warnings as modeled in Figure 2.1.

Since Essence is still being developed, we want to ensure that the guiding documentation is flexible and can be updated easily. The documentation pages of Essence terminology are written in markdown files. The use of markdown files allows for quick updates in case of changes to the terminology of the methodology. The documentation includes descriptions of terminology and other relevant documentation related to Essence.

2.2 Design Space for the Essence Configuration Table

In [20], the authors propose different design choices when designing canvas-based VITs for innovation. They describe six different design parameters for innovation canvases and their possible choices. These design parameters include the following:

Process Step: is where the canvas is located within the innovation process. The process steps include further categorization of either "fuzzy front end" (FFE), "new product development" (NPD), or "commercialization" (C).

Media: the input media of the canvas. The media can be anything from handwriting/sketching to physical cards or computer support.

Sequence: how the sequence of the tool is enforced. Either by strict numbering, guiding arrows, or no guidance.

Instructions: the description of the element and ontology of the tool. The instructions can be printed inside the elements, outside the elements, as a supplementary guide, or without instructions.

Elements: describes how many elements are presented in the tool.

Design Specifications: the design-specific elements used in the tool.

We use this framework to describe how the guiding features change the design space of the Essence configuration table. The design space is presented in Table 2.1.

<table>
<tr><th>Parameter</th><th colspan="9">Possible Choices</th></tr>
<tr><td>Process Step</td><td>FFE
Research</td><td>FFE
Synthesis</td><td>FFE
Ideation</td><td>NPD
Refining</td><td>NPD
Prototype</td><td>NPD
Test</td><td>C
Assessment</td><td>C
Business Model</td><td>C
Sales</td></tr>
<tr><td>Media</td><td colspan="2">Handwriting/Sketching</td><td colspan="2">Post-it</td><td colspan="2">Stickers</td><td>Cards</td><td colspan="2">Computer Support</td></tr>
<tr><td>Sequence</td><td colspan="3">Numbers</td><td colspan="3">Arrows</td><td colspan="3">None</td></tr>
<tr><td>Instructions</td><td colspan="2">None</td><td colspan="3">Printed within Element</td><td colspan="2">Printed outside Element</td><td colspan="2">Supplementary Guide</td></tr>
<tr><td>Elements</td><td colspan="2">3-6</td><td colspan="2">7-8</td><td colspan="2">Excatly 9</td><td>10-14</td><td colspan="2">15+</td></tr>
<tr><td>Design Specifics</td><td colspan="2">Icons</td><td colspan="2">Checklists</td><td colspan="2">Color-Code</td><td>Visual Metaphor</td><td colspan="2">Integrated Framework</td></tr>
</table>

Table 2.1: The Design Space for the Essence Configuration Table

To describe the effect the four proposed guiding features have on the design space of the configuration table, the table is updated. The new table is presented in Table 2.2. The **media** is changed from *handwriting/sketching* (either digitally or physically) to *computer support* following the digitalization of the configuration table. Implementing the guiding feature *relationships between cells* changes the **sequence** to *arrows*, as these relationships are indicated by arrows in Digital Essence. Initially, the instructions for the configuration table, such as the patterns and descriptions of each cell, are provided in [1]. By implementing the *guiding documentation* feature, the **instructions** is changed to *printed outside element*. The feature *relationships between statements* and *change-advis* introduces three new **design specifics**, namely *icons*, *checklists* and *integrated framework*. The icons indicate relations between statements, and checklists display warnings from the change-advis. The integrated framework is achieved by creating a user interface, which assists the user in establishing relationships between statements.

<table>
<tr><th>Parameter</th><th colspan="9">Possible Choices</th></tr>
<tr><td>Process Step</td><td>FFE
Research</td><td>FFE
Synthesis</td><td>FFE
Ideation</td><td>NPD
Refining</td><td>NPD
Prototype</td><td>NPD
Test</td><td>C
Assessment</td><td>C
Business Model</td><td>C
Sales</td></tr>
<tr><td>Media</td><td colspan="2">Handwriting/Sketching</td><td colspan="2">Post-it</td><td colspan="2">Stickers</td><td>Cards</td><td colspan="2">Computer Support</td></tr>
<tr><td>Sequence</td><td colspan="3">Numbers</td><td colspan="3">Arrows</td><td colspan="3">None</td></tr>
<tr><td>Instructions</td><td colspan="2">None</td><td colspan="3">Printed within Element</td><td colspan="2">Printed outside Element</td><td colspan="2">Supplementary Guide</td></tr>
<tr><td>Elements</td><td colspan="2">3-6</td><td colspan="2">7-8</td><td colspan="2">Excatly 9</td><td>10-14</td><td colspan="2">15+</td></tr>
<tr><td>Design Specifics</td><td colspan="2">Icons</td><td colspan="2">Checklists</td><td colspan="2">Color-Code</td><td>Visual Metaphor</td><td colspan="2">Integrated Framework</td></tr>
</table>

Table 2.2: The Design Space for Digital Essence

3 | Research Method

In this chapter, we describe our choice of research methods. The methods used are inspired by those used in the field of DSR. Firstly, we elaborate on our choice of overall research method. Secondly, we describe the design options for Digital Essence and the methods used to evaluate our prototype. Thirdly, we elaborate on how our research method addresses the research questions described in chapter 1.

3.1 Choice of Methods

This thesis aims to establish guiding features and generate knowledge regarding usage patterns and dangling information in VITs. It seeks to investigate the effect of minimizing dangling information when using digital versions of VITs. As the research in this thesis is within the area of VITs, we are adopting research methods from Design Science Research (DSR). DSR is a research orientation that concerns inventing or building new artifacts for solving problems[17]. Researchers and practitioners in DSR often produce IT artifacts that can be classified as VITs [6].

In a related work [30], the authors derive design requirements and produce a software-based prototype of a tool supporting reflection in design thinking projects. Their work focuses on *"[...]the lack of guidelines of how software can be used to support (collaborative) reflection during design thinking projects."* [30, p. 5]. They build an IT artifact based on requirements derived from theory as well as relevant work and experts in the field. The IT artifact is evaluated based on *ex-post* evaluation with qualitative analysis of data gathered through data logging and workshops with master-level students of Information Systems in an *artificial* setting[34]. Their research method is derived from [27], which describes a DSR methodology for Information Systems Research. Since [30] researches the designing and building a VIT similar to Digital Essence, this thesis takes inspiration from this work to establish a suitable research method. The phases of our research method can be seen in Figure 3.1. The method consists of five phases, where findings in one phase contribute to the next. We iterate over the different phases several times as each phase extends our knowledge base.

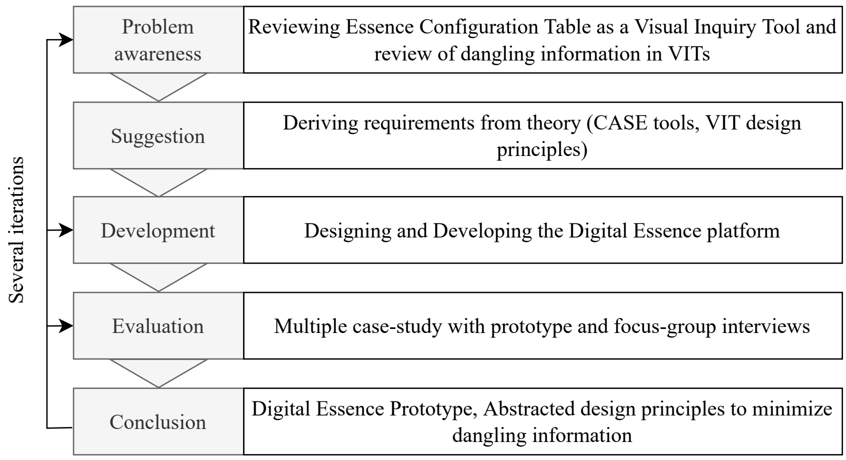


Figure 3.1: Design science research method[30]

Previous work [14] focused on the Problem awareness, Suggestion, and Development phase. In [14, ch. 4], we reviewed the Essence configuration table as a VIT, identifying strengths and weaknesses. Based on these findings, we derived requirements from CASE tools theory and VIT design principles. These requirements were conceptualized into the guiding features presented in chapter 2. Implementing the features in practice resulted in the development of the IT artifact called *Digital Essence*.

In this thesis, we reiterate the first three phases by investigating dangling information in other VITs, proposing a fourth guiding feature, and implementing it in Digital Essence. The main focus of this thesis is to evaluate the guiding features in practice and conclude upon the research questions. The Evaluation and Conclusion phases validate the guiding features and research questions.

3.2 Design Options for Digital Essence

In [23], the authors present a framework for choosing design options when building and evaluating VITs. Each VIT design phase is categorized into what the authors call *meta-dimensions*. These meta-dimensions represent critical phases a designer goes through. They include *Design Problem*, *Design Process*, *Design Solution*, and *Design Evaluation*. Each meta dimension includes another dimension representing an area of concern in a given phase, e.g., Design Problem includes *Design Purpose* and *Design Element*. For each dimension, different *Characteristics* can be chosen. The Characteristics are based on 24 visual inquiry tools developed in design science and action design research, with 15 empirical examples[23]. We use this framework to outline the design options chosen for Digital Essence. The selected options are highlighted (in gray) in Table 3.1.

MD	Dimension	Characteristics					EX
Design Problem	Design Purpose	Collaborative Ideation	Design a New Artifact	Analyze / Support Process	Alignment		No
	Design Element	Data	Gamification	Digital Transformation	Requirement Engineering		No
		Organizational Phenomenon	Ideas	Research Process	Services	Public Value	
		Business Models	Mobile Applications	Artificial Intelligence	Digital Platforms	...	
Design Process	Design Method	ADR		DSR			Yes
	Design Philosophy	Ontology-Based		Requirements / Principle-Based			Yes
	Design Requirement Source	Interviews	Survey	Workshops	Litterature		No
Design Solution	Design Origin	New		Adapted			Yes
	Design Medium	Print-Out		Digital Template	Application		No
	Design Output	Stand-Alone		Part of Toolkit			Yes
Design Evaluation	Evaluation Strategy	Case Study	Workshops	A/B Test	Focus Group	Questionnaire	No
	Evaluation Criteria	Usability	Practicability	Impact	Usefulness		No
		Efficacy	Effectiveness	Efficiency	Elegance	Ethicality	

Table 3.1: Design Options [23, p.6]

Design Problem The options within the Design Problem are the same as for the Essence configuration table, namely *collaborative ideation*, *ideas*, *requirement engineering* and *research process*. The reason is that this study is not concerned with giving the configuration table a new purpose but rather improving upon the design to minimize dangling information.

Design Process The design method used in developing Digital Essence lies within *DSR* as mentioned above. The overall design philosophy in the Essence configuration table is an *Ontology-based* approach where each cell has a name and a meaning. The design requirements sources for the configuration table are derived from software innovation literature and pragmatism[1, p. 16].

Design Solution Regarding the design solution, the configuration table is *adapted* to a digital version, namely Digital Essence, to utilize and implement the proposed guiding features. As some of the guiding features are based on user interactions, the design medium is a digital *application*. Even though the Essence configuration table is part of other tools in the Essence methodology, such as Prospect Scenarios, Digital Essence is a *stand-alone* VIT.

Design Evaluation For the evaluation strategy, a combination of *case study* and *focus groups* are selected, as these two strategies complement each other. We can conduct a quantitative study through a case study of the groups using Digital Essence. In the case study, we can collect data for identifying and classifying their behavior in Digital Essence. With focus group interviews, we can conduct a qualitative study, investigating the internal reflection process of groups with different user behavior. This interview format is primarily a non-governing interview style, focusing on getting different perspectives from the participants. The moderator's purpose is to present the topics for discussion and facilitate the joint discussion and reflection of the participants.

The master-level students are working on their projects in groups. Focus group interviews enable us to investigate the reflection and joint inquiry within these groups. As the projects are different, conducting multiple focus group interviews can be beneficial. *A/B tests* are not selected as an evaluation method, as dividing the students between two different versions of Digital Essence could affect their coursework. We deem that *Questionaires* and *Workshops* would not provide the needed depth and quality of data to answer RQ2 and RQ3.

In this thesis, we focus on building the right thing rather than building the thing right, as we aim to investigate the impact of the guiding features in VITs. Based on this, we focus on the evaluation criteria *usefulness*, *impact*, and *effectiveness*.

3.3 Remarks on the Evaluation Setup

The participants of the case study and focus group interviews are master-level students following the Software Innovation course at Aalborg University. All students following the course have been instructed to use Digital Essence as a digital companion to the course. During this thesis, we have worked as teaching assistants in the Software Innovation course, where we have had the opportunity to assist the students in their work with Digital Essence. By acting as teaching assistants, we have gained insights into the course's curriculum and thus better contextualize the student's work with Digital Essence.

3.4 Evaluation Strategy for the Case Study

Measuring a group's work with dangling information and the reflection process is complex, as perceiving dangling information is subjective. How can we identify when users are reflecting and if what they are reflecting upon is related to dangling information?

3.4.1 Measuring reflection

When users use Digital Essence to insert, edit, or delete statements in their project, it can be hard to identify whether the statements are based on existing or new knowledge. The group might use Digital Essence to document their project if the statements are based on existing knowledge. If the group creates statements using Digital Essence to reflect and generate new knowledge, then Digital Essence serves its purpose as a visual inquiry tool. We seek to measure the level of *team reflexivity* when using Digital Essence, as this can help us evaluate if the guiding features bring value and how they are used. In [36], the author describes team reflexivity as the "*extent to which team members collectively reflect upon the team's objectives, strategies, and processes, as well as their wider organizations and environments, and adapt them accordingly.*" In terms of Essence, we can use this definition as the extent to which the team members collectively reflect upon the configuration table cells and the interconnectedness between statements and update them accordingly to adapt to changes.

Some measures for reflexivity in teams exist in the literature as described in [21]. One of the early reflexivity measures used in research is Carter & West[7]. Here the authors describe team reflexivity measures as eight items, where three of them are relevant to this thesis. The focus items are outlined below, including their relevance to Digital Essence.

The team often reviews its objectives: The more sessions a team has, the more they may reflect on previous statements and relationships.

The team modifies their objectives in light of changing circumstances: If the users 'react' to changes in their project and the configuration table, they probably do so because they reflect on the changes. This reaction is called a pivot in Essence.

The team often reviews its approach to getting the job done: The team reflects on their use of the configuration table and the terminology of Essence.

Some items presented in [7] we are not able to measure with Digital Essence. An example of this is the item: *"The team regularly discusses whether the team is working effectively together"* since this is out of the scope of the Essence methodology.

To address RQ1, we will define a series of user actions and metrics. The user actions will be used to identify different user behavior in Digital Essence, and the metrics will be used to classify these behaviors.

3.4.2 User Action Types

To investigate the differences between the groups in Digital Essence, we will use the actions a person can perform within the platform as the basis of comparison. Examples of such actions are insertion, deletion, and modification of statements, number of sessions, and classification of statements. The user actions are defined in Table 3.2. In this section, we use the terms team and group interchangeably to describe the users working with Digital Essence.

User Action in Digital Essence	
D1: Number of inserted statements	D7: Number of standard statements
D2: Number of updated statements	D8: Number of future statements
D3: Number of deleted statements	D9: Number of uncertain statements
D4: Number of created relations	D10: Number of sessions
D5: Number of deleted relations	D11: Number of statements that follows configuration table patterns
D6: Number of wiki look ups	

Table 3.2: User Actions in Digital Essence

3.4.3 Metrics

We will use a series of metrics to classify the differences between the groups in Digital Essence. These metrics are based on measures for reflexivity, dangling information, and user action types. As described in chapter 2, we have introduced four guiding features to reduce dangling information. The metrics we define will be related to the different guiding features and can, in combination with each other, help to address RQ1. These metrics are as follows:

- (M1) The percentage of *standard* statements without relationships.

