POLITECNICO DI MILANO

Master in

Building Information Modelling

BIM A+ — European Master in Building Information Modelling

Method for Big Data Management in EPC Firms During Engineering Phase

Supervisors:

Prof. Alberto Pavan
Mr. Davide Gatti
Mr. Enrico Demaria

Author:

Joe Said

a.a. 2022/2023

AUTHORSHIP RIGHTS AND CONDITIONS OF USE OF THE WORK BY THIRD PARTIES

ACKNOWLEDGEMENTS

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of ethics and conduct of Politecnico di Milano.

SOMMARIO

L'industria della Modellazione delle Informazioni per la Costruzione (Building Information Modelling, BIM) è da tempo concentrata sullo sviluppo di software, standard e interoperabilità, il che ha permesso un notevole progresso negli ultimi cinque anni. Questa focalizzazione spesso trascura il volume e la varietà dei dati generati sia dai singoli modelli BIM che dai progetti federati. Questa ricerca è stata commissionata per affrontare il problema dei big data disponibili in un'azienda di Ingegneria-Approvvigionamento-Costruzione (Engineering-Procurement-Construction, EPC) con sede in Italia. Con l'obiettivo di sostituire i processi manuali di controllo di qualità con un approccio automatizzato, questa ricerca è riuscita a sviluppare un metodo strutturato per estrarre, formattare e analizzare i dati utilizzando pratiche di gestione dei big data, note come Big Data Management (BDM), applicate ai dati BIM. Questa metodologia è stata ottenuta dopo un processo iterativo di creazione di diversi pannelli di controllo, o dashboard, con Power BI e segue un processo in otto fasi: esplorazione, mappatura, estrazione, formattazione, visualizzazione, automazione, test e iterazione. Le dashboard create per questa ricerca includono quattro controlli qualitativi e un riepilogo quantitativo generale delle prestazioni del progetto. La prima dashboard qualitativa verifica l'allineamento della posizione del modello nella struttura delle cartelle dell'Ambiente Dati Comune (Common Data Environment, CDE) e del modello federato, utilizzando la piattaforma Autodesk Construction Cloud (ACC) e il modello federato Navisworks. La seconda dashboard qualitativa è stata creata per verificare gli elementi singoli del modello per l'allineamento con gli standard aziendali e le convenzioni di marcatura. La terza crea un collegamento tra l'Ambiente Dati Comune online dell'azienda e la documentazione di costruzione per assicurare che gli elementi del modello siano caricati in entrambi e le quantità siano allineate. L'ultima dashboard qualitativa utilizza ancora il connettore dati ACC / BIM 360 per Power BI ed effettuata un controllo completamente automatizzato per le revisioni di Squad Check, evidenziando le prestazioni del reparto e le revisioni mancanti. L'unica dashboard quantitativa ha raccolto tutti i dati dai progetti in un formato riepilogativo per aiutare i manager di livello medio a prendere decisioni informate sullo stato di avanzamento del progetto e disporre di un quadro storico per guidare i futuri sforzi dell'azienda. I dashboards risultanti sono ora in uso e hanno permesso di risparmiare molte ore di lavoro ai coordinatori BIM dell'azienda grazie al grado di automazione che offrono. Sebbene questa ricerca abbia adottato un approccio non convenzionale rispetto alla tradizionale ricerca accademica, ponendo un'enfasi sui risultati tangibili, è stata un'esperienza estremamente positiva. Ciò è dovuto al fatto che ha permesso all'autore di creare una metodologia collaudata per la gestione dei big data BIM basata sulle iterazioni nella creazione delle dashboard con Power BI. Le future ricerche dovrebbero concentrarsi sulla capacità di automatizzare l'estrazione dei dati BIM in modo che possano essere correttamente importati in un software di analisi senza la necessità di formattare manualmente nessuno dei dati.

Parole chiave: ACC, Big Data Management nel BIM, EPC, Power BI.

ABSTRACT

The Building Information Modelling (BIM) industry has long been focused on the development of software, standards, and interoperability that the industry has moved forward greatly in the last five years. This focus often neglects the volume and variety of the data created from both individual BIM models and federated projects. This research was commissioned to deal with the big data generated in an Engineering-Procurement-Construction (EPC) firm based out of Italy. With a focus on creating Power BI Dashboards to replace manual quality checks with automation of data checks, this research was able to create a methodical approach to extracting, formatting and analyzing data by utilizing Big Data Management (BDM) practices with BIM data. This methodology was obtained after an iterative process of creating several Power BI Dashboards and follows an eight-step process: exploration, mapping, extraction, formatting, visualization, automation, testing, and reiteration. The dashboards created for this research include four qualitative checks and an overall quantitative summary of project performance. The first qualitative dashboard checks the model location alignment in the Common Data Environment (CDE) folder structure and the federated model; utilizing the Autodesk Construction Cloud (ACC) platform and the Navisworks federated model. The second qualitative dashboard was created to check singular model elements for alignment with company standards and tagging conventions. The third qualitative dashboard creates a bridge between the companies' web based CDE and the construction documentation to ensure model items are loaded in both and the quantities are aligned. The final qualitative dashboard also utilized the ACC Data Connector to Power BI and had a fully automated check for Squad Check Reviews, highlighting department performance and missing reviews. The only quantitative dashboard garnered all data from projects into a summary format to help mid-level managers make informed decisions on project progress and have a historical context to help guide future endeavors for the company. The resulting dashboards are now in use and have saved many work hours for the BIM coordinators of the company due to the degree of automation they provide. Although this research worked contrary to typical academic research due to the focus on deliverables, it created a very beneficial workflow. This was because it allowed the author to create a tried and tested methodology for BIM Big Data Management from the iterations of Power BI Dashboard creation. Future research should focus on the ability to automate the extraction of BIM data to be correctly imported into an analytics software without the need to manually format some of the data.

Keywords: ACC, Big Data Management in BIM, EPC, Power BI.