As described above, dangling information is subjective. A potential way of measuring dangling information is by investigating the number of standard statements without relationships. This metric indicates the amount of dangling information in a prospect and thus provides us some insights into the effectiveness of the guiding features in reducing dangling information. We must address whether users use the guiding feature for *relationships between statements*. If not, detecting dangling information based on this metric becomes unattainable. In this case, the focus group interview should be used to discover the group's work with dangling information.

- (M2) The percentage of inserted statements that follow a configuration table pattern.

We hypothesize that by following the patterns described in [14, sec. 2.5], the information in the configuration table is more accurate to the users understanding of the problem and solution space. We thus want to investigate to what degree the users follow these patterns when completing the configuration table. These patterns are illustrated by the guiding feature *relationships between cells*.

- (M3) The distribution of Standard, Uncertain and Future statements.

The distribution of statement classifications is relevant to dangling information, as we believe that dangling information can be reduced by either relating statements to each other or classifying them appropriately. Classifying a statement as "future" or "uncertain" indicates that the users have reflected upon a given statement's position and impact on the prospect.

- (M4) Number of changes done in relation to warnings.

Another indication of a reflection process could be if the users know that they consider the related statements when they update a statement. Reflection in a team can occur when a team reviews their project in the light of changes. The change-advis guiding feature helps the users be aware when they edit related statements. They receive a warning about it and can act upon it. After taking the appropriate action, they can resolve the warning.

- (M5) Number of sessions for a prospect.

If the users use the configuration table as documentation, they will likely fill out most of the table in a few sessions. Using the configuration table as a documentation tool suggests less reflection and inquiry. A measure of reflexivity is that a team often reviews its objectives. This metric could provide valuable insights into how engaged a team is. A high number of sessions could indicate a high engagement. Thus the users are more likely to have been actively using the configuration table, which could indicate that they are more reflective than if they have only a few sessions.

- (M6) Percentage of inserts compared to deletions/modifications of statements per session.
- (M7) Percentage of added relations compared to deleted relations per session.

We want to measure the ratio between inserted and deleted/modified statements and created/deleted relationships to further investigate teams' engagement during a session. We hypothesize that the level of reflection and engagement is higher when users introduce new knowledge in statements or relationships and change or remove existing knowledge. Modifying or deleting a statement or relationships could result from the team learning something new regarding their prospect.

4 | Visual Inquiry Tools

This chapter describes visual inquiry tools and defines the principle of dangling information. Firstly, it investigates dangling information in two other VITs: Business Model Canvas and Value Proposition Canvas. Secondly, it describes the characteristics and consequences of dangling information.

4.1 Tools for Inquiry

John Dewey defined *inquiry* as *"[...]the controlled or directed transformation of an indeterminate situation into one that is so determinate in its constituent distinctions and relations as to convert the elements of the original situation into a unified whole"*[11, p. 104-105]. In other words, inquiry involves the controlled or directed transformation of a situation that lacks clear definitions and relationships into a specific one. By doing so, the elements of the original situation are brought together in a unified whole. When inquiry occurs, the transaction between the problematic situation and the resolution changes how we perceive the elements of both the problem and the resolution. When we take action to resolve a problematic situation, it changes the said situation. Having tools to effectively promote inquiry when working with highly complex tasks such as business modeling or software innovation is valuable. It can help structure, prioritize, and communicate information. Tools for promoting inquiry are widely used in many different domains[4].

Visual Inquiry Tools (VITs) are often canvas-based tools with visual elements that seek to facilitate inquiry into some problem space. Various VITs target business development, software development, and innovation domains. In [16], Horn and Webber detail how such tools facilitate ideation into a problem and potential resolutions for it, in addition to promoting information sharing and joint inquiry. The authors further describe how tools can be used to understand the interconnectedness between problems by investigating how the different chunks of information within the problem are related. They define some of the characteristics of what they call *wicked problems*, a term originally defined by Rittel & Webber [28], which essentially are complex problems with interrelated problems. Understanding the interconnectedness in solving these wicked problems suggests that the connections between elements in tools such as VITs are vital to achieving the intended outcome of its use.

4.2 Dangling Information

The Essence configuration table is decisive for inquiring into a software-based innovation project's complex problem and solution space. It both represents information and generates new information via visual inquiry. As with other VITs, it consists of interconnected information that, if correctly arranged, constitutes a coherent model of the problem and solution space. However, VITs can also be prone to what we in this project define as *dangling information* issues, where components are unclearly labeled or not correctly related. The dangling information can lead to confusion and incorrect interpretation of the information presented in the VIT, similar to *dangling pointers* in programming, which can cause bugs and crashes in the system. In programming, a dangling pointer points to an arbitrary memory location. When a program attempts to access the memory location pointed to by a dangling pointer, it can result in unexpected behavior, crashes, or other errors. Similarly, in a VIT, pieces of information can be (implicitly) linked to information that either does not exist anymore or is outdated. A VIT's dangling information can result in unexpected or incorrect interpretations. Users should be precise with the information they introduce in the VIT, or they might draw incorrect conclusions.

In the case of the Essence configuration table, Capabilities (entries in the Capabilities cell) without relation to any of the Values (entries in the Value cell) infers that the proposed Contribution has a Capability that is not used in any of the Values, meaning that it does not correspond to any Value in the project. This *dangling information* could result in developers implementing a capability that does not provide any value.

Figure 4.1 illustrates pieces of information in a graph where A, B, C, and D are interconnected, but E has no related information. Information or data like E is what we describe as *dangling information*. In this project, we use dangling information to describe information without connecting to other pieces of information, whether formal or informal, relationships between data. Further, we use data and information interchangeably to describe the factual statements entered into VITs.

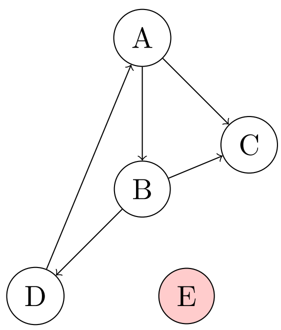


Figure 4.1: Dangling Information in a graph

4.3 Dangling Information in VITs

In [14, ch. 4] we described how dangling information could occur in the Essence configuration table. In this section, we further investigate the principle of dangling information within popular VITs, to determine the relevance of this principle.

4.3.1 Business Model Canvas

The Business Model Canvas (BMC) is a strategic management template for business development. Osterwalder proposed BMC, based on the work from his Ph.D. dissertation [24]. The BMC covers four primary areas of a company, namely *Product*, *Customer interface*, *Infrastructure management* and *Financial aspects* [24, p.42]. These areas are further broken down into a set of nine interrelated elements, also described as *Business Model Elements*(BME) [24, p.43]. In Osterwalders dissertation, he created a *Business Model Ontology*, which he describes as a set of elements and their relationships, to describe a firms money earning logic. The ontology contains the nine business model elements of the BMC. Osterwalder provided a format for describing each BMEs, presented in Figure 4.2.

Name of BM-Element	NAME
Definition	Gives a precise description of the business model element.
Part of	Defines to which pillar of the ontology the element belongs to or of which element it is a sub-element
Related to	Describes to which other elements of the ontology an element is related to.
Set of	Indicates into which sub-elements an element can be decomposed.
Cardinality	Defines the number of allowed occurrences of an element or sub-element inside the ontology.
Attributes	Lists the attributes of the element or sub-element. The allowed values of an attribute are indicated between accolades {VALUE1, VALUE2}. Their occurrences are indicated in brackets (e.g. 1-n). Each element and sub-element has two standard attributes which are NAME and DESCRIPTION that contain a chain of characters {abc}.
References	Indicates the main references related to the business model element.

Figure 4.2: Description of a business model element [24, p.47].

The format presented in Figure 4.2 consists of seven attributes, one of which describes the BMEs' relations to other BMEs. This attribute is helpful for the work in this project, as this infers the presence of relationships within the original work behind the BMC, and thus the possibility of dangling information to occur. These relationships are further presented in Figure 4.3, which illustrates how BMEs are related to other BMEs across the four main areas of a company.

VPD Companion is a software platform with a digital version of the BMC, developed by Osterwalders company Strategyzer. This tool allows users to collaborate on and fill out the BMC. The canvas is filled out in the form of sticky notes of different types, including *Experiment*, *Insigt*, *Hypothesis*, *Evidence* and, *Standard*. These notes can also be related to one another, as depicted in Figure 4.4, in which the blue lines indicate the relationships between the notes. The VPD Companion provides multiple guiding features, including guiding documentation for each segment, context-specific notes, and connections between notes.

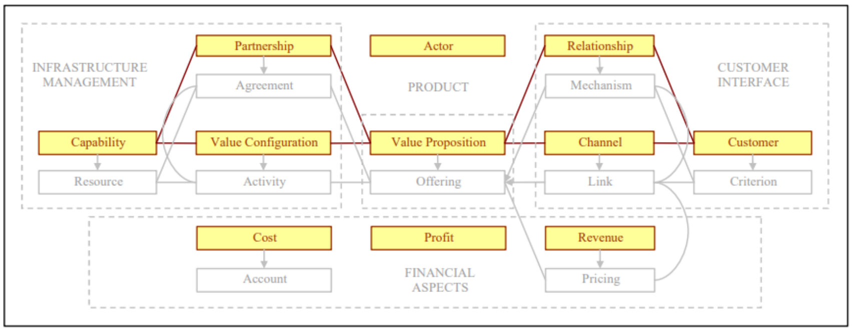


Figure 4.3: The Business Model Ontology [24, p.44].

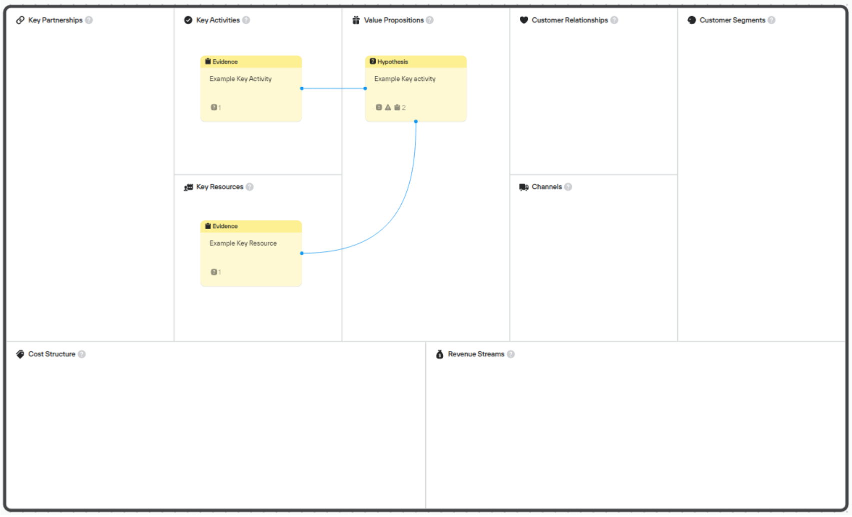


Figure 4.4: VPD Companion - Example of BMC

4.3.2 Value Proposition Canvas

The Value Proposition Canvas (VPC) is a VIT for creating products and services that fit customers' needs. It serves to find a fit between the customer segment and the value propositions that new products or services should provide. VPC consists of a customer profile and a value map, each with three segments. Figure 4.5 illustrates the VPC, where on the left canvas is the value map, and on the right is the customer profile.

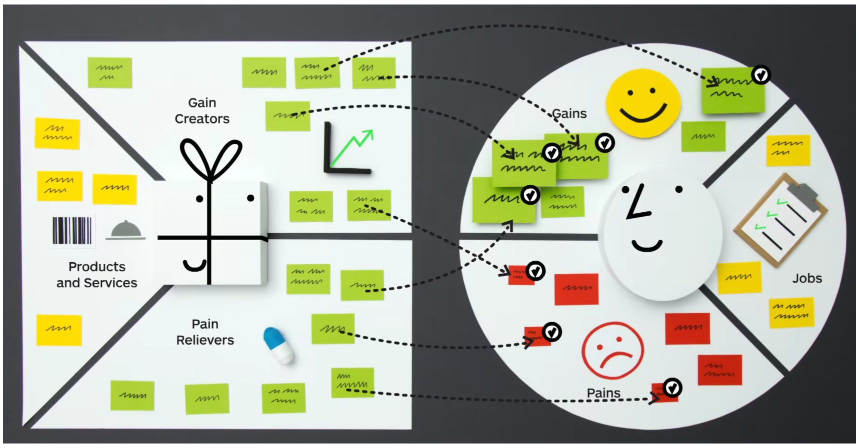


Figure 4.5: Value Proposition Canvas with connections [2, video].

In the VPC, the users are supposed to insert small cards with text describing their current knowledge for each segment provided by the tool. The customer profile consists of Gains, Pains, and Jobs, and the value map consists of Gain Creators, Pain Relievers, and Products and Services. The value map serves to understand how a company's products and services address the customers' needs, wants, fears, and aspirations through the gain creators and pain relievers offered by the products and services.

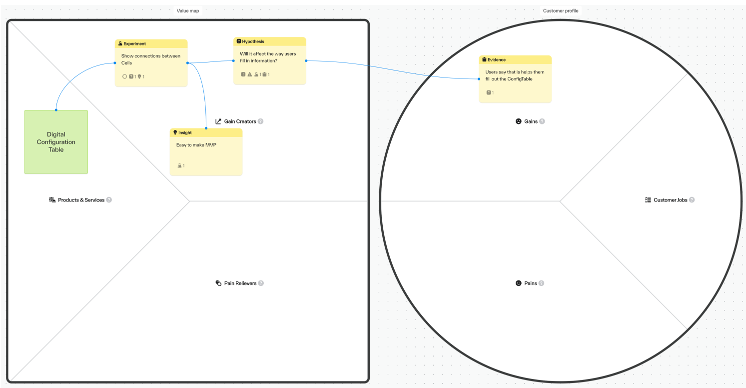


Figure 4.6: VPD Companion - Example of VPC

As shown in Figure 4.5, the different gain creators and pain relievers should be related to the gains and pains of the customer profile. These relationships help users understand how a given product or service supports their users. The tool VPD Companion, described in subsection 4.3.1, can also create a digital VPC. It provides a physical frame to input digital sticky notes into each segment. Furthermore, it helps the users provide context to each digital note via the specific note types. In Figure 4.6, a small demonstration of the tool is presented, in which sticky notes are used to describe pieces of information and how they are related.

4.3.3 Essence Configuration Table

The configuration table is a visual inquiry tool presented in Essence's software innovation methodology and is the subject of this thesis. The purpose of the table is to both depict existing knowledge and explore new within the problem and solution space of the innovation project. The table depicted in Figure 4.7 consists of 4 columns corresponding to the four views of Essence and three rows corresponding to three levels of concern, described in [14, p. 5-8]. The rows and columns constitute 12 cells, each focusing on a specific part of the problem or solution space. The configuration table is designed so that the cells are interconnected. For example, the Capabilities cell is related to the Value cell, as the different capabilities of a system are used to create value in the solution. These relationships also mean that information within each cell should be related to information in other cells. The relationships within the configuration table are further described in [14, p. 5-8]. To ensure the configuration table depicts the correct information regarding the problem and solution space, users must be aware of and maintain these relations throughout the innovation project.

Core Concern	Situation	Contribution	Solution	Valuation
Rationale: Why?	(P) Problem	(L) Leverage	(R) Resolution • Prospect • Warrant • Backing	(CR) Criteria for Rationale
Strategy: What?	(O) Outer Environment	(I) Inner Environment	(Q) Qualification • Reservation • Rebuttal	(CI) Criteria for Strategy
Tactics: How?	(M) Manifestations	(C) Capabilities	(V) Value	(CV) Criteria for Tactics

Figure 4.7: Essence Configuration Table [1, p. 21]

4.4 Consequence of Dangling Information

In this section, we outline the consequences of dangling information. Here we outline potential dangling information for each VITs considered in this chapter. Following this, we describe some general remarks on why dangling information should be reduced.

4.4.1 Essence Configuration Table

As the purpose of the configuration table is to both depict existing knowledge and explore new within the problem and solution space of an innovation project, it is important to ensure that the knowledge depicted aligns with the users' current understanding of both spaces. If the information depicted differs from the users understanding, it can affect the innovation project. As previously stated, Capabilities without relation to Value may result in developers implementing features that do not contribute to the problem's solution. Another example could be a lacking relation between the problem and resolution. As an innovation project aims to solve a particular problem, it would be problematic if the stated problem is unrelated to the proposed resolution. These issues were discussed in further detail in [14, ch. 4].

4.4.2 Business Model Canvas

The Business Model Canvas (BMC) describes the logic of how a firm intends to deliver value and make money, inquiring into the nine business model elements [31]. In Osterwalders dissertation [24], he described how different BME related to each other. These relations indicate how everything within the firm accumulates into the sellables of the firm, which in turn generate money by reaching the customers. Dangling information in the Business Model Canvas impacts the firm's money-earning logic, as inconsistencies within the canvas could mean lost opportunities or poorly invested capital. An example could be a value proposition without relations to customer relations or channels. Neglecting this relationship would mean no means of distributing the value proposition, thus a lost opportunity to generate income.

4.4.3 Value Proposition Canvas

As shown in Figure 4.5, the two canvases presented in the VPC are closely connected. Gain creators are connected to the gains, and the pain relievers are the customer's pains. Together with connections between the two canvases, the individual segments also have some interconnectedness. The products and services in the value propositions are (and should be) directly connected to gain creators and pain relievers. Similarly, in the customer segment, the jobs are where the pains and gains of the customers originate. Dangling information can occur if, for example, any gain creators are not connected to something in the products and services. The users may describe gains that they do not provide. The missing connection can result in misunderstandings of what the company offers as gains to the customer segment.

4.4.4 The Characteristics of Dangling Information

Based on the examples above, dangling information in VITs can misrepresent knowledge, which may lead to missed opportunities or resources wrongly used. Apart from these consequences, some dangling information may exist in VIT.

VITs mentioned above do not dictate which kinds of information are permitted. The users can write information relevant to the time of writing, but also ideas to explore or information that may be useful in the future. Essence works with classifying information where uncertain information or information considered in the future is in *cursive*. In Digital Essence, to expand on this, three classifications of information are implemented; *standard*, *future*, *uncertain*. When writing, users might not know the nature and implications of the *future* and *uncertain* statements. They may not have any natural relationships to other statements in the project. These examples of dangling information should be allowed to exist until the users can resolve their relationships naturally or conclude upon their relevance. Similar information abstractions are also implemented in the VPD companion of BMC and VPC. In these online VITs, users can classify information as *hypothesis*, *experiment*, *insight*, and *evidence*.

Apart from the consequences of dangling information, which can be deduced based on the presence of relationships within the investigated VITs, another possible result of dangling information is the promotion of inquiry. An assumption in this thesis, which will be examined in the case study, is that dangling information in VITs promotes further inquiry as users actively try to reduce it. This process may mature the users' knowledge and understanding of both spaces and thus improve the outcome of the VIT. Understanding that not all dangling information is created equal means that reducing dangling information may be a source of inquiry. How the reduction of dangling information promotes inquiry will be investigated in later chapters of this thesis.

5 | System Auditing

To address the research questions, we have implemented a system auditing module that provides historical data from each group to analyze their user behavior in Digital Essence. Data gathering is essential for identifying different usage patterns in Digital Essence. This chapter aims to describe the implementation of the system auditing module and how it is used to gather historical data from all participating groups. This chapter includes details about the audit database table and the types of information it can store. It will also cover implementation details for some metrics described in section 3.4.

5.1 System Auditing in Digital Essence

In previous work [14], we presented a conceptual description of a system auditing feature for saving historical data in Digital Essence. This feature aims to collect and analyze historical data from the participating groups in Digital Essence, which is used to address the research questions. As described in [14, page 34-35], the system auditing works by storing changes in a single database table, with an associated *timestamp*, *user id* and *prospect id*. This table enables us to explore and filter the historical data based on the time, the different users, and each prospect. We can store all the database transactions in Digital Essence in the system auditing module.

Not all transactions help determine differences in usage patterns. We deem the most useful for determining usage patterns the user action types described in subsection 3.4.2. The user action types focus on capturing relevant data points generated through our guiding features in Digital Essence and are selected based on how relevant we believe they are for answering our research questions.

Figure 5.1 is a diagram illustrating the properties of the audit table in which the historical records are stored. The audit table contains valuable metadata such as *id*, *userId*, *prospectId*, and *createdAt*. This metadata filters and calculates metrics based on each user and prospect. Together with the metadata, the table contains a reference to the original entity. The reference consists of a foreign key (*entityId*) and the name of the table (*model*). Lastly, the audit table contains two properties, *oldValues* and *newValues*, which are used to store the transformation the record goes through based on the action performed by the user.

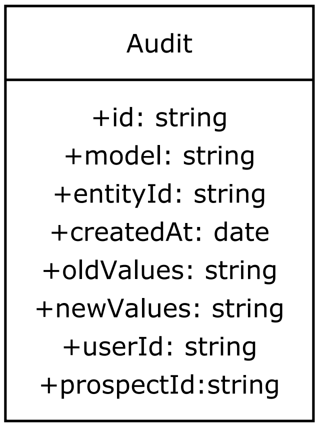


Figure 5.1: Audit table details

An example of how the transaction is stored in the audit table is shown in Figure 5.2. Here we see that 'Transaction 1' is made by 'user 1', which adds a new statement to 'cell 1'. After a while, 'user 2' changes the statement, and we can now see that the old value is 'new statement,' and the new value is 'changed statement.' This tracking with old and new values enables us to go back in time and replay all user actions the user transaction performs.

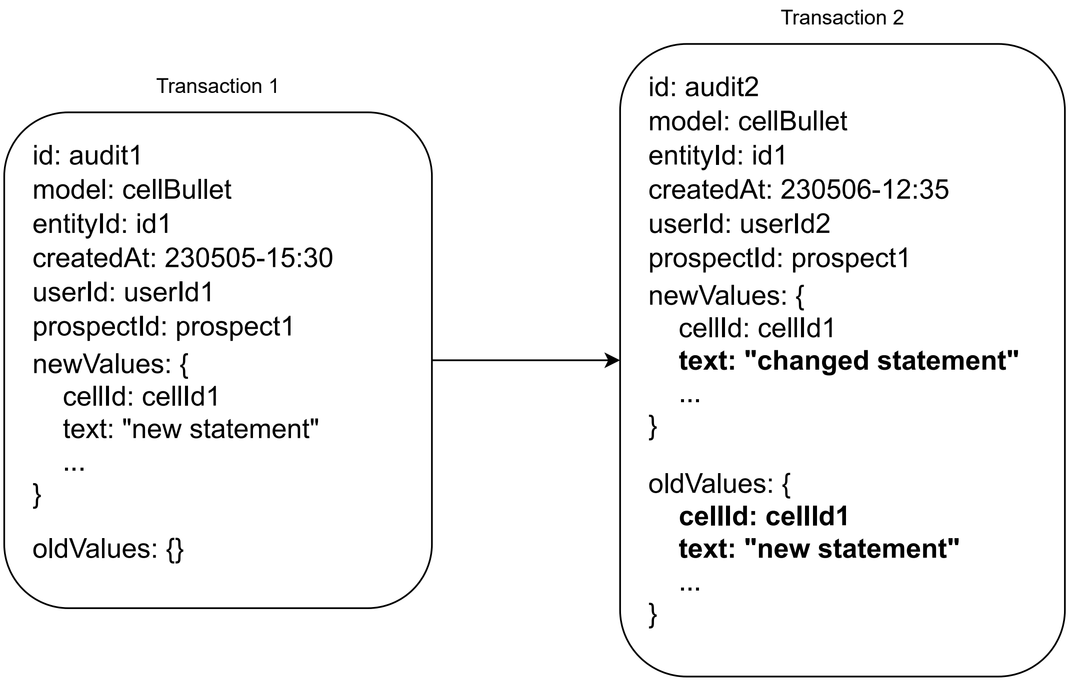


Figure 5.2: Transactions in the Audit table

5.2 Implementation of System Auditing

The data points stored in the audit table are used to investigate the metrics detailed in section 3.4. As pointed out in [14, p. 34-35], one of the disadvantages to our historical data storage solution is that all records are generalized following the same structure and stored within the same table.
To accommodate these shortcomings of the data storage implementation, we have implemented a pre-processing application in .NET (C#), which is responsible for handling the deserialization process and making the structured historical data available through an API. This API is an additional module in collecting tools for Digital Essence. By separating the deserializing and algorithmic processing demands from the visualization logic, we stay true to having our applications in a decoupled and module-based architecture, as previously detailed in [14, p. 29-31]. The decoupling allows us to select different technologies for data visualizations, such as tables and graphs, without worrying about the historical data preparation process.

To visualize the historical data and perform the necessary calculations to answer the metrics, we have investigated two technologies we have used before, namely Jupyter Notebooks[19] and Grafana[22]. Jupyter Notebooks imply a code-oriented approach in the Python programming language, in which the visualizations are created and modified through code. These notebooks are highly flexible and allow us to customize the visualizations to our needs. Grafana is a web application following a low-code approach, where different visualization media are predefined and configurable. Grafana allows us to easily select and test different visualization media by supplying the data source and performing minimal configurations. Grafana dashboards and Python notebooks can be reproducible implementations with access to the API in the pre-processing application.

We choose to implement the visualizations of metrics in a Grafana dashboard, allowing us to quickly evaluate results from the experiment and modify the visualization logic accordingly.

This resulted in the overall architecture shown in Figure 5.3. Here the arrows indicate the direction of the data flow between the modules. The Digital Essence Database provides two types of data: audits and user actions. The data is provided to the Pre-processing Application, which then prepares and calculates the metrics. These metrics are further exposed to Grafana, where two dashboards are implemented. One dashboard provides an overview of the metrics for all prospects, and one selects a single prospect.

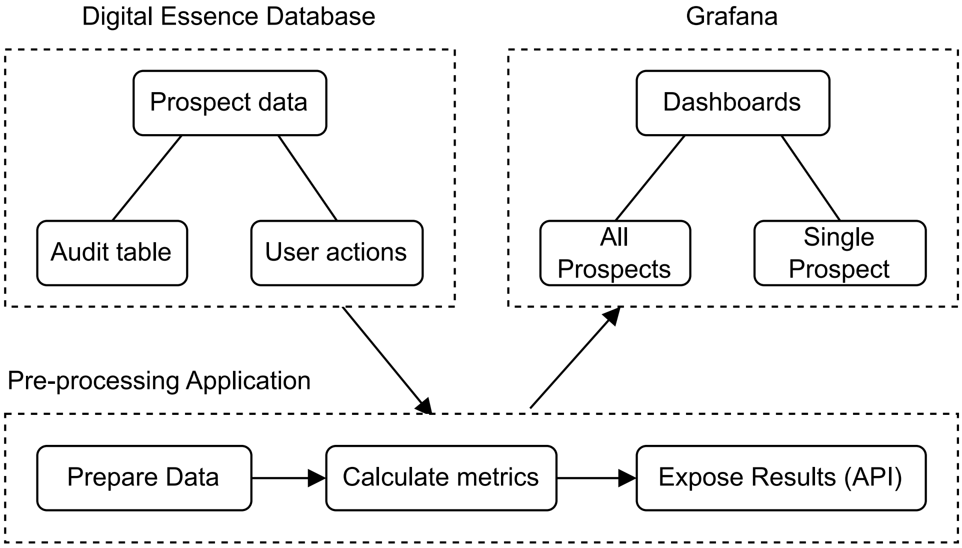


Figure 5.3: Digital Essence - High-level System Architecture

5.3 Implementation of Digital Essence Metrics

Most metrics presented in section 3.4 are fairly straightforward to measure based on the data format. However, the following two metrics are not trivially calculated based on the available data:

- (M2) The percentage of inserted statements that follow a configuration table pattern
- (M4) Number of changes done in relation to warnings

These two metrics are difficult to answer, as they measure a series of actions done by any user in a given prospect. Additionally, a time aspect should be considered, as the time the action takes place matters. This section elaborates on how these two metrics are calculated.

5.3.1 Number of changes done in relation to warnings

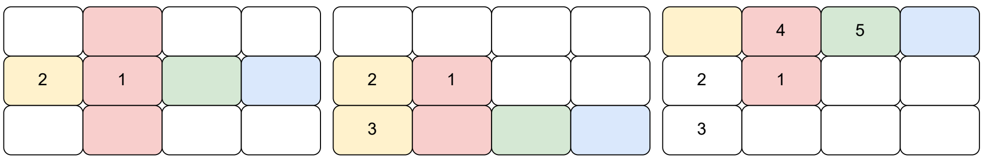
This metric describes how many change warnings the users have reacted to before it is resolved. A change warning is presented to the users whenever a statement with one or more relationships is modified. A change warning is generated for each of these relationships to encourage the users to evaluate these related statements to ensure the statements are still consistent with the users' current understanding of the problem and solution space. When a user evaluates a change warning, it can be marked as *resolved*, and the warning is removed.

Three difficulties must be considered when answering this metric. Firstly, we cannot determine the appropriateness of modifying statements concerning a change warning. A user may have corrected a spelling mistake and not reflected if the content or meaning of the statement should have been changed. Secondly, modifying a statement that triggers the change warnings may not require further changes. In this case, users may still reflect on the statements to ensure everything is correct without performing any changes. In this case, this reflection cannot be measured through the metric. Lastly, some users may prefer to resolve change warnings before making any statement changes. We cannot measure this, as it would require us to determine a general time frame in which we look for changes.

In our solution for measuring this metric, we are not considering the content of a change, as this would require some form of sentence analysis. Additionally, not being able to measure reflection leading to no changes is something we have accepted, as it is simply out of scope in this thesis to measure solely based on the data obtained through system auditing.

5.3.2 The percentage of inserted statements that follow a configuration table pattern

This metric measures the number of statements created following a configuration table pattern. These patterns are described in detail in [14, ch. 2, p. 7-8]. Measuring if and how the users follow these patterns is difficult, as multiple users may be inserting statements interchangeably, and statements may not necessarily be building on top of the immediately previous statement.



(a) Pattern for 1 (b) Pattern for 2 (c) Pattern for 4

Figure 5.4: Example sequence of filling out the configuration table

Figure 5.4 is an illustration of a series of inserted statements in the configuration table, starting with *1* and ending with *5*. The coloring is used to indicate the table's different patterns. In this example, the statements *1*, *2*, and *3* are placed in cells related to each other. The statement *4* breaks the chain, as the cell of *4* and the cell of *3* are not part of the same pattern. However, the insertion of *4* still follows the same pattern as *1*. We believe that filling out the configuration table will result in natural breaks, where the insertion of a statement does not follow a pattern of the immediate statement. Instead, it builds on top of an earlier statement, thus following the pattern of that, as the case is for *4*, which builds on *1*.

Algorithm 1 is a pseudo-code description of how we measure the number of statements inserted according to patterns. The algorithm accepts a list of all inserted statements and outputs the number of statements inserted according to patterns. It evaluates all the statements, and for each statement, it evaluates if that statement follows a pattern of any previous statement.

Algorithm 1 Determine the number of statements inserted according to patterns

Input: *statements*, All inserted statements
Output: n, Number of statements inserted according to patterns

```
visited ← []
n ← 0
for s in statements do
    if visited is empty then
        visited ← s
        continue
    end if
    if visited.last.cell equals s.cell then
        continue
    end if
    if s.cell is related to any in visited then
        visited ← s
    else
        n ← visited.length
        visisted ← []
    end if
end for
```

The pitfall of this approach is that if an inserted statement that breaks the immediate pattern has no relation to any previously inserted statements, it could be classified as the start of following a new pattern. An example of this would be if *4* in Figure 5.4 is unrelated to *1*. In this case, both *1* and *4* share a pattern, but as the statements are unrelated, a user could start a new chain of insertions unrelated to the previous.

6 | Quantitative Study

This chapter presents our quantitative study, consisting of a case study (see section 3.2). The participants are students following the Software Innovation Course at Aalborg University. The purpose is to address RQ1, in which we identify and classify different usage patterns of the guiding features in Digital Essence. The approach consists of performing a Principal Component Analysis (PCA)[18] based on user action types presented in 3.4. Following the PCA, we perform a clustering analysis to cluster the groups based on their placement in the PCA. Lastly, we classify the clusters based on the metrics presented in 3.4.