Erasmus Mundus Joint Master Degree Programme – ERASMUS+

vi   European Master in Building Information Modelling BIM A+

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

This page is intentionally left blank

1. INTRODUCTION

Building Information Modeling (BIM), also known as Virtual Design in Construction (VDC) has been steadily improving in the last five years. The improvement is defined by an increase in software availability and interoperability, a larger number of elements in a model that need to transfer reliable data to the construction site, and an more project users and accessibility to all parties of a project. This increase of data availability and model enrichment has created a vast opening in the sector for data management. With more data available, most Architecture, Construction, and Engineering (AEC) firms are still not equipped with the knowledge for Big Data Management (BDM) practices and end up having some unreliable data that is not checked for quality. It is near impossible to check the data manually now due to its sheer size so there is a research need to create good practices and methodologies to manage the data quality.

Practices of big data management have been utilized throughout the information technology sector for many years now, the purpose of this research is to gap the bridge between the two industries and create a methodology for AEC firms to create quality checks using any data analytics and visualization platform they choose. This research was initiated by an Engineering-Procurement-Construction (EPC) firm looking to harmonize the big data checks and data quality for the company. As part of the research, the author created five Power BI dashboards utilizing the data available from more than 15 projects for the company. The five dashboards' requirements can be summarized in Table 1:

Table 1 – Dashbaord Requirements for Research

Dashboard	Requirement
Quality 1 – Model Alignment Check	Used to ensure models are stored in the correct location in the Common Data Environment (CDE) and the federated model.
Quality 2 – Models Attribute Check	Required to match company standards for elements' properties. Used to show modelers errors and how to fix them.
Quality 3 – Tags Data Alignment Check	Makes sure items from the models are loaded in the company database and that quantities match.
Quality 4 – Squad Check Reviews	Highlights missing reviews and helps identify departments performance and speed in Squad Checks.
Quantity 1 – Civil Quantities	Give historical background and ongoing progress of material use and model development.

Given the nature of this research, the work on the dashboard was immediate and preceded many of the literature review studies. Although this seems counterintuitive, it was very beneficial to start research with a blank slate of methodologies. During the research there were many variations of methodologies used to create the dashboards with mistakes, rework, and deleted work; but due to this iterative process, the author was able to create a methodology that fits the AEC industry while still utilizing the influence of the Data Analytics industry. After reviewing other research within the same sphere, there was an apparent gap between the data analytics research and the BIM research. The data analytics research is

too advanced for BIM technologies and assumes fewer anomalies than the BIM data availability. But the data analytics sector can ignore these anomalies since the data is so much larger than typical BIM data which cannot ignore anomalies. On the other hand, the BIM research does not fully utilize all the capabilities of data analytics in working with the data. The real gap was before the data was formatted, where the BIM data needs to be adjusted for the big data analytics procedures. This will be further discussed in the methodology of this research.

This paper follows a standard procedure of reading, but it is worth noting that the actual work was done in the opposite order than the reading of this paper. After the introduction in Chapter 1, Chapter 2 looks at past literature in the subject matter. The chapter starts with defining the fundamentals of Big Data Management and what previous research focuses on to avoid mistakes. Then the author investigates a few research projects on using BDM with BIM in different industries, this helps guide the gap in the research. The discussion then delves into the BIM standards and abilities to export properties. Finally, the chapter ends with a focus on visualization for user interactions based on graphic design principles.

Chapter 3 is the methodology conceived by the author during the research period. The methodology follows an eight-step process of: data exploration, mapping, data extraction, data formatting, visualization, automation, testing, and reiteration. Exploration was introduced by the author to help organize the mindset of future creators of data visualizations. The purpose of exploration is to understand the data on a deeper level by understanding the type of data extracted, the different sources the data comes from and any connections they might have to each other and to the required output. Mapping focuses on the required results and where the data for said results originates from. Mapping helps in identifying the connections and required steps to create the required checks. The next step of extraction is required for BIM implementation so that the data can be continuously extracted or extracted to the correct location for analysis. Formatting the data is an important phase that helps achieve the results discussed in the mapping step by manipulating the data using three techniques: cleaning, creating or linking the data. Visualization is integral to successful data analysis because it is what the user interacts with and can be the cause for a useless dashboard if not fully thought out. Visualization can be broken down into creating a layout, selecting the correct charts, building the filters required, and choosing the appropriate color scheme. The next step of automation is a working progress in the BIM sector since some data is not directly transferable to data analysis tools. Automation can be either manual, semi-automatic, or automatic based on the steps required to refresh a dashboard. Next, testing is required so that a select pool of users can give feedback and further understanding into what they expect from the dashboard, which leads to the final step of reiteration. Reiteration cannot be ignored since, more often than not, the dashboard needs to be updated after a round of testing to accommodate the users' requirements, and reiteration is used when the data gets larger and more complex to accommodate for any problems that may arise.

After the methodology is explored, Chapter 4 focuses on the case study of creating the five dashboards explained in Table 1 for the EPC firm. Each dashboard goes through the eight-step process outlined in Chapter 3. Chapter 4 starts with an introduction into MAIRE Tecnimont and the company's model creation process. Quality 1 focuses on the alignment of models in a project from the CDE folder structure to the Federated Model. Quality 2 focuses on the element attribute rules created by the company and checks that the properties for each element are correct. Quality 3 compares main items for projects from two data sources: from the BIM model and the companies' CDE manually uploaded by the project

leaders. Quality 4 utilizes cloud data connector to check progress of the Squad Checks. Quantity 1 uses historical insight into the quantities of main materials in projects based on the type of object being built. Chapter 4 also shows how the dashboards are visualized and used.

The final chapter, Chapter 5, concludes the research by re-visiting the work done and the achieved results of the project. Furthermore, the author recommends some future research into the topic and future applications that may benefit any further endeavors in BDM for BIM. Due to the time constraint and focus on development, the actual work on the project was chronologically the opposite of this paper; the case study work was done first then the methodology was extracted from lessons learned from creating the dashboards, then the work was reviewed and re-organized with the help of literature review.

This page is intentionally left blank

2. LITERATURE REVIEW

The research method of this project was not typical of academic research given the focus was on delivering useful dashboards to the client. Therefore, the literature review was done concurrently with the methodology of the project and focused on methods of solving problems and understanding Big Data Management (BDM) rather than delving into the literature before work starts. The current chapter will focus on defining some terminology that will help with the further reading of this research and some similar works that focused on BIM and Big Data.

2.1. Big Data Management

While working, the author wanted to understand the best methods for analyzing data and managing big data. This did not necessarily need to come from a BIM standpoint, and Gordov's "Analytical Review of Data Visualization Methods in Application to Big Data" helped guide the workflow in the project. The first plan of action was to understand what type of data the author was dealing with from the client. It was clear from reading the paper that the data included a large volume and high velocity, meaning many data points that get added constantly. The remainder of the paper focused on understanding which visualizations might benefit this type of data and some problems a person might face when dealing with big data and ways to avoid them. A few useful tips that helped guide this research included: avoiding visual noise when one array of data cannot be distinct from others, avoiding a large image perception because the human mind cannot capture large data clearly without guidelines and marks to help them gain insight, and avoiding information loss by limiting the amount of filtering a user can manipulate in a project. To avoid these problems, the authors suggested including more than one visual per page to be able to define the entire data set together. Also, creating a dynamic change in the factors affected in a visual allow some visuals to adapt to the filters used by the author to create a clearer picture of the data; whether it's digging deeper into a category or expanding several categories together, this can be achieved by creating dynamic filters and pages. (Gordov & Gubarev, 2013)

2.2. BIM and BDM

Understanding the difficulties and solutions a researcher might find when working with big data, the next step was to understand how Big Data Management (BDM) can be incorporated into BIM data. The last few years have introduced a varying amount of research focused on utilizing the BDM practices into BIM data management and understanding the data underlying BIM models. For example, research conducted by Ali and Bandi found a direct correlation between the key attributes of big data and the categories associated with BIM data. This research was able to associate attributes of BIM models with big data attributes and recommended the further investigation of utilizing BDM in the extraction and use of BIM models (Ali & Bandi, 2021). This may be true for BIM models, but similar research found it useless to use singular models for BDM. Research conducted by Correa found that BDM should not be used in for all BIM models and can be focused solely on the use of BIM in Smart Cities due to the large number of data associated with cities. The conclusion of the research showed that big data from City Information Modelling (CIM) may utilize the big data attributes and framework, but smaller projects should avoid this because it tends to lose the granularity of data points which are significant in smaller projects (Correa, 2015).

Though the two research topics are similar it was clear to the author that the research done in 2015 did not take into account the attribute heavy future of BIM or the granularity of building elements in projects. Another point that benefits the author was that the current research conducted focused on large EPC energy plants which included multiple models within each project and can be categorized as a small "city" built for the sustainability of the plant. With this in mind, it was clear that the BDM principles discussed earlier were useful and necessary to the management of the BIM data in this research.

This leads to research focused on the BIM Big Data framework research which assisted the author with the mental compartmentalization of how to organize the data. Research conducted by Farghaly et al. focused on the framework for asset management for BIM data. This research helped the author greatly and introduced a framework that helped validate the methodology for this research. The focus of the framework can be found in Figure 1. Looking closely, there are four levels of data management: first on collecting and extracting data, then processing and storing data, after that analyzing the data and finally incorporating and visualizing the data. It not only focused on the data available, but the ability to integrate knowledge from sources such as Geographical Information Systems (GIS), Internet of Things (IoT), asset management, energy management, and much more. This research focused on the asset management of building elements but incorporating other aspects to help control and visualize the data flow of BIM models can greatly benefit from the research conducted by Farghaly et al. (Farghaly, et al., 2017).
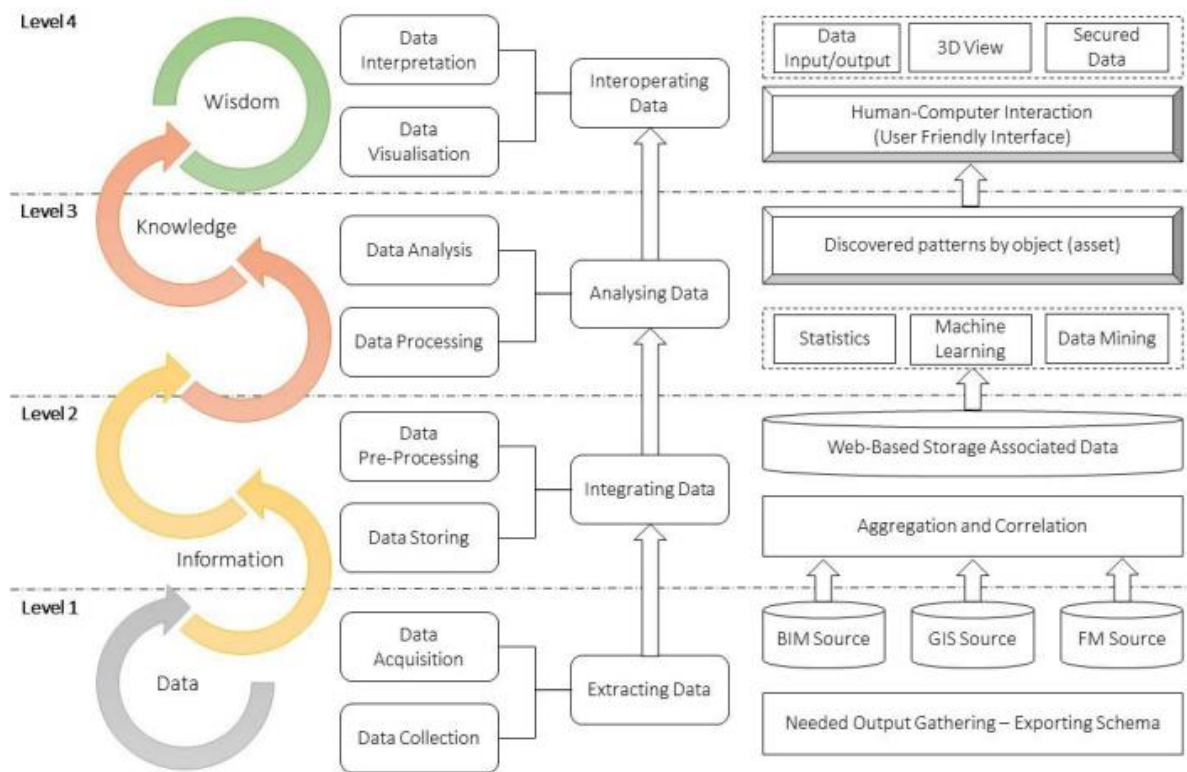