6.1 Principal Component Analysis

The dataset explored in this quantitative study is based on the case study. The dataset comprises the user action types described in 3.4. These variables are observed for each participating group. The observed variables in the dataset consist of eleven dimensions, each representing a user action type. Determining whether some groups have similar behavior can be challenging based on multiple variables.

One way of exploring and interpreting such datasets is with Principal Component Analysis, a technique to reduce dimensionality while minimizing information loss. By reducing the dimensionality, the PCA increases the interpretability of the data[18, p. 1]. PCA serves to produce linear combinations of the original dimensions to generate new axes, which are called the *principal components* (PC).

The first principal component is given by a linear combination of dimension D_1 to D_{11}:

$$PC1 = a_1 D_1 + a_2 D_2 + \cdots + a_{11} D_{11}$$

Such that it accounts for the most variance in the dataset, but with the constraint that

$$a_1^2 + a_2^2 + \cdots + a_p^2 = 1$$

to not over-fit any given variable[15].

6.2 Performing the PCA

To perform the PCA, we have used the Python library *sklearn*, which includes a module to quickly and efficiently explore a dataset[10]. The implementation for our PCA can be found in Appendix D.

To interpret the results of a PCA, we need to determine how many principal components to examine. To determine the number of PCs to use in our exploration of the data, we can calculate the total variance explained for each PC[15].

Figure 6.1 shows the principal components on the x-axis and the variance explained for each on the y-axis. The blue line represents the average variance explained. We see that PC1 explains over 40% of the variance in the dataset.

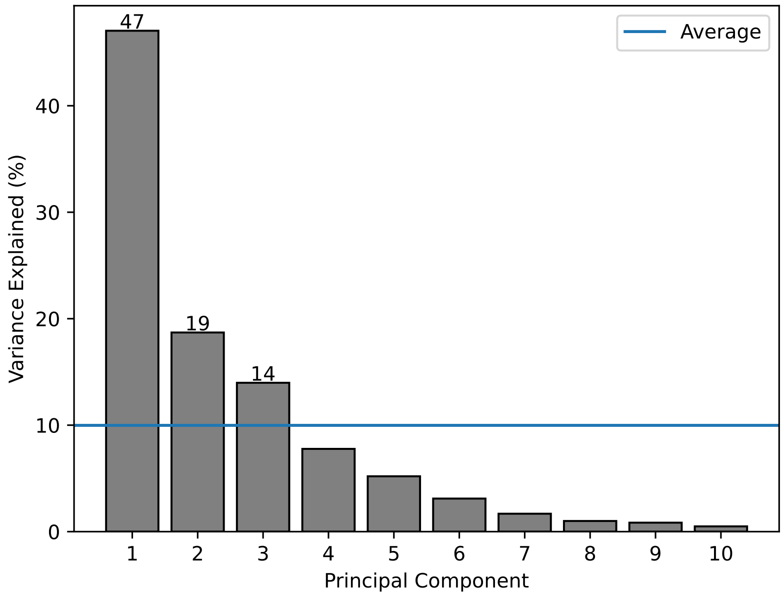


Figure 6.1: Scree plot

To determine the appropriate number of PCs for our PCA, it is common to use a predefined percentage of total variance explained. A common practice is to use 70% of total variability as a cut-off point [18, p. 4].

By interpreting Figure 6.1, we see that the first three PCs provide a variability above 70%. Next, we want to understand how each of the eleven dimensions contributes to the three PCs. The impact of the dimensions is revealed by the loadings presented in Table 6.1. The loadings describe the correlation between the dimensions and the PCs. A positive loading corresponds to a positive correlation, and a negative loading corresponds to a negative correlation. Significant positive and negative loadings indicate that a given variable has a strong effect on the PC.

	D2	D10	D9	D3	D4	D8	D7	D6	D5	D1	D11
PC1	**0.389**	**0.374**	**0.358**	**0.356**	0.225	0.225	0.178	0.261	0.293	0.291	0.286
PC2	-0.159	-0.085	-0.095	-0.263	**0.508**	**0.424**	**0.584**	0.168	-0.084	-0.193	-0.195
PC3	-0.086	0.133	-0.376	0.003	-0.239	0.315	-0.161	**0.543**	**-0.492**	**0.306**	0.147

Table 6.1: Loadings for PC1, PC2 and PC3

In Table 6.1 the large loadings are bolded. We can interpret that dimensions 2, 3, 9, and 10 positively correlate to PC1. These are the number of updating, deleting, and uncertain statements and the number of sessions. We interpret dimensions 4, 7, and 8 positively correlated with PC2. These are creating relations and the number of standard and future statements. Lastly, PC3 correlates with dimensions 1, 5, and 6, which are the number of inserted and deleted statements and the number of look-ups in the wiki.

To visualize the correlations between dimensions and PCs, we have created a distance biplot presented in Figure 6.2. The dots represent prospects, and the vectors represent the loadings of each dimension. The loadings dictate the position of each prospect in the three-dimensional space. The prospects are colored to indicate their placement in the three-dimensional space further. To illustrate the depth in the graph, we use a darker green color to represent a high value on the PC2 axis, whereas a lighter green color represents a low value.

To better understand the placement of the prospects, we have performed a clustering analysis using the k-means algorithm[9]. To determine the appropriate number of clusters to be used in the k-means algorithm, we use the elbow method[33]. Following this method, we obtain the graph presented in Figure E.1, from which we can determine the turning point of the curve at 5 clusters. The clustering analysis results are illustrated in Figure 6.3, where a unique color represents each cluster.

Following a brief investigation of the prospects within each cluster, we have determined that the prospects within the yellow cluster should be discarded. The common denominator for these prospects is that they have very little activity in Digital Essence. To further explore the prospects in the remaining four clusters, we selected a prospect from each and analyzed their metrics. The four selected prospects are marked in Figure 6.3.

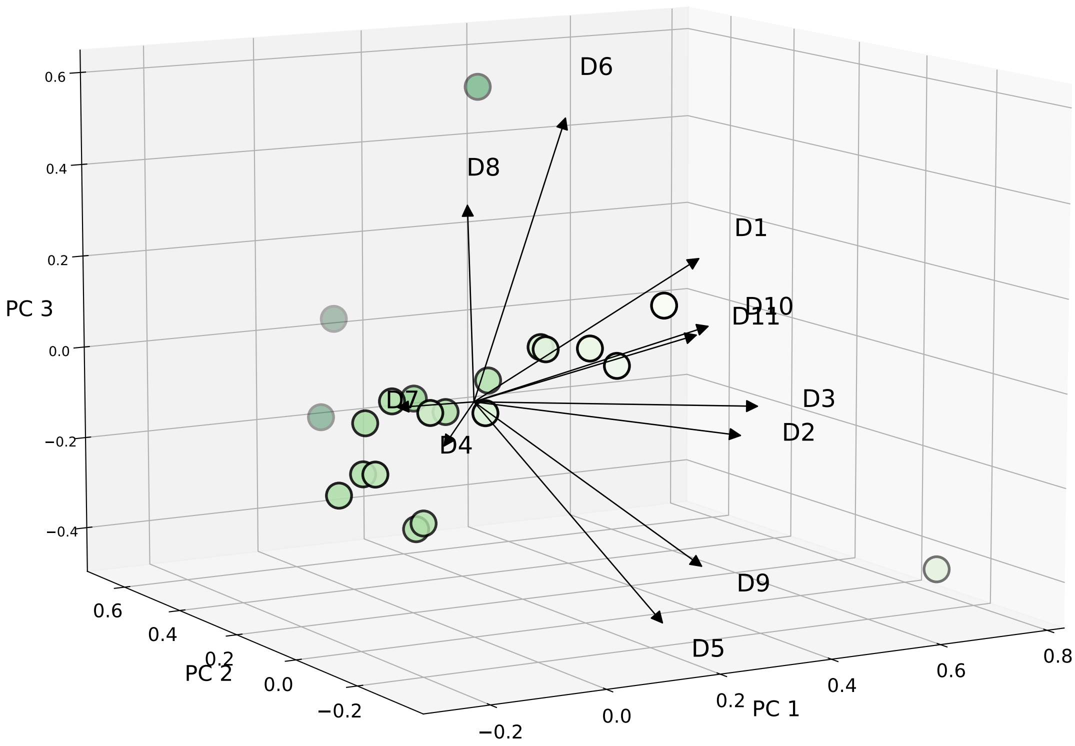


Figure 6.2: Distance Biplot

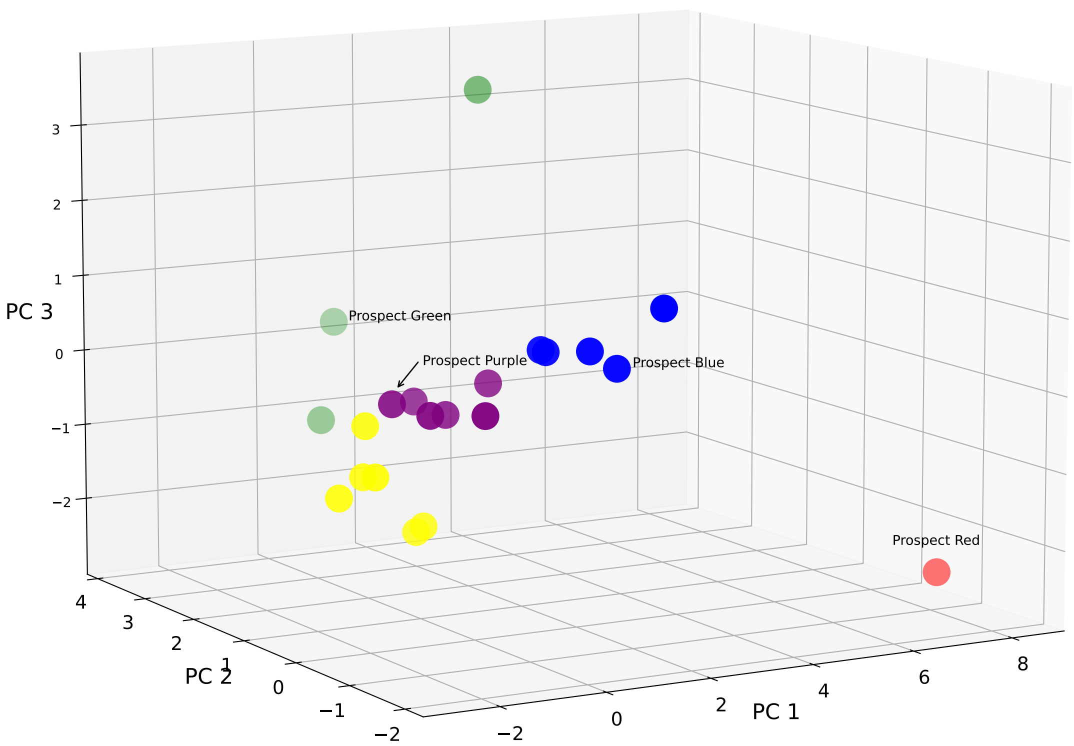


Figure 6.3: Clustering of prospects

6.3 Results of the Quantitative Study

The PCA and clustering revealed five clusters based on the data gathered in the case study. As previously mentioned, only four of the clusters will be further examined. We have selected a prospect from each to identify and characterize the behavior of these clusters. These four prospects are evaluated based on the metrics presented in section 3.4.

Table 6.2 presents the metrics for each of the four prospects. We can identify different behaviors and see that the prospect's usage of the guiding features and Digital Essence generally varies.

	M1	M2	M3	M4	M5	M6	M7
Blue	100%	98%	(91%, 0, 9%)	0	5	54.9%	0%
Red	46.9%	98%	(78%, 15%, 7%)	0	8	45%	58.9%
Green	5.56%	16%	(96%, 2%, 2%)	0	3	73.9%	100%
Purple	100%	25%	(67%, 13%, 20%)	0	2	72.7%	0%

Table 6.2: Metrics from the four prospects

Based on the metrics presented in Table 6.2, we can summarize the behavior of the four prospects in Table 6.3.

	Standard stmt w/o relations	Follows pattern	Statement classification	Change-advis	Sessions	Statements: Insert vs Modify	Relationships: Insert vs Delete
Blue	Many	Exhaustively	Little usage	No usage	Many	Even ratio	None created
Red	Some	Exhaustively	Medium usage	No usage	Many	Even ratio	Created & deleted
Green	Few	Little	Little usage	No usage	Few	Primarly inserts	Only created
Purple	Many	Little	High usage	No usage	Few	Primarily inserts	None created

Table 6.3: Summarization of the metrics for the four prosepcts

6.4 Interpretation of the Results

Based on the classification of the behavior of the different groups in Table 6.3, we can identify four categories with different usage patterns:

- **Documentation**. Groups that primarily insert statements in Digital Essence. It can be seen as the first and lowest level of reflection. Here the users only use the configuration table as a common place for project documentation. These groups have a few sessions in Digital Essence, where they insert their current knowledge gathered elsewhere. This category includes prospect Purple.
- **Interconnected documentation**. Groups with similar behavior to *Documentation*, but which also establishes relationships between their statements. It can be seen as the second level of reflection. These groups also use Digital Essence as a documentation tool and document their project's interconnectedness. This category includes prospect Green.
- **CRUD**[1]. Groups that both insert, update, and delete statements in Digital Essence. It can be seen as the third level of reflection. The users use the configuration table to explore and reflect on their problem and solution space. These groups have more sessions than the previous categories, where they update the configuration table with existing knowledge and use the patterns to discover new knowledge and ensure the internal consistency of their project. This category includes prospect Blue.
- **Interconnected CRUD**. Groups that actively try to reduce and reflect upon their dangling information in Digital Essence. It can be seen as the last and highest level of reflection. These groups are an extension of *CRUD*, where they use the configuration table to explore and reflect on their problem and solution space. They validate and explore their project by following the patterns and explicitly relating statements to each other. This category includes prospect Red.

[1] Create, Read, Update and Delete

7 | Qualitative Study

This chapter presents our qualitative study, investigating if the usage categories identified in chapter 6 impact how the groups using Digital Essence reduce dangling information (RQ2). Additionally, we investigate if the pursuit of reducing dangling information improves inquiry (RQ3). This study consists of focus group interviews (see section 3.2), with two groups selected from different categories.

7.1 Choice of Groups

We have selected groups from different categories to gain further insights into user behavior and reflection. The categories we are investigating are *CRUD* and *Interconnected CRUD*, as they are the most reflective categories. Group Blue is in the *CRUD* category and consists of five people, and Group Red is in the *Interconnected CRUD* category, consisting of two people[13, p. 34].

7.2 Interview Guide

We have chosen to divide the interview into three phases. The interview guide is presented in Appendix A. The left column shows the thematics of the interview, and the right column shows the questions.

The first phase is briefing the informants and asking preliminary questions to prepare the participants. The preliminary questions focus on contextualizing the group's project. The first phase helps us gain some insights into their previous experience with tools similar to Digital Essence.

The second phase of the interview consists of an assignment, where the participants are asked to continue the work on their configuration table in Digital Essence. The participants are asked to reflect upon their configuration table and if any statements should be revised or related. This phase is used to observe how the participants are working with Digital Essence[13, p. 38].

The third phase of the interview consists of questions investigating RQ2 and RQ3. Here the participants are asked questions regarding their reflection process, how they approach and use Digital Essence, and its effect on their teamwork. In addition to how they have used the guiding features. The interview ends with an opportunity for the students to give feedback on Digital Essence.

7.3 Results of the Qualitative Study

The results of the two focus group interviews can be found in Appendix B and Appendix C. Both groups have experience with project management tools, such as SCRUM, and tools to explore a project and possible solutions, such as wireframes and the golden circle (why, how, what).