Figure 1 – BIM Big Data Framework for Asset Management (Farghaly, et al., 2017)

2.3. BIM Attributes and Classifications

With the framework of the research understood, it is now imperative to understand the attributes that are extracted from a BIM model and the universal standards and classifications that might be used with different BIM models. There are a multitude of standards and classification systems available, research comparing the many types are available and helped give insight to what to expect when collecting data from a BIM model. Research conducted in 2016 first compared the differences between the classification systems of that time. It is best summarized by Table 2: (Afsari & Eastman, 2016)

Table 2 – Comparison Between Four Classification Systems Based on the Established Criteria (Afsari & Eastman, 2016)

Classification systems	OmniClass	MasterFormat	UniFormat	Uniclass
Country of origin	North America	North America	North America	UK
Produced by	CSI and CSC	CSI and CSC	CSI and CSC	CPIc and NBS
Language	English	English	English	English
Purpose and properties	Organization, sorting and retrieval of product information for all objects in the built environment in the project life cycle.	A master list for organizing construction work results, requirements, products and activities. Mostly used in bidding and specifications.	For arranging construction information, organized around the physical parts of a facility known as functional elements mainly used for cost estimates.	For all aspects of the design and construction process. For organizing library materials and structuring product literature and project information
Framework	ISO 12006-2, ISO 12006-3, MasterFormat, UniFormat, EPIC	Industry practice and gradual development	ISO 12006-2 , Professional judgment	ISO 12006-2 , SfB, CAWS, EPIC, CESMM
Grouping principle	faceted	hierarchical	hierarchical	faceted
Organization and taxonomies	15 inter-related tables categorized by number and name. A combination of Table 21, Table 22 & Table 23 allows for classifying a product precisely.	One table with a series of six numbers and name: Level one with 50 divisions (2004 version) each is made up of level two, level three, and sometimes level four numbers and titles for more detailed areas of work results.	One table with alphanumeric designations and titles in five levels: level one is in nine categories separated by their special function. Level 2 separates them into constituent parts, level 3, 4 and 5 further subdivide them.	The division among facets is based on the alphabet in 11 tables and within each facet by decimal scale up to 6 digits. Table G, J, K and L can be used for classifying product models.

Knowing the standard classification system helps organize the author's extracting method depending on the classification system used for a BIM model. With the attributes added to the BIM models, the extraction method of getting the data usually follows the Industry Foundation Class (ifc) system introduced by buildingSMART. This standard is very useful in exporting the attributes specified by the classification standard chosen by the user. This ifc method of export contains layers of data schemes that allow for the export and unification of all BIM models from different software sources. Knowing this gave the author an understanding of how data can be exported and used for BDM and then further utilized in visualizations for this research (Liebich, et al., 2020).

2.4. Power BI Interface and Language

When creating a Power BI report, there is a process of importing and formatting data that requires the knowledge of the use of Power Query M formula language in Power Query and Data Analysis Expressions (DAX) in Power BI. It is best to explain the importance of each and the documentation available for people to understand and use the material available.

"Power Query is a data transformation and data preparation engine." Figure 2 shows the process from data to Microsoft software, as seen all data must pass through the Power Query so it is imperative to learn how to use the Query. Power Query's ability to merge and append data as well as format it helps in creating multiple functions for a large volume of data quickly. And to do this and format data, a user must learn the Power Query M Language, which is a functional, case sensitive language. (Microsoft, 2023)
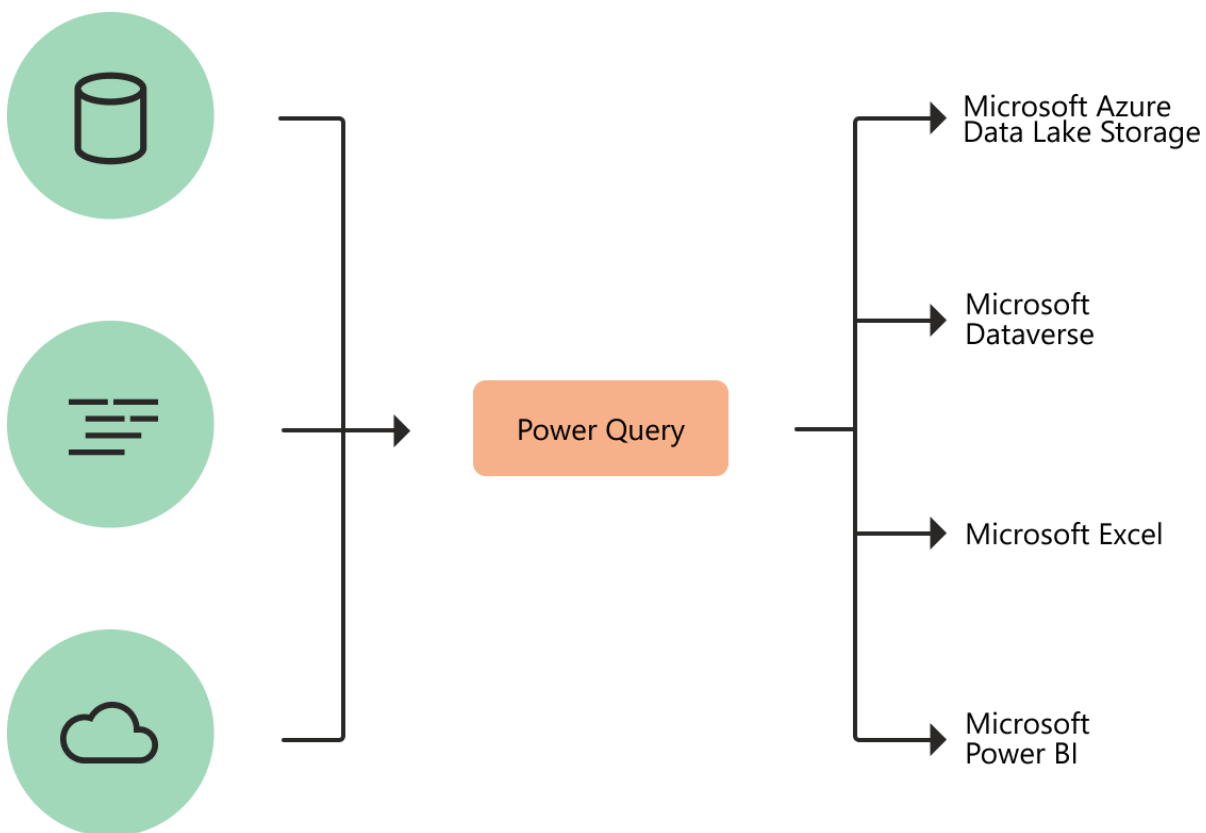
Figure 2 – Power Query Diagram (Microsoft, 2023)

Once the data is formatted and inside the Power BI platform, there is still possibility to format the data more using Power BI DAX language. A reader might first think, "why not just use Power Query for all formatting?" A good question, but when linking multiple data sources, it is better to create functions and operations that link the data in Power BI using DAX language. The ability to link data from different sources without changing the source data is very powerful and allows for comparing data from multiple sources. (Microsoft, 2023)

2.5.   Visualization

Once the author understood the process of gathering data and the attributes that might face a user when extracting data from BIM models. The final critical step was to understand how to visualize the data. As explained in Gordov & Gurabev's paper, there are many ways to exhibit big data depending on the variety and volume of your data. But the author found this to be too old to accommodate modern users (Gordov & Gubarev, 2013).

Instead, the author chose the Google Material Design webpage as a guidance to creating the dashboards. "Material Design is an adaptable system of guidelines, components, and tools that support the best practices of user interface design." Although this website focuses on web applications and apps User Interface (UI), it can be applied to dashboards as well. The Materials Design webpage is all inclusive and has categories such as: Foundations, Styles, and Components that the author utilized. (Google, 2023)

"Foundations inform the basis of any great user interface, from accessibility standards to essential patterns for layout and interaction." The author gained a lot of insight into the accessibility, adaptive design, and materials used in a layout to be able to transfer it over directly to the Power BI Dashboard. "Styles are the visual aspects of a UI that give it a distinct look and feel. They can be customized by changing your Material theme." Styles help in visualizing a correct color scheme and distances between elements. It also focuses on how to represent icons, create motion inside the interface, define shapes and scales, and select the adequate typography for the use-case. All this information helps design a dashboard that is specific to the user, and after this is set, the components added can be organized from Material Design. "Components are interactive building blocks for creating a user interface. They can be organized into categories based on their purpose: Action, containment, communication, navigation, selection, and text input." Some components were even used in this research's dashboard to help create elements that are designed with the user in mind. For example, **Figure 3** shows an example of the navigation bar that was then utilized for the remainder of this research and in the dashboards submitted to the client in the case study (Google, 2023).
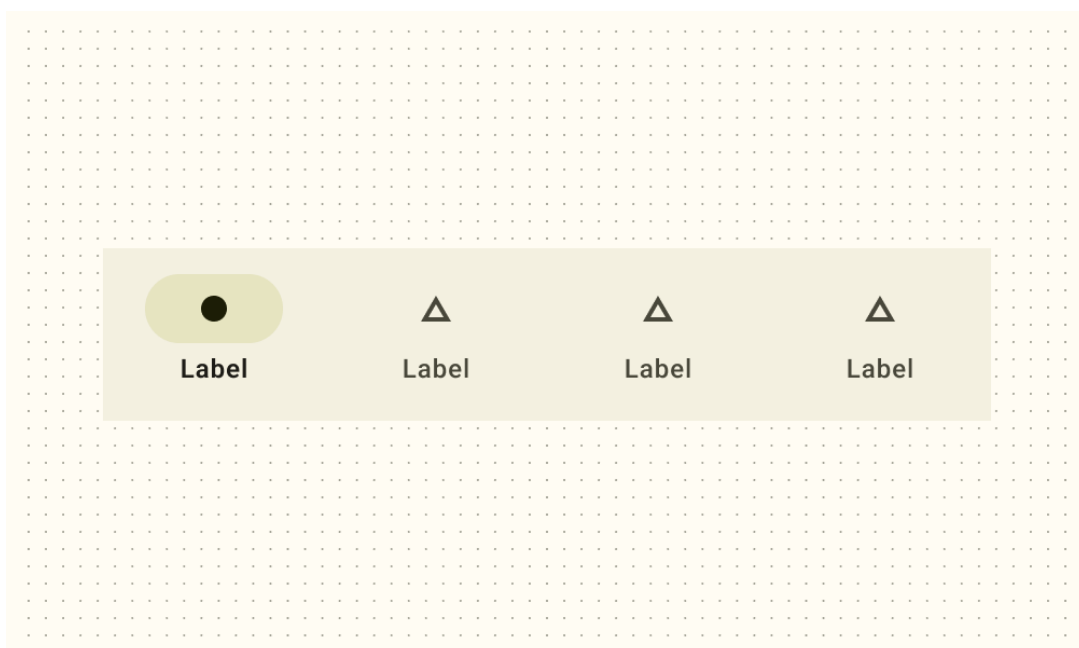
Figure 3 – Materials Design Navigation Bar (Google, 2023)