7.3.1 Key takeaways from the interview with Group Blue

- (RQ2) Found Digital Essence most valuable at the beginning of their project, for their brainstorming session, and by the end of their project, to fact-check against the information they had written down. They found that the twelve cells and their relation in the configuration table *forced* them to reflect upon and evaluate different aspects of their project, which they otherwise would not have.
- (RQ2) During the assignment, they reviewed their existing statements in the configuration table in no specific order.
- (RQ2) The guiding features *relationships between cells* and *guiding documentation* proved most useful for them. They combined these two guiding features to ensure the validity of the statements they created and better understand the configuration table and its terminology.
- (RQ2) Did not find immediate value in establishing relationships between statements. They needed help determining the appropriate time to establish these relationships. For this group, the *relationships between cells* and *guiding documentation* were enough in terms of ensuring the internal consistencies of their project and thus reducing what they thought to be dangling information.
- (RQ3) During the assignment, the group identified a point of interest where they were uncertain. They found a statement in the configuration table that was irrelevant to their scope of work (dangling information). By evaluating their statements in the configuration table, they identified a piece of dangling information, which led to further inquiry into their delimitations.
- (Potential new guiding feature) Found it difficult to determine an appropriate entry/starting point in the configuration table. When the participants were met with twelve empty cells, they found it challenging to get started. The participants want to be guided toward a more appropriate entry point based on their project type.

7.3.2 Key takeaways from the interview with Group Red

- (RQ2) Used the configuration table to maintain an overview of their project and as a medium to communicate and explore different ideas. They found that the configuration table challenged their ideas and proposed solution, as they were forced to review different aspects according to the different cells and their relationships in the table. They concluded that Digital Essence helped eliminate assumptions and explore tacit knowledge in their project.

- (RQ2) Actively used the classification of statements in their configuration table. They chose to move all *future* classified statements to a new prospect to gain an overview of the future of their project.

- (RQ2) The guiding features *relationships between cells*, *relationships between statements* and *guiding documentation* proved most useful for them. They combined all three guiding features to properly understand the configuration table and ensure the internal consistency of their project.

- (RQ2) The relationships between statements helped them reflect upon the interconnectedness of their project. However, they needed help determining when to establish and evaluate relationships between their statements. They were divided between creating relationships and statements simultaneously or sequentially.

- (RQ3) During the assignment, they evaluated the configuration table according to the amount of dangling information in each cell. They identified parts of the configuration table where they had inconsistencies and where they were in disagreement. This disagreement led to a discussion of certain aspects of their project, including to whom their contribution was relevant.

- (Potential new guiding feature) Found it difficult to determine an appropriate entry/starting point in the configuration table. The participants want to be guided toward a suitable entry point based on their project type.

7.4 Interpretation of the Results

The focus group interviews show that the usage category impacts how they identify, reflect, and reduce dangling information. Our summary and findings from the focus group interviews with group Red and Blue have been validated through what some call ecological validation[3, p. 106]. Both groups have agreed that the summaries reflect the discussions from the interviews and the usage category they have been placed in.

For RQ2, we can interpret that the two groups work with reducing dangling information in different ways. Group Blue focuses on reducing dangling information when they insert new statements by ensuring the new statements' validity. Here they use the relationships between cells to properly understand how the statements are related based on the cells. They also use the guiding documentation to ensure the purpose and terminology of each cell. This group followed no specific order for reviewing their statements; however, when they reviewed statements in a specific cell, they used the cell's relationships with other cells to validate them.

Group Red also focuses on reducing dangling information when they insert statements, similar to Group Blue. However, Group Red used the relationships between their statements to identify dangling information. When reviewing their configuration table, they identified cells with a high number of dangling information and then proceeded to review the statements of these cells.

For RQ3, we have identified that both groups experienced an improvement in their inquiry into their projects after identifying dangling information. Group Blue found that they had a dangling piece of information in their configuration table, which led to them realizing they had to revise the written material in their project. Group Red found that by reviewing dangling information, they could more precisely specify their target audience and thus improve inquiry.

To further investigate RQ2 and RQ3, it would be relevant to conduct focus group interviews with groups in the *Documentation* and *Interconnected Documentation* categories to explore how groups in these categories work with dangling information and its impact on their reflection process.

Both groups experienced difficulties determining an appropriate starting point in the configuration table. We see this as a potential need for a new guiding feature, which aids groups in getting started with the configuration table.

8 Discussion

This chapter discusses our adopted research method and its impact on our ability to answer the research questions fully. It further discusses our approach for identifying and classifying the usage categories of Digital Essence, using user actions and metrics, and lastly, Essence as the basis for our research into guiding features and dangling information.

8.1 Research Method

The research presented in this thesis studies the impact of dangling information on reflection and inquiry in Visual Inquiry Tools. To investigate this, we have adopted the Design Science Research (DSR) approach, which is well-suited to our research objectives. DSR offers a systematic and structured approach to designing, building, and evaluating IT artifacts. This approach allows us to contribute to VIT design knowledge. Specifically, we contribute design knowledge of how VITs can implement guiding features to help users reduce and reflect on dangling information.

Our research method is derived from within the field of DSR from a study that has a substantial similarity to ours. We derived five phases that constitute our method. Each phase includes relevant elements of the DSR framework[6], namely *Relevance* from the *Environment*, *Rigor* from the *Literature* and using theory to *Design* and *Evaluate* our IT artifact Digital Essence. This framework makes DSR suitable for the research presented in this thesis. However, another approach also used when developing IT artifacts is Action Research (AR). AR is a research approach designed to establish a connection between actions and problem-solving[8]. AR involves researchers and practitioners cooperating to identify problems and improve systems and processes within a specific context. This practice is done through a collaborative and participatory approach. AR aims to understand and improve the practices, processes, and systems to drive practical change in a real-world scenario. The AR approach could have been used to address RQ2 & RQ3 since it investigates the impact of the features in Digital Essence on the user's reflection and inquiry. However, this thesis focuses on designing and evaluating the guiding features of Digital Essence, and the DSR approach emphasizes evaluating the effectiveness and utility of IT artifacts. Thus, the DSR approach aligns more closely with our research objectives.

The specific research method chosen in this thesis consists of five phases, each contributing to address the research questions in chapter 1.

8.1.1 Problem Awareness

The problem awareness phase investigates the relevance of dangling information in VITs. In this phase, we reviewed the Essence configuration table as a VIT and found some inconsistencies with the VITs design principles. We used the term *dangling information* to describe information in a VIT that has no relationships to other information. To validate the relevance of our findings, we investigated dangling information in other VITs. We concluded that these VITs suffer the same issue as the configuration table, where the model ontology only implicitly describes relationships between elements. We presented Business Model Canvas and Value Proposition Canvas as two VITs allowing users to have dangling information.

We chose to investigate two similar canvas-based VITs, of the same authors. Both BMC and VPC are very similar in their structure and use. Even though these tools are widely used, the lack of dissimilarities can limit our contribution's ability to be generalized and, thus, its relevance. Alternatively, we could have investigated VITs of different domains and authors. However, we believe that other canvas-based VITs will have a similar structure and arrangement as the VITs we investigated and the configuration table.

8.1.2 Suggestion

This phase comprises the foundation of our study by establishing the requirements of the guiding features used to reduce dangling information. The requirements are based on CASE tool theory and VIT design principles[35, 4]. We proposed four guiding features (see chapter 2). Each feature was created assuming that it will help users reduce dangling information and that the pursuit of reducing dangling information will improve inquiry. The literature used to design the guiding features is based on a limited selection of research and is somewhat dated as CASE tools are not studied as frequently. However, the CASE tools theory aligns well with our research objectives. The CASE tool theory provides a guided philosophy that fits well with the flexible nature of VITs and provides valuable knowledge about how such guiding features can be created.

8.1.3 Development

The guiding features are aimed at Digital Versions of VITs. Because of this, to investigate RQ1 and, in turn, RQ2 and RQ3, we needed to develop a digital version of the configuration table and implement the guiding features.

As we wanted to investigate the usage patterns of the guiding features in Digital Essence through a case study, we needed to invest time and resources into making the system production-ready. This effort included ensuring the robustness of the system and its ability to accommodate concurrent users. Additionally, we implemented CI/CD pipelines, enabling us to quickly resolve any errors that may arise during the case study. These improvements were made to ensure that the performance of Digital Essence had no negative impact on the case study. The time spent on getting the system production-ready has not directly impacted our research and our ability to answer the research questions. However, these investments were necessary not only for the success of the case study but also as the students relied on Digital Essence as part of their coursework.

8.1.4 Evaluation

Our evaluation strategy is influenced by the design options presented in Table 3.1. These design options constitute different evaluation strategies used for various VITs. We performed a case study and focus group interviews to address our research questions.

The case study allowed us to investigate RQ1 by gathering quantitative data, which we explored through PCA. We identified four usage patterns, which used different combinations of the guiding features. To answer RQ2 and RQ3, we conducted focus group interviews with two groups in different clusters in the PCA analysis. In these interviews, we explored the reflection and inquiry process of two groups (RQ3) and how they used the guiding features to reduce dangling information (RQ2). We encountered difficulty with the data generated through the case study because we could not determine the whole truth. We cannot determine the group's thought process behind their decisions based on the data.

Additionally, we realized that at the beginning of the case study, not all groups were familiar with all the implemented guiding features and the value they might provide. This lack of knowledge could mean that some of the groups' behavior in Digital Essence is more shallow than it could have been if they knew all features from the beginning. A potential solution to this problem could be to conduct a workshop before the case study, in which the participants are introduced to the implemented guiding features and their purpose. A workshop could also provide insight into the participant's thoughts on Digital Essence, both on the user interface and their user experience.

8.1.5 Conclusion

In chapter 6, we identified four usage patterns based on the PCA (RQ1). As discussed above, our findings in the PCA could have been different if we had performed a workshop introducing the guiding features before the case study. Through the focus group interviews, we investigated whether the identified usage patterns influenced the group's work with dangling information (RQ2). We discovered that the two groups used different combinations of guiding features to reduce dangling information. Group Blue explored their prospect's internal consistencies by ensuring the statements' correctness through the *connection between cells* and *guiding documentation*. Group Red used the same guiding features as Group Blue and explicitly defined the relationships between their statements. This distinction means that the usage categories *CRUD* reduces dangling information implicitly, whereas *Interconnected CRUD* reduces it explicitly.

To further investigate RQ2, we should conduct focus group interviews with group Green and Purple. These additional interviews can show how the groups belonging to *Documentation* and *Interconnected Documentation* reduce dangling information. We could also further enrich our examination of RQ2 by evaluating multiple prospects from each usage pattern.

8.2 Metrics

To address RQ1, we first defined eleven dimensions. These dimensions constitute different user actions and are used to identify usage patterns in Digital Essence. Secondly, we established eight metrics providing context for the user actions to classify these patterns. These metrics were selected based on assumptions about the user's behavior in Digital Essence.

The metrics have been a good entry point for classifying the usage patterns of Digital Essence. They have provided us with insights into how the participants use the platform. However, some metrics are inadequate. It has been challenging to determine user behavior programmatically. An example of this is the identification of statements inserted according to a pattern (M2). Our method of detecting this behavior does not consider that the inserted statements may not be immediately related. When studying this behavior, one can either be strict and insist on inserting information strictly based on a pattern or adopt a more flexible approach, taking into account all the affected cells. In both cases, there is a risk of losing some precision.

We proposed one metric for measuring dangling information in prospects (M1). This metric is, however, limited to only being relevant for groups using *relationships between statements*. It is based on the assumption that users of Digital Essence establish relationships between their statements, allowing us to determine which statements are dangling. By answering RQ1, we identified that not all groups use the guiding feature *relationships between statements*. Identifying better quantitative measuring alternatives is challenging because dangling information is subjective to a specific group and their understanding of a project. Following this discovery, a suitable approach for gaining insights into dangling information for a given prospect would be to conduct more qualitative measures such as further focus group interviews.

Through our focus group interviews, we have deduced that our guiding feature, *Guiding Documentation*, was favored and well-utilized among the participants. However, we do not have a metric to enrich the classification of the usage patterns of this feature. The reason for the lack of metrics regarding the documentation is that it is difficult to measure time. We want to investigate how much time users spend on the documentation and how often they access different pages. However, too many potential confounding factors are associated with these metrics, so we have not included them. Examples of such factors are users forgetting to close the documentation after use.

Following the results of the metrics presented in Table 6.2, we see that none of the selected prospects have conducted any changes when encountering warnings (M4). This lack of use could indicate that none of the groups have used the guiding feature *change-advis*. Alternatively, it could indicate that the groups resolve warnings before making any changes or that most changes are grammatical errors or rephrasing.

8.3 Essence

We have chosen Essence's configuration table for our case study, analysis of guiding features, and their impact on dangling information and reflection in VITs. The advantages of selecting Essence are that through the Software Innovations course, we have access to students whom we can use in the case study. Additionally, our supervisor is the author of Essence, which has provided us with insights into how the configuration table is constructed and intended. However, the configuration table is less widely used than the Business Model Canvas. The configuration table's maturity and audience size could make our findings difficult to generalize. It would be beneficial to expand our work by investigating the effectiveness of our guiding features in other VITs to strengthen our results and improve generalizability.

9 Conclusion

This thesis explored the concept of dangling information, presented a series of guiding features, and investigated usage patterns and the impact of guiding features on reducing dangling information in Digital Essence.

(RQ1) What usage patterns are associated with our guiding features?

In conclusion, this research investigated the usage categories associated with Digital Essence guiding features. Through quantitative data analysis using Principal Component Analysis and clustering, we identified the following usage categories: *Documentation*, *Interconnected documentation*, *CRUD*, *Interconnected CRUD*. These categories span from documentation usage, where users primarily use Digital Essence as a documentation tool, to reflective usage, where Digital Essence is used for joint inquiry.

(RQ2) Do the usage patterns influence how dangling information is reduced?

Based on the quantitative study, the identified usage categories work with dangling information differently by using different combinations of the guiding features. Based on the qualitative study, consisting of focus group interviews with groups from the *CRUD*, *Interconnected CRUD* usage categories, we found that these categories work with reducing dangling information, but *Interconnected CRUD* does so in an explicit manner, whereas *CRUD* does it implicitly.

(RQ3) Does reducing dangling information help users during inquiry?

From our focus group interviews, we identified that both groups experienced an improvement in their inquiry after identifying potential dangling information. Both groups experienced that identifying dangling information resulted in helpful discussions, which improved inquiry into their problem and solution space.

We have not been able to validate the impact of *change-advis* on dangling information in our case study and focus group interviews. This lack of evidence leads us to not propose it as part of the guiding features in this study. We believe further research into the *change-advis* feature is necessary to conclude its relevance.

Addressing the generalizability of the proposed guiding features, we have identified a series of conditions that must be met for a given VIT. For the *guiding documentation*, a VIT should enforce some degree of model ontology, which impacts the tool's usage. For *relationships between statements* and *relationships between cells*, the model ontology should include relationships between the elements in the canvas. These three guiding features are applicable in the case of BMC and VPC investigated in this thesis. Both VITs contain a model ontology that includes relationships between the elements.

In summary, the study identified distinct usage categories and found that reducing dangling information helps users in inquiry. However, the relevance of the "change-advis" feature requires further research. These findings improve the design and implementation of digital visual inquiry tools by making the model ontology more explicitly available for the users.

10 | Future Work

This chapter presents a series of future works worth pursuing. Firstly, a new guiding feature from the focus group interviews is described. Secondly, a suggestion for improving our measure for dangling information is described by considering the timeliness and age of statements in Digital Essence. Lastly, a study of why groups use VITs as documentation tools rather than tools of inquiry is outlined.

10.1 Extending the Guiding Features

The focus group interviews revealed a need for a new guiding feature, aiding users in determining a good place to begin in the configuration table. Both groups needed help determining an appropriate starting point for their innovation projects. In [29], the author suggests different innovation projects influenced by technological development or potential new markets (technology-push or market-pull). These innovation projects are helpful when determining the point of entry in the configuration table. As an example, a *technology push* innovation project could start in the *contribution* or *solution* columns of the configuration table. In contrast, a *market pull* project could start in the *problem* column.