This page is intentionally left blank

3. METHODOLOGY

This chapter aims to explore the methodology a user can use to create any visualization of big data BIM elements or BIM libraries. The methodology follows the authors own attempts at creating multiple Power BI dashboards further discussed in Chapter 4 and is based on the multiple iterations done to create the dashboards. The resulting methodology was found to be the most efficient and successful method in creating the dashboard; it follows 8 basic steps explained in Figure 4 below:
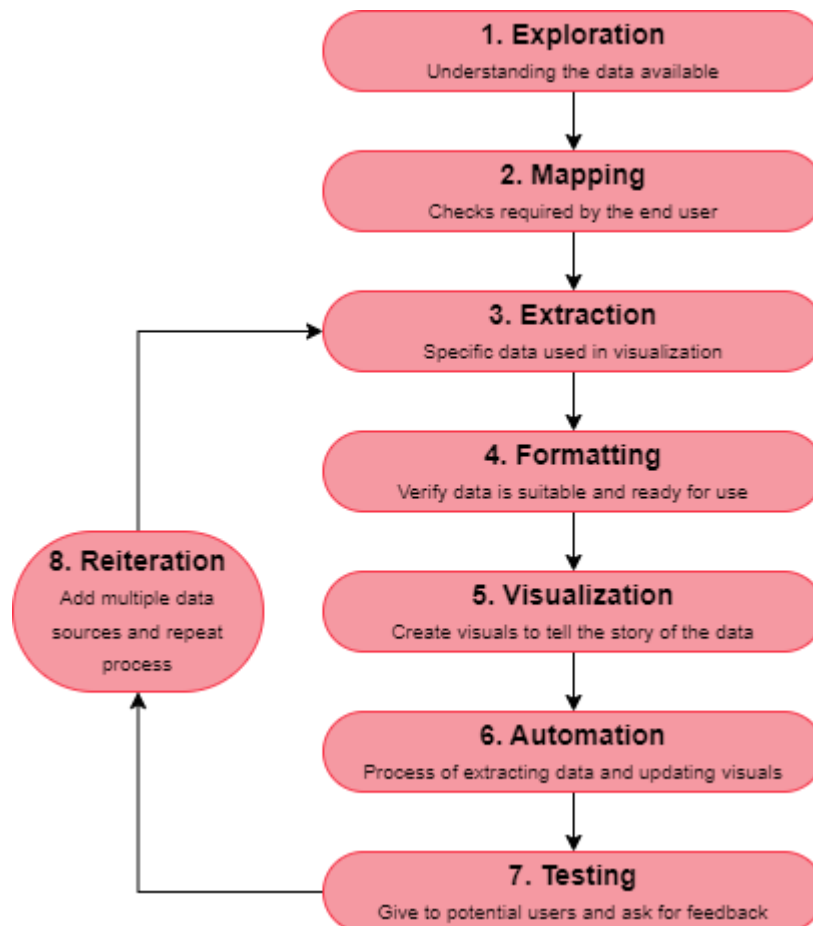
Figure 4 – Methodology Steps

It is worth noting that this chapter takes into consideration the prior knowledge of the reader's background with basic BIM software, methodologies, and standards. This includes Revit, ACC / BIM 360, ifc, Uniformat, Uniclass and more.

3.1. Data Exploration

So, what then is *Data Exploration*? Simply put it is just an understanding of what data is available and can be used, not a deep dive into the data. The author found it easier to communicate with the clients or potential users with a basic understanding of the location of parameters, columns available, the type of data (text, bool, number), the quantity of data, and the consistency of the data across different sources. Although it is simple, its value should not be overlooked, as it saves time by catching the inability of

the data to portray the users' requirements during the discussion phase and before the data is being manipulated and processed by the author.

The first step to data exploration is to open the data and make sure to know where it is stored. Usually, data can come in a database form such as an excel workbook or csv file and this is easy to visualize. BIM models are not as clear cut at first glance, but all the data extraction of a Revit model or ifc file tend to follow a basic database format as well. With the ifc's schema split into the four conceptual layers it might be easier for a user to extract their models in ifc first and examine it as either a text or using helpful software's to view IFC data such as "ifcQuickBrowser" which has benefits to viewing the data since it will link any referenced entities within an entity without having to scroll through the entire ifc file. (Liebich, et al., 2020)

It is also worthwhile to know the columns and type of data available in a database. This helps in knowing whether calculations can be made on some columns, if they are numbers for example, or whether some columns can only be used as text only even if they have some quantitative measure. An example of this would be columns with a quantity and its unit in the same data point, this cannot be read as just a number and therefore will require further manipulation in the future. Knowing this information can help guide the data analyst in preparing for the data extraction.

The quantity of data is also important to know. If the data source includes a large quantity of data, it is advised to use only a subset or filtered set of the data to ensure accuracies and ease in creating the dashboard. Given most dashboards required multiple projects to be included it is also wise to make sure the data exported from different projects are exported in the same way. Exploring this can include but is not limited to checking column names and order, checking sheet/table names to match, and alignment of the data. Especially with projects for large multinational companies, there are many inconsistencies between different projects in different places in the world, even simply the automatic unit of measure can be imperial or metric and this needs to be put into consideration for the remainder of the work.

The specificity of the AEC industry allows some crossover between big data development and the development of AEC software. Specifically, with the Autodesk Construction Cloud (ACC) Data Connector into Power BI. This ACC Data Connector allows for all project information in an enterprise to be directly linked into a Power BI dashboard and automatically updated. The availability of this has been instrumental in the development of real time dashboards that measure quality, but a prior reading of the Data Schema and Design is crucial to the data exploration phase. It helps create an understanding of the available data to be utilized in the dashboard and the links between different columns in the schema. An example of this can be shown in Figure 5 below from the ACC Data Connector reviews schema which is later used in the squad checks dashboard in Chapter 4: (Autodesk, Inc., 2022)

reviews

Describes the basic information of created reviews.

ordinal_position	column_name	data_type	constraints	notes
1	id	string: UUID		PK: UUID of a review.
2	sequence_id	number		The sequence id of a review.
3	instance_id	string: UUID		The instance id in the workflow service.
4	bim360_account_id	string: UUID		BIM 360 HQ Account ID.
5	bim360_project_id	string: UUID		BIM 360 HQ Project ID.
6	workflow_id	string: UUID		The workflow id of a review.
7	status	enum: string	**Possible Values:** • initiating • open • closed • void • failed • closing • processing	The status of a review.
8	review_name	string		The name of a review.
9	memo	string		The note of a review.
10	created_by	string		Id of user who create the review
11	next_due_date	datetime: ISO 8601		The next due date of a review.
12	created_at	datetime: ISO 8601		Creation time of the reivew.
13	updated_at	datetime: ISO 8601		Update time of the review.
14	docs_count	number		The count of reviewing documents of a review.
15	approved_count	number		The count of approved documents of a review.
16	rejected_count	number		The count of rejected documents of a review.
17	workflow_name	string		The name of a workflow.
18	started_at	datetime: ISO 8601		The time of the review started.
19	finished_at	datetime: ISO 8601		The time of the review finished.

Figure 5 – Autodesk Construction Cloud reviews Schema Columns (Autodesk, Inc., 2022)

3.2. Checks Mapping

Once the data is consumed by the user it is imperative to understand the end result before indulging in the work. Mapping requires either a conversation with the client or a better understanding of the end user's requirements. Once the author understands the specifics required at the end of the project, the author can then create a map that has the data, the required information and the final visuals. A connection between these three lists will help guide the remaining work of the dashboard. Chapter 4 shows a detailed example of the mapping across all projects but Figure 6 shows two examples to map the process:
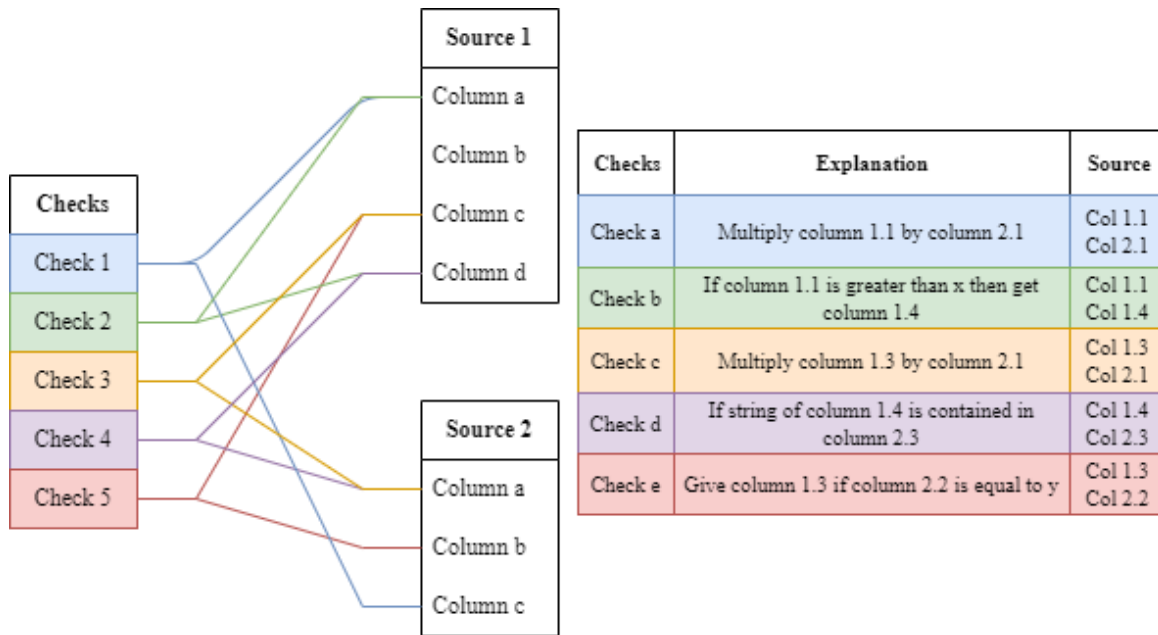
Figure 6 – Example Mapping Parallel Plot & Table

The parallel plot allows the author to visually understand the frequency of use of different data sets and which checks might require more time and development, while the table gives specifics of how each check is accomplished.

In the context of BIM and the EPC sector, these columns and checks can relate to quantitative or qualitative checks as required. Generally, a user can check if each element is classified correctly under the Uniformat standard and that would require to crosscheck two sources: the model element parameters and the Uniformat standard. This can be especially useful for big firms that have their own naming conventions that might change across projects, which allows for different sources for each project to be checked but still show an overall image of the company in one dashboard.

Another reason to map early is to know which columns might not be useful to be extracted. This will both ease the work on the author and will cause efficiencies in the data extraction and updating process of the project.

3.3. Data Extraction

To understand the methods for extracting data, the data needs to be precisely defined especially when it comes to BIM. BIM models have many properties, attributes, and metadata which can lead to an overburden of data that causes delays and inaccuracies in exporting and loading the data. It is imperative to export the correct data as required by the mapping done in the previous section. Now there are several methods of exporting data from BIM modeling software, but the purpose of this research is not to teach exporting methods but to help the reader understand the most efficient step to have data extracted that can be used by big data analysis and management.