10.2 Outdated Information

To further explore the concept of dangling information, future research could examine the concept of outdated information. This thesis defines *dangling information* as "components that are unclearly labeled or not correctly related." However, outdated information could also be considered dangling as it is no longer valid for a project.

The focus group interviews revealed that both groups examined their existing statements when asked to continue the work on their configuration tables. Group Red reviewed their existing statements without relationships, whereas Group Blue adopted an ad-hoc approach. We believe a suitable approach for reviewing existing statements in the configuration table for all four usage categories could be evaluating them according to their creation time. This approach would suggest reviewing the oldest statements first to ensure their relevance for the project.

10.3 Further Focus Group Interviews

In this thesis, we only conducted focus group interviews with groups from the *CRUD* and *Interconnected CRUD* usage categories. We believe it could be beneficial to interview groups from the *Documentation* and *Interconnected Documentation* categories, to investigate their work with dangling information. It could also be interesting to investigate why some groups use VITs only as documentation tools rather than for inquiry to determine an approach for cultivating them to use VITs in the manner they are intended. Thus, try moving them towards the *CRUD* and *Interconnected CRUD* categories.

Bibliography

[1] I. Aaen. *Essence - Problem Based Digital Innovation (Unpublished book draft)*. Aalborg: Department of Computer Science, Aalborg University, 2023.

[2] Strategyzer AG. *Value Proposition Vancas - Official Template (Video)*. `https://www.strategyzer.com/canvas/value-proposition-canvas`. 2022.

[3] Lotte Bøgh Andersen, Kasper Møller Hansen, and Robert Klemmensen. *Metoder i statskundskab*. Rosinante&Co., 2010.

[4] Hazbi Avdiji et al. "A design theory for visual inquiry tools". In: *Journal of the association for Information Systems* 21.3 (2020), p. 3.

[5] Hazbi Avdiji et al. "Designing tools for collectively solving ill-structured problems". In: (2018).

[6] Jan vom Brocke, Alan Hevner, and Alexander Maedche. "Introduction to design science research". In: *Design science research. Cases* (2020), pp. 1–13.

[7] Susan M. Carter and Michael A. West. "Reflexivity, Effectiveness, and Mental Health in BBC-TV Production Teams". In: *Small Group Research* 29.5 (1998), pp. 583–601. DOI: `10.1177/1046496498295003`. eprint: `https://doi.org/10.1177/1046496498295003`. URL: `https://doi.org/10.1177/1046496498295003`.

[8] Dalila Cisco Collatto et al. "Is Action Design Research Indeed Necessary? Analysis and Synergies Between Action Research and Design Science Research". In: *Systemic Practice and Action Research* 31.3 (2018), pp. 239–267. DOI: `10.1007/s11213-017-9424-9`. URL: `https://doi.org/10.1007/s11213-017-9424-9`.

[9] scikit-learn developers. *K-means*. `https://scikit-learn.org/stable/modules/clustering.html#k-means`. 2023.

[10] scikit-learn developers. *sklearn.decomposition.PCA*. `https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html`. 2023.

[11] John Dewey. "Logic: The theory of inquiry, 1938". In: (1999).

[12] Dina Elikan and Yves Pigneur. "A visual inquiry tool for brand identity". In: (2019).

[13] Bente Halkier and Sven-Erik Torhell. *Fokusgrupper*. Liber Stockholm, Sweden, 2010.

[14] Hattel, K., Winkler, K. S., Thomassen, M. S. E. "Digital Essence". 9th semester project report. Dept. of Computer Science, Aalborg University, 2023. URL: https://projekter.aau.dk/projekter/files/512499066/trioen_cs_22_sd_9_04.pdf (visited on 02/28/2023).

[15] Steven Holland. *Principal Components Analysis - GEOL 8370*. http://strata.uga.edu/8370/lecturenotes/principalComponents.html. 2023.

[16] Robert E Horn and Robert P Weber. "New tools for resolving wicked problems". In: *Mess Mapping and Resolution Mapping Processes* (2007).

[17] Juhani Iivari and John R Venable. "Action research and design science research-Seemingly similar but decisively dissimilar". In: (2009).

[18] Ian T Jolliffe and Jorge Cadima. "Principal component analysis: a review and recent developments". In: *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences* 374.2065 (2016), p. 20150202.

[19] Project Jupyter. *Project Jupyter | About Us*. https://jupyter.org/about. 2023.

[20] THORING Katjaab, Roland M MUELLER, and BADKE-SCHAUB Petraa. "Exploring the design space of innovation canvases". In: *Transformations* (2019), p. 593.

[21] Udo Konradt et al. "Reflexivity in teams: A review and new perspectives". In: *The Journal of psychology* 150.2 (2016), pp. 153–174.

[22] Grafana Labs. *Grafana | Query, visualize, alerting observability platform*. https://grafana.com/grafana/. 2023.

[23] Frederik Möller and Barbara Steffen. "'License to VIT'-A Design Taxonomy for Visual Inquiry Tools". In: *Proceedings of the 55th Hawaii International Conference on System Sciences*. 2022.

[24] Alexander Osterwalder. "The business model ontology a proposition in a design science approach". PhD thesis. Université de Lausanne, Faculté des hautes études commerciales, 2004.

[25] Alexander Osterwalder et al. "Business Model Generation: A handbook for visionaries, game changers and challengers". In: *African journal of business management* 5.7 (2011), pp. 22–30.

[26] Alexander Osterwalder et al. *Value proposition design: How to create products and services customers want*. John Wiley & Sons, 2015.

[27] Ken Peffers et al. "A design science research methodology for information systems research". In: *Journal of management information systems* 24.3 (2007), pp. 45–77.

[28] Horst WJ Rittel and Melvin M Webber. "Dilemmas in a general theory of planning". In: *Policy sciences* 4.2 (1973), pp. 155–169.

[29] Edward B Roberts. “What we’ve learned: Managing invention and innovation”. In: *Research-Technology Management* 31.1 (1988), pp. 11–29.

[30] Thorsten Schoormann, Julien Hofer, and Ralf Knackstedt. “Software tools for supporting reflection in design thinking projects”. In: (2020).

[31] Strategizer. *Building Blocks of Business Model Canvas*. `https://www.strategyzer.com/business-model-canvas/building-blocks`. 2023.

[32] Strategyzer. *Team Alignment Map*. `https://www.strategyzer.com/resources/canvas-tools-guides/team-alignment-map`. 2022.

[33] Robert Thorndike. “Who belongs in the family?” In: *Psychometrika* 18.4 (1953), pp. 267–276.

[34] John Venable, Jan Pries-Heje, and Richard Baskerville. “FEDS: a framework for evaluation in design science research”. In: *European journal of information systems* 25 (2016), pp. 77–89.

[35] Iris Vessey, Sirkka L Jarvenpaa, and Noam Tractinsky. “Evaluation of vendor products: CASE tools as methodology companions”. In: *Communications of the ACM* 35.4 (1992), pp. 90–105.

[36] M.A. West. “Reflexivity, revolution and innovation in work teams”. English. In: *Product development teams*. Ed. by M. Beyerlein, D. Johnson, and S. Beyerlein. Advances in interdisciplinary studies of work teams. United States: JAI Press, 2000, pp. 1–29. ISBN: 0-7623-0360-3.

A | Interview Guide

Briefing	• **Samtykke** optagelse, anonymisering og tilbagetrækning af deltagelse - ingen rigtige eller forkerte svar. • **Ikke en evaluering af Essence generelt.** • **Præsentation af os** Specialestuderende på Software. Vi har udviklet Digital Essence som en del af vores speciale.Formålet med projektet: at udforske, hvordan nogle funktioner i digitale værktøjer kan øge refleksion og 'inquiry'. • **Tidsrammen** en times tid - er der en bagkant, vi skal være opmærksomme på?
Tematikker	**Interviewspørgsmål**
	Kan en af jer kort præsentere gruppen og jeres projekt?
	(*Så har vi en idé om hvad deres projekt indebærer..*)
Indledende spørgsmål	Har I erfaringer med lignende værktøjer til at undersøge problemer?
	Hvordan finder I frem til om I er på vej den rigtige retning i jeres projekt?

Opgave: Vi vil gerne spørge jer lidt ind til jeres oplevelse med Digital Essence og hvordan I har brugt konfigurations tabellen i kurset. Vi vil gerne tage udgangspunkt i jeres projekt. Nu har vi en opgave til jer; I skal kigge på jeres konfigurations-tabel her på min computer - I får ca. 10 min. til at arbejde på den - skal der være flere forbindelser? Skal noget rettes? Skal noget slettes? Osv.

	- Nu har I lige siddet med værktøjet, hvordan er det at bruge?
Brugen af Digital Essence	- Kunne I bruge værktøjet nemt og effektivt? Hvis ikke, hvad var udfordringerne, du stod over for?
	- Hjalp værktøjet jer med at organisere jeres tanker eller ideer bedre?
	- *Hvis ja, kan I give et eksempel?*
	- Hjalp værktøjet jer med at få udfyldt tabellen?
Refleksion (RQ2)	- Det er et nyt værktøj, jeg ikke har brugt før på et projekt. Nu har I prøvet at bruge det - får det jer til at reflektere mere eller mindre over jeres projekt?
	- *Kan I fortælle os mere om den refleksionsproces uden Digital Essence?*
	- Hvordan understøtter eller hindrer Digital Essence jeres evne til at reflektere over information i jeres projekt?
	- Har der været nogle situationer hvor I har været usikker på hvordan information skulle forbindes?
Dangling Information (DI)	- *Eksempel?*
Reduktion af DI	- *Hvordan kom I videre? (hjælp)*
	- Har I brug for i jeres projekt at lave forbindelser mellem information?
	- Når I laver nye "statements", hvordan ved I så hvad de skal forbindes til?
	- Laver "statements" uden at forbinde dem til andre med det samme?
	- *Vi kan se i data, hvad de gør, men hvorfor?*
	- Fandt I, at relationerne mellem cellerne og statements fik jer til at reflektere over jeres projekt?

Inquiry (RQ3)	- Er der nogle steder I bruger future eller uncertain til at klassificere jeres statements? - *Hvorfor?* - Hvornår synes I at jeres konfigurations tabel er god? - Nu skal I være helt ærlige, kommer I til at bruge Digital Essence mere frem mod aflevering af jeres miniprojekt? - *I hvilke stadier af projektet er Digital Essence mest relevant for jer?* - Hvordan påvirker det jeres gruppearbejde, at nogen har SWI og bruger Digital Essence og andre ikke gør? *Var der nogen specifikke funktioner i værktøjet, som I fandt nyttige eller som I gerne vil se forbedret?*
Afrunding af interview	
Opsummerende: **- Er der andre tanker eller feedback, som I gerne vil dele?** **- Vi vil gerne takke jer for jeres tid og fordi I deltog i fokusgruppen.**	

B | Focus Group Blue

B.1 Part 1: Introduction

The group has previously used tools such as golden circle (why, how, what) and SCRUM boards.

B.2 Part 2: Assignment

The group evaluates the statements in the different cells. There is no specific order in which the cells and 'statements' are considered. During the assignment, the participants discuss when to create relationships between statements. They are divided between establishing relationships once they know the validity of the 'statements' or along the way when making 'statements.' The group did not make any relationships between statements in this assignment. The participants used the documentation throughout the assignment to clarify different terms.

In the assignment, the participants are reflecting upon the manifestations of their problem, where they realize that they are not in agreement about certain delimitations in their project. One person thought they made a delimitation about not supporting a particular capability (in their project), whereas the others realized they had forgotten to write about it. By evaluating their statements in the configuration table, they identified a piece of dangling information, which led to further inquiry into their delimitations.

B.3 Part 3: Follow-up questions

They found that having a dedicated user interface for the configuration table made it easier and more attractive. According to the participants, Digital Essence is suitable for reflecting upon their project. By filling out the cells, they were forced to reflect on different aspects of their project, which they otherwise would not have. Additionally, they found that Digital Essence is a source of truth for the various topics discussed throughout the project. They thought that Digital Essence brought the most value at the beginning of a project, by aiding in the brainstorming and reflection process, and at the end of the project, for fact-checking.

The guiding features they used most in their project are guiding documentation and relationships between cells. The documentation has helped them better understand the terminology used in the configuration table and has been used frequently. The relationships between the cells enabled them to fill out the table better and understand how the project's different parts are related.

As for difficulties, they found that the starting point in the configuration table needed to be clarified. In the course, they were presented with one column at a time, starting with the Situation column. For their project, however, they found that starting with another column might have been more accessible and more intuitive.

The group did not establish relationships between their statements, as they found it hard to identify when to do so and the benefits of it. They found that the relationships between the cells and relationships between the statements essentially were the same. After a discussion with the moderator, the group concluded that having relationships between 'statements' could help identify dangling information in their project and thus avoid working on specific aspects of the project, which provided no value.

C | Focus Group Red

C.1 Part 1: Introduction

The group has some experience with other tools and techniques, such as scrum, wireframes, interviews, and standup meetings but has yet to use them in a structured manner.

C.2 Part 2: Assignment

The group is evaluating the statements in their Digital Essence prospect. They are considering the cells based on the number of dangling information in each. During the assignment, they reflect upon the validity of some of their statements. They are evaluating the relationships they have established and if the classification of their 'statements' is correct. In the assignment, they encounter a few misalignments, which leads to discussions about the group's different points of view. The discussions lead to an agreement about certain delimitations in the project.

C.3 Part 3: Follow-up questions

They found that Digital Essence helped them maintain an overview of their project and served as a medium to communicate and explore different ideas. The configuration table challenges their proposed solution, as they are forced to evaluate various aspects based on the cells in the table. They found that Digital Essence was most beneficial for them in their project's early and late stages to aid in the reflection and validation process. However, they pointed out that Digital Essence might be even more helpful next semester, as they would have learned all the theory behind Essence by then. One of the key benefits of Digital Essence is that it eliminates assumptions by the way information is being divided between cells, which leads to tacit knowledge being openly discussed within the group.

The group has actively been using all of the guiding features except the change-advi. The guiding documentation helped them clarify and understand the Essence terminology, and the relationships between cells helped them navigate through the configuration table. They found that relationships between statements helped them reflect upon how the different pieces of information are related to their project.

As for difficulties, they found that the entry point in the configuration table took time to pinpoint. They requested a "getting started" guide, which proposed different entry points in the configuration table based on different types of projects. They needed help determining the appropriate time to create relationships between their statements. They discuss whether they should create relationships continuously or wait until multiple 'statements' have been made.

D | PCA - Jupyer Notebook

In [1]:
```
# Setup
from mpl_toolkits.mplot3d import Axes3D
from matplotlib.patches import FancyArrowPatch
from mpl_toolkits.mplot3d import proj3d
import numpy as np
class Arrow3D(FancyArrowPatch):
    def __init__(self, xs, ys, zs, *args, **kwargs):
        super().__init__((0,0), (0,0), *args, **kwargs)
        self._verts3d = xs, ys, zs

    def do_3d_projection(self, renderer=None):
        xs3d, ys3d, zs3d = self._verts3d
        xs, ys, zs = proj3d.proj_transform(xs3d, ys3d, zs3d, self.axes.M)
        self.set_positions((xs[0],ys[0]),(xs[1],ys[1]))

        return np.min(zs)
arrow_prop_dict =dict(mutation_scale=20,
                      arrowstyle='-|>',
                      color='k',
                      shrinkA=0,
                      shrinkB=0)
arrow_prop_dict2 =dict(mutation_scale=10,
                      arrowstyle='<-',
                      color='k',
                      shrinkA=0,
                      shrinkB=0)
```

Loading data

Loading the data from the `data.json` in the `data` folder and show the structure of the workspace dataset.

In [2]:
```
import json
import pandas as pd
import numpy as np
with open('../data/data.json') as json_file:
    workspace_data = json.load(json_file)

workspace_df = pd.DataFrame(workspace_data)

features = ["insertedStatements","updatedStatements",
            "deletedStatements","createdRelations",
            "deletedRelations","wikiLookups",
            "standardStatements","futureStatement",
            "uncertainStatements","sessions",
            "patternsFollowed"]

workspace_df.head()
```

Out[2]:

	workspaceId	insertedStatements	updatedStatements	deletedStatem
0	cle5jbyax000bl508dxu7umfw	45	9	
1	cle5k109a000eju08qp5gvnro	59	69	
2	clefcx6ku0007id08ivfhcbyg	40	20	
3	clefdpwf6001hid08xplr3vx8	33	10	
4	clefi348f001xlb08irzw5ydv	24	9	

Normalizing the features in the dataset.

This is done with a standard scaler which is done by removing the mean and scaling to unit variance.

In [3]:
```
from sklearn.preprocessing import StandardScaler
x = workspace_df.loc[:, features].values
x = StandardScaler().fit_transform(x) # normalizing the features
```

The normalized dataset look as follows:

In [4]:
```
feat_cols = ['feature'+str(i) for i in range(x.shape[1])]
normalised_ws = pd.DataFrame(x,columns=feat_cols)
normalised_ws.tail()
```

Out[4]:

	feature0	feature1	feature2	feature3	feature4	feature5	feature6	f
16	-0.211938	-0.804182	-0.876746	-0.548754	-0.353553	-0.531972	-0.724616	-0.
17	1.441181	0.272332	1.552571	-0.548754	-0.353553	0.628136	-0.724616	-0.
18	-1.801476	-0.804182	-0.876746	-0.548754	-0.353553	-1.133510	-0.724616	-0.
19	0.678203	-0.400489	0.018266	-0.548754	-0.353553	-0.746807	-0.724616	-0.
20	-0.211938	-0.131360	-0.493170	2.756196	-0.353553	-0.789774	2.339906	1.

Finding the number of principal components

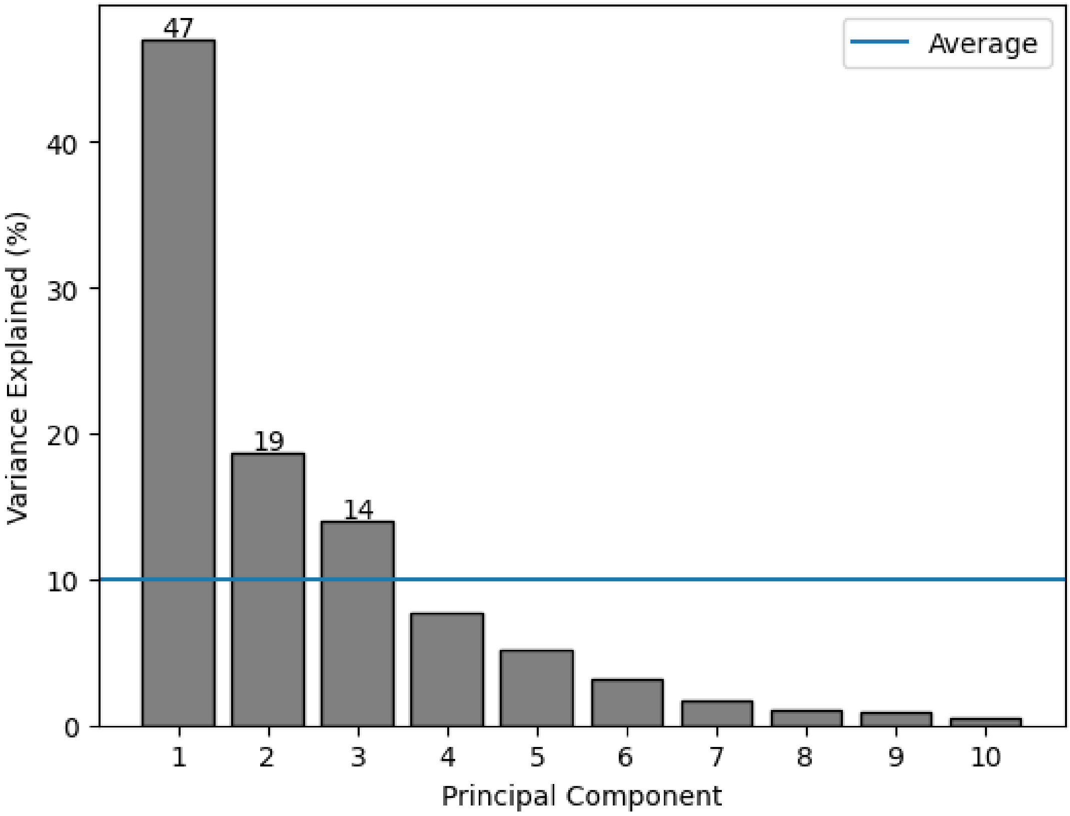
To find the number of principal components we are going to use in the analysis, we can use what is called a 'scree' plot.

This shows the Variance Explained (%) (VE-%) for each principal component. Here we can see what PC attributes to the most Variance Exlpained.

We want to pick all PCs that have higher than average VE-%.

In [5]:
```
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt
```

```
pca = PCA(n_components=10)
pp = pca.fit_transform(x)
x1 = [i for i in range(1,11)]
y1 = pca.explained_variance_ratio_

plt.xticks(x1)
plt.xlabel('Principal Component')
plt.ylabel('Variance Explained (%)')
plt.bar(x1, y1*100, color='grey', edgecolor='#000000')
for i in range(len(x1) - 7):
    plt.text(i + 1, y1[i]*100 + 0.2, round(y1[i]*100), ha = 'center')
plt.axhline(y=np.nanmean(y1*100))
plt.legend(['Average'], loc='upper right')
plt.savefig('../output/workspaces_pca_scree.pdf', bbox_inches='tight')
plt.savefig('../output/workspaces_pca_scree.png', bbox_inches='tight')
```



Choosing to work with 3 PCs

The plot shows that there is a big cutoff after PC1, but PC2 and PC3 are still above average. We looked at using 1, 2 or 3 PCs and found that including PC1, PC2, and PC3 gave of the best results.

The next step is to decompose the dataset into 3 PCs.

In [6]:
```
from sklearn.decomposition import PCA
pca_ws = PCA(n_components=3)
principalComponents_ws = pca_ws.fit_transform(x)
```

```
In [7]: loadings = pd.DataFrame(pca_ws.components_.T,
                                columns=['PC1', 'PC2', 'PC3'],
                                index=["D1","D2","D3","D4","D5",
                                       "D6","D7","D8","D9","D10","D11"])
        loadings = loadings.reindex(["D2", "D10", "D9", "D3", "D4",
                                     "D8", "D7", "D6", "D5", "D1", "D11"])
        loadings.round(3)
```

Out[7]:

	PC1	PC2	PC3
D2	0.389	-0.159	-0.086
D10	0.374	-0.085	0.133
D9	0.358	-0.095	-0.376
D3	0.356	-0.263	0.003
D4	0.225	0.508	-0.239
D8	0.225	0.424	0.315
D7	0.178	0.584	-0.161
D6	0.261	0.168	0.543
D5	0.293	-0.084	-0.492
D1	0.291	-0.193	0.306
D11	0.286	-0.195	0.147

```
In [8]: print("Covariance Matrix:")
        cov_df = pd.DataFrame(pca_ws.get_covariance()).head()
        cov_df.columns = [i for i in range(1,12)]
        cov_df.head()
```

```
Covariance Matrix:
```

Out[8]:

	1	2	3	4	5	6	7	8
0	0.921635	0.605445	0.628662	0.057502	0.269602	0.550082	-0.008021	0.31159
1	0.605445	1.128596	0.790048	0.326403	0.666139	0.411373	0.201704	0.28846
2	0.628662	0.790048	1.072820	0.161929	0.574630	0.398025	0.039838	0.20518
3	0.057502	0.326403	0.161929	1.111458	0.414586	0.290155	0.811596	0.56370
4	0.269602	0.666139	0.574630	0.414586	1.065422	0.014134	0.281863	0.06751

The dataset is now reduces to three PCs:

```
In [9]: principal_ws_Df = pd.DataFrame(data = principalComponents_ws,
                                       columns = ['principal component 1',
                                                  'principal component 2',
                                                  'principal component 3'])
        principal_ws_Df.tail()
```

Out[9]:

	principal component 1	principal component 2	principal component 3
16	-1.415153	-0.549247	0.085997
17	1.553532	-1.906086	1.531939
18	-2.623173	-0.038284	-0.994938
19	-0.893876	-1.043761	0.162168
20	0.371272	3.286181	-0.994136

In [10]:

```
print('Explained variation per principal component: {}'
      .format(pca_ws.explained_variance_ratio_))
print(sum(pca_ws.explained_variance_ratio_))
principal_ws_Df['workspace'] = workspace_df['workspaceId']
principal_ws_Df.head()
```

```
Explained variation per principal component: [0.47044367 0.18702993 0.1398067
2]
0.7972803147804117
```

Out[10]:

	principal component 1	principal component 2	principal component 3	workspace
0	1.164426	3.805754	0.170221	cle5jbyax000bl508dxu7umfw
1	8.276532	-0.878307	-2.587956	cle5k109a000eju08qp5gvnro
2	2.436290	2.205530	3.508760	clefcx6ku0007id08ivfhcbyg
3	-0.538910	0.008270	-0.094116	clefdpwf6001hid08xplr3vx8
4	-1.639528	-0.066740	0.136391	clefi348f001xlb08irzw5ydv

Plotting the results

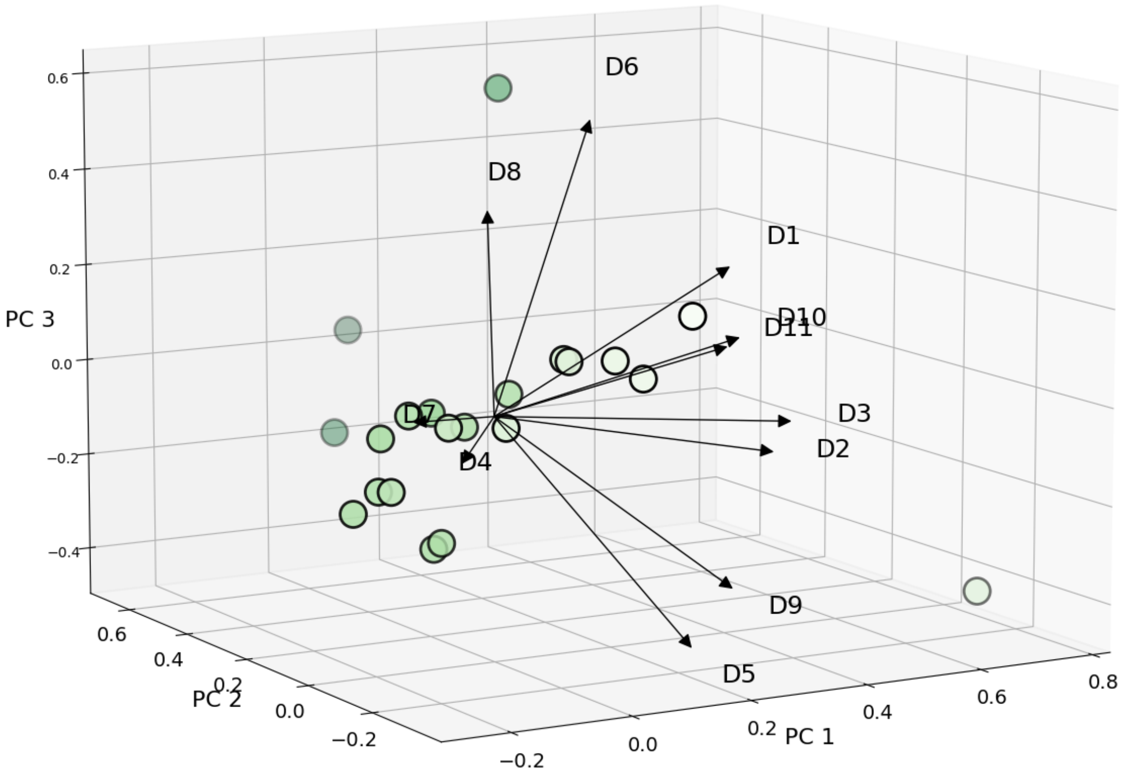
We can now look at the Distance Biplot to see how the different dimensions affect the placement of teams in the 3d-space.

In [11]:

```
from matplotlib import pyplot as plt

loadings = pca_ws.components_.T * np.sqrt(pca_ws.explained_variance_)
loadings_matrix = pd.DataFrame(loadings,
                               columns=['principal component 1',
                                        'principal component 2',
                                        'principal component 3'],
                                        index=features)
ldngs = pca_ws.components_
feats = ["D1", "D2", "D3", "D4", "D5",
         "D6", "D7", "D8", "D9", 'D10', 'D11']

PC1 = pp[:, 0]
PC2 = pp[:, 1]
PC3 = pp[:, 2]
```

```
scalePC1 = 1.0/(PC1.max() - PC1.min())
scalePC2 = 1.0/(PC2.max() - PC2.min())
scalePC3 = 1.0/(PC3.max() - PC3.min())

def get_coordinates(workspaceId: str):
    return (
        principal_ws_Df.loc[principal_ws_Df['workspace'] == workspaceId]['pr
        principal_ws_Df.loc[principal_ws_Df['workspace'] == workspaceId]['pr
        principal_ws_Df.loc[principal_ws_Df['workspace'] == workspaceId]['pr

fig = plt.figure(figsize=(15,15))
ax = fig.add_subplot(projection='3d')

# Data for three-dimensional scattered points
zdata = principal_ws_Df['principal component 3']
xdata = principal_ws_Df['principal component 1']
ydata = principal_ws_Df['principal component 2']
ax.set_xlabel('PC 1', fontsize=16)
ax.set_ylabel('PC 2', fontsize=16)
ax.set_zlabel('PC 3', fontsize=16)
ax.scatter3D(xdata * scalePC1,
             ydata * scalePC2,
             zdata * scalePC3,
             c=ydata,
             cmap='Greens',
             s=350,
             linewidths=2,
             edgecolors='black')

for i, feats in enumerate(feats):
    a = Arrow3D([0, ldngs[0, i]],
                [0, ldngs[1, i]],
                [0, ldngs[2, i]],
                **arrow_prop_dict)
    ax.add_artist(a)
    ax.text(ldngs[0, i] * 1.15,
            ldngs[1, i] * 1.15,
            ldngs[2, i] * 1.15,
            feats, fontsize=18)

ax.azim = 240
ax.elev = 10
plt.yticks(fontsize=14)
plt.xticks(fontsize=14)

plt.savefig('../output/workspaces_pca_3d_biplot.png')
plt.savefig('../output/workspaces_pca_3d_biplot.pdf')
import os
os.system('pdfcrop %s %s &> /dev/null &'%('../output/workspaces_pca_3d_biplo
                                          '../output/workspaces_pca_3d_biplo
```

Out[11]: 0

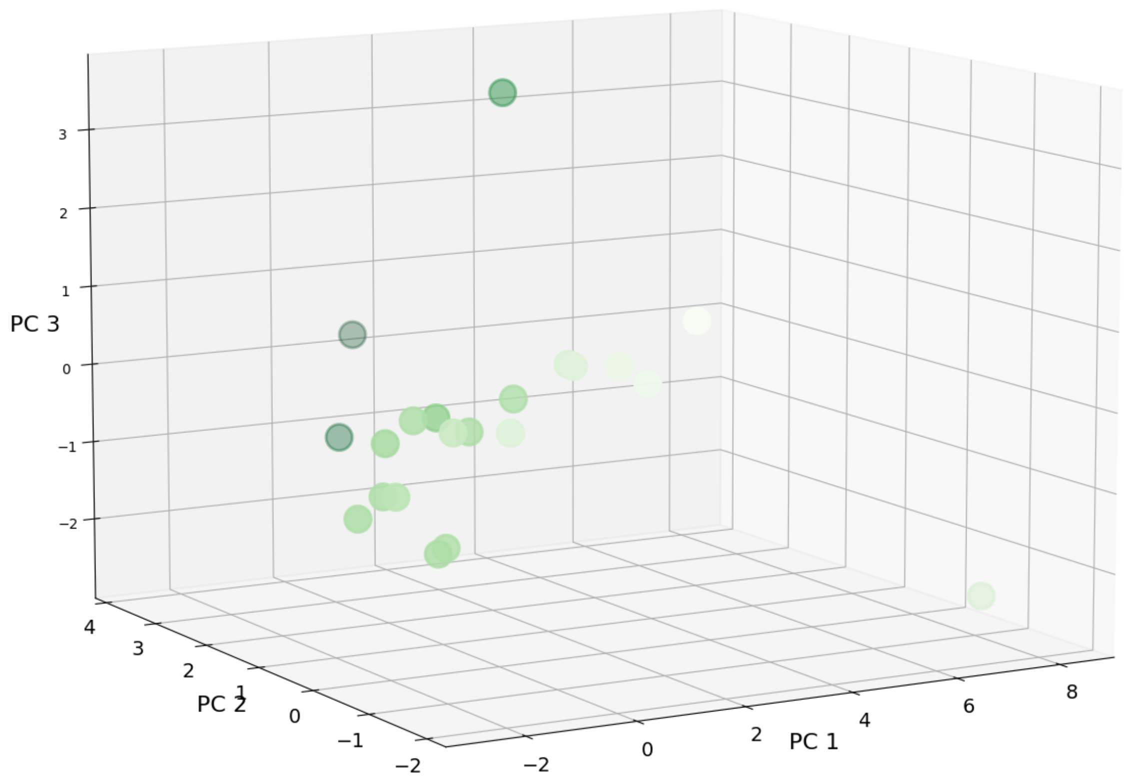


A plot where we can see teams. Ligther green means 'lower' on the PC2-axis.

In [12]:
```
import matplotlib.pyplot as plt

def get_coordinates(workspaceId: str):
    return (principal_ws_Df.loc[principal_ws_Df['workspace'] == workspaceId]
            principal_ws_Df.loc[principal_ws_Df['workspace'] == workspaceId]
            principal_ws_Df.loc[principal_ws_Df['workspace'] == workspaceId]

fig = plt.figure(figsize=(15,15))
ax = fig.add_subplot(projection='3d')

# Data for three-dimensional scattered points
zdata = principal_ws_Df['principal component 3']
xdata = principal_ws_Df['principal component 1']
ydata = principal_ws_Df['principal component 2']
ax.set_xlabel('PC 1', fontsize=16)
ax.set_ylabel('PC 2', fontsize=16)
ax.set_zlabel('PC 3', fontsize=16)
```

```
ax.scatter3D(xdata, ydata, zdata, c=ydata, cmap='Greens', s=350, linewidths=
ax.azim = 240
ax.elev = 10
plt.yticks(fontsize=14)
plt.xticks(fontsize=14)

plt.savefig('../output/workspaces_pca_3d_depth.png')
plt.savefig('../output/workspaces_pca_3d_depth.pdf')
import os
os.system('pdfcrop %s %s &> /dev/null &'%('../output/workspaces_pca_3d_depth
```

Out[12]: 0



Clustering into segments

Finally we can run a clustering algorithm to see if the groups can be clustered into different segments.

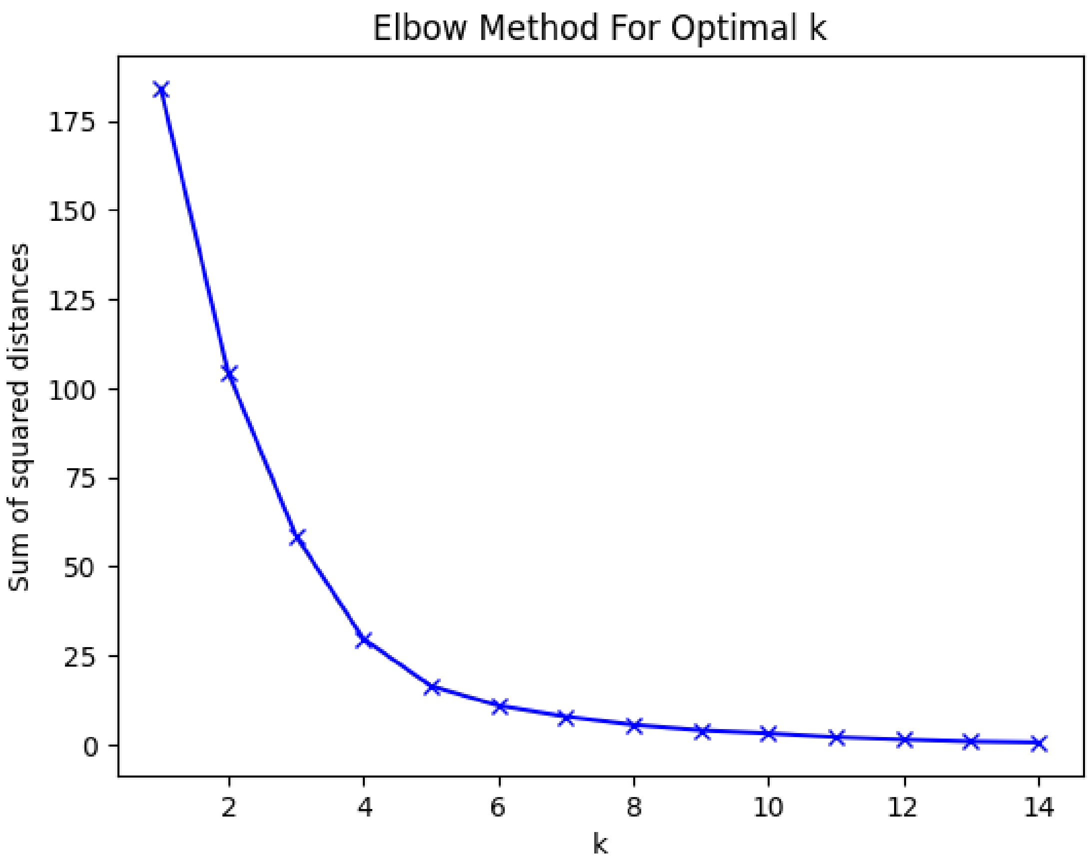
Here we are using kmeans as is works well together with pca-reduction.

In [13]:
```
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score, silhouette_samples
X = np.array(list(zip(principal_ws_Df['principal component 1'], principal_ws
# determining the maximum number of clusters
# using the simple method
limit = int((X.shape[0]//2)**0.5)

# determining number of clusters
# using silhouette score method
for k in range(2, limit+1):
    model = KMeans(n_clusters=k, n_init=10)
    model.fit(X)
    pred = model.predict(X)
    score = silhouette_score(X, pred)
    print('Silhouette Score for k = {}: {:<.3f}'.format(k, score))
```

```
Silhouette Score for k = 2: 0.655
Silhouette Score for k = 3: 0.510
```

In [14]:
```
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
X = np.array(list(zip(principal_ws_Df['principal component 1'], principal_ws
sum_of_squared_distances = []
K = range(1,15)
for k in K:
    km = KMeans(n_clusters=k, n_init=10)
    km = km.fit(X)
    sum_of_squared_distances.append(km.inertia_)

plt.plot(K, sum_of_squared_distances, 'bx-')
plt.xlabel('k')
plt.ylabel('Sum of squared distances')
plt.title('Elbow Method For Optimal k')
plt.savefig('../output/workspaces_pca_3d_elbow.png')
plt.savefig('../output/workspaces_pca_3d_elbow.pdf')
import os
os.system('pdfcrop %s %s &> /dev/null &'%('../output/workspaces_pca_3d_elbow
```

Out[14]:
```
0
```

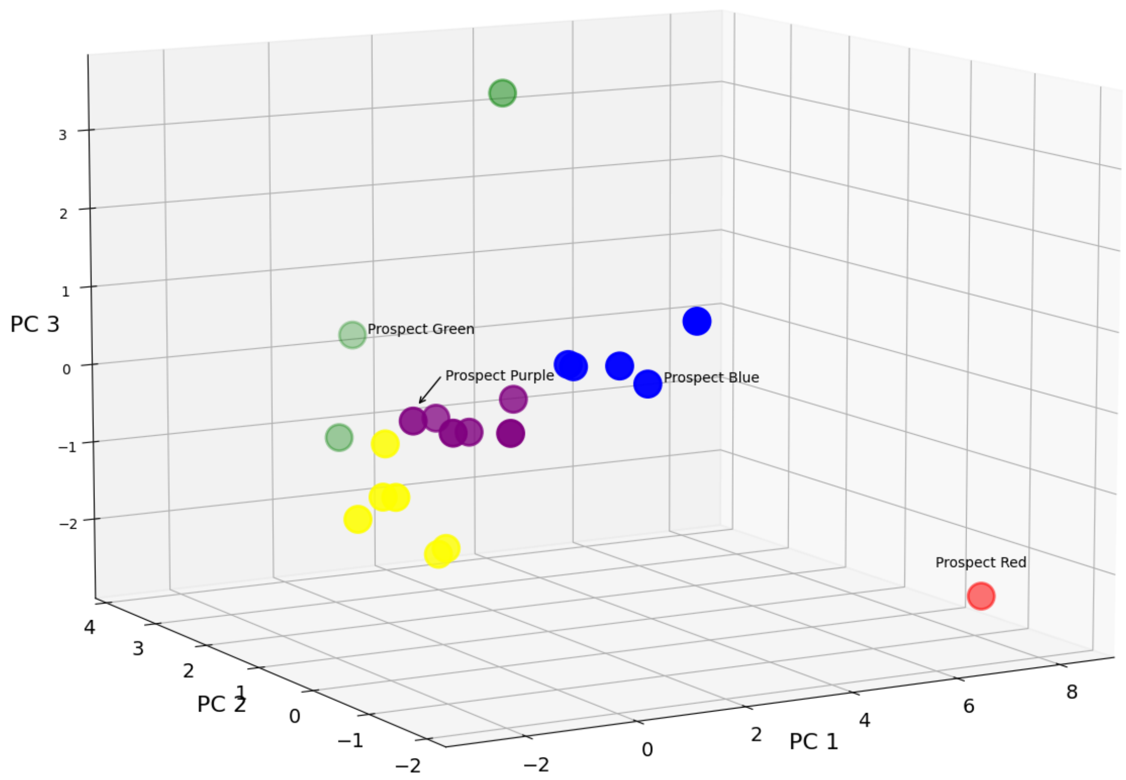


In [15]:

```
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

X = np.array(list(zip(principal_ws_Df['principal component 1'], principal_ws
kmeans = KMeans(n_clusters=5, random_state=0, n_init="auto").fit(X)

def get_coordinates(workspaceId: str):
    return (principal_ws_Df.loc[principal_ws_Df['workspace'] == workspaceId]
            principal_ws_Df.loc[principal_ws_Df['workspace'] == workspaceId]
            principal_ws_Df.loc[principal_ws_Df['workspace'] == workspaceId]

fig = plt.figure(figsize=(15,15))
ax = fig.add_subplot(projection='3d')
def set_color(x):
    if x == 0:
        return "purple"
    elif x == 1:
        return "blue"
    elif x == 2:
        return "red"
    elif x == 3:
        return "green"
    else:
        return "yellow"

colors = [set_color(x) for x in kmeans.labels_]

# Data for three-dimensional scattered points
```

```
zdata = principal_ws_Df['principal component 3']
xdata = principal_ws_Df['principal component 1']
ydata = principal_ws_Df['principal component 2']
ax.set_xlabel('PC 1', fontsize=16)
ax.set_ylabel('PC 2', fontsize=16)
ax.set_zlabel('PC 3', fontsize=16)
ax.scatter3D(xdata, ydata, zdata, c=colors, s=350, linewidths=2)
ax.azim = 240
ax.elev = 10
plt.yticks(fontsize=14)
plt.xticks(fontsize=14)
A = get_coordinates("clepe39ze0004l2083efqmees") # (SWI Exercise)
B = get_coordinates("cle5k109a000eju08qp5gvnro") # Lilleper (Dynamic nationa
C = get_coordinates("cle5jbyax000bl508dxu7umfw") # ML Research (Titled)
D = get_coordinates("clefi348f001xlb08irzw5ydv") # UX designer community dev
ax.text(A[0] + 0.3, A[1], A[2], "Prospect Blue")
ax.text(B[0] - 0.8, B[1] + 0.1, B[2] + 0.4, "Prospect Red")
ax.text(C[0] + 0.3, C[1], C[2], "Prospect Green")
#ax.text(D[0], D[1], D[2] + 0.2, "Prospect D")
a = Arrow3D([D[0] + 0.1, D[0] + 0.5],
            [D[1], D[1]],
            [D[2] + 0.2, D[2] + 0.5],
        **arrow_prop_dict2)
ax.add_artist(a)
ax.text(D[0] + 0.5 * 1.15,
        D[1] * 1.15,
        D[2] + 0.4 * 1.15,
        "Prospect Purple")

plt.savefig('../output/workspaces_pca_3d_clusters.png')
plt.savefig('../output/workspaces_pca_3d_clusters.pdf')
import os
os.system('pdfcrop %s %s &> /dev/null &'%('../output/workspaces_pca_3d_clust
```

Out[15]: 0



PC 3
3
2
1
0
−1
−2
Prospect Green
Prospect Purple
Prospect Blue
Prospect Red
PC 2
4
3
2
1
0
−1
−2
PC 1
−2
0
2
4
6
8

E | Clustering of prospects

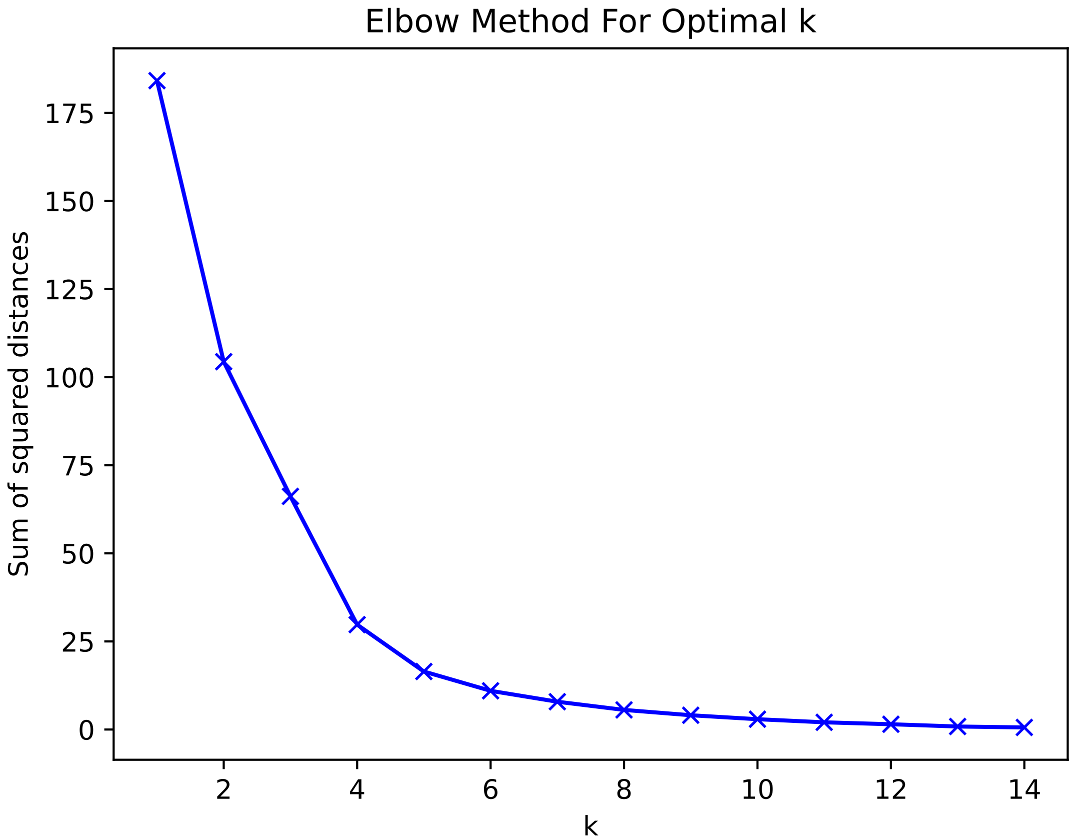


Figure E.1: elbow method plot