Starting from this step, there will probably be iteration required when new databases or different sources with the same data are added. This happened multiple times throughout the process of this research,

sometimes the client wants to add new projects, old projects, export from different sources to streamline the process without requiring the conversion of model folders on ACC / BIM 360 to be exported first; there are many different changes that might happen. The important thing is to be able to replicate the data extraction process in the future and build on what was done before to be able to combine multiple projects.

When extracting data for the purpose of creating visualizations of project progress, there are a few steps to ensure are correct. If folders are being used as sources, all files inside that folder need to have the same structure, meaning the same sheet names, the same columns structure and other points discussed in Section 3.1. Some clients might ask for extraction from their federated models on Navisworks. Again, creating an exporting template to ensure each csv export matches the ones already in the project is crucial.

Another method of extracting data is through the automatic extractions available to users through Power BI connectors such as the Autodesk Construction Cloud (ACC) Data Connector or the Microsoft Fabric platform. These platforms make extraction a simple click and sign-in since there is a backend streamline that directs the data into specific queries to be used directly by the user. The automation of these extractions makes the work easier in extracting but might not allow for the customization and control of the extraction when the user creates their own extraction methods. With big data this might create problems with the sheer amount of data imported into the dashboard and the best way to mitigate this is by creating sample projects on these platforms to begin working with until the dashboard is ready for the entire enterprise/folder to be imported.

Once the data is exported and ready to be imported into an analytics platform to be visualized, the author recommends a quick glance at the data before any formatting. This will help the reader prepare for the process of formatting the data.

3.4. Formatting

Formatting data might be the most important step in analyzing the data and presenting it in a simple and concise manner. Although simplicity is required, formatting might be the most complex part of building the analytics. The process of formatting data can be split into three basins: cleaning, creating, and linking. The order of these steps is not necessarily chronological and there can be steps that go back and forth depending on the specific requirements of a project. In the following paragraphs, the author will attempt to discuss each "basin" in depth, but there are many different methods to achieve the results a client may be asking for and some methods might not be fully discussed here.

3.4.1. Cleaning

Cleaning as defined by the author takes place when data needs to be adapted or fixed singularly. This means that it does not depend on other data points but just needs adjustments on a singular column. This can be anything from formatting text to converting units of measure. Table 3 shows several examples of different methods of cleaning the data:

Table 3 – Cleaning Data Examples

Type	Explanation	Example
Text	Manipulation of text in data points such as trimming, extraction, renaming, or grouping.	Extracting the name of a project from model elements can be useful in mapping each element to a project.
Number	Modifying numbers in a singular way, any arithmetic or geometric calculations.	To unify projects units, all units can be changed to metric by multiplying by their relative conversion ratio.
Error	Removing errors that may be caused by wrongly inputted data points. If the data does not follow the data type it would be wise to remove and flag those errors.	Removing duplicates from columns that need to be mapped to just one project. Or if text is inserted into the "volume" column.
Filtering	Filter the dataset based on specific requirements in a column.	Requiring just one project to be analyzed or requiring a certain minimum threshold to be used for "weight" in a visual.
Anomaly	If a data point has such a large variation that it causes a skew in the visualization it is better to set upper and lower boundaries for some columns. Given the data is big it becomes a negligible loss of data.	Some quantities might be inputted incorrectly and having a limit to how much cubic meters of concrete can be poured in a day allows these inconsistencies to not affect your final presentation.

3.4.2. Creating

Another method of data formatting is creating, which creates new data based on the data available. This type of formatting is the more common way to create the required calculations and equations that help drive the visualization. There is a plethora of examples of how to use it and the more important ones used in this research will be further explored in the next chapter with the case studies, but Table 4 below shows the generally used examples of data creating:

Table 4 – Creating Data Examples

Type	Explanation	Example
Addition	Often new data points need to be added to the dataset.	Adding a new column defining the structure type in the dataset to be able to use as filter later in the project.
Condition	Creating columns based on other columns conditions.	Weight ranges can be classified as light, medium, or heavy weight steel.
Combination	Combining two columns to create new columns.	Concatenating model object code and family can give a unique identifier for each object.
Split	Splitting a column into multiple columns to extract some information.	Splitting model names to extract project name and discipline.
Transformation	Transforming columns into new columns that can reference other columns.	Changing soil conditions rules based on depth of excavation.

3.4.3. Linking

The final way to format data is by linking data to each other. This focuses more on the connection between different data sources which then allows for the cleaning and creating of additional data with conditions from different sources. This sounds complex but it can be clarified with a very simple example. Sometimes a contractor has two different databases for model elements that will be built, one modelled by the BIM department and another by the estimation department as they are setting up their bid. A contractor would want to compare the quantities set up by these two departments with ease across all their projects. The solution to this is by linking the object name columns from the two databases and then creating a comparison between the two quantities through that link. This sort of linking can also be used to set up rules which may differ from project to project and can be changed by the project users throughout the lifetime of the project. This will help the user test multiple scenarios on the fly without having to recreate multiple visualizations. Linking can be very powerful when combined with the cleaning and creating formatting tools and with the three being interchangeably used back and forth, a data analyst can create powerful visualizations that give true insight to an EPC firm's model health and quality.

3.5. Visualization

The work done so far can be done correctly, but the inability to present it correctly can negate all the positive progress. When working on visualization, a data analyst must change hats from a scientist to an artist who is deeply involved in the ease and accessibility of their work for others to use. The main objective of this subchapter is to discuss the selection of the layout, charts, filters, and color scheme.

3.5.1. Layout

When creating the dashboards, the author created a generic layout that can be repeated for different dashboards. The goal was to create a dashboard that was easily navigable, clear and with enough space to let the charts have significance. The first thing to create was the background and title bar, when thinking of this, the simpler the better to not take any attention away from the charts. Figure 7 shows the generic layout followed by an explanation of the elements within the layout.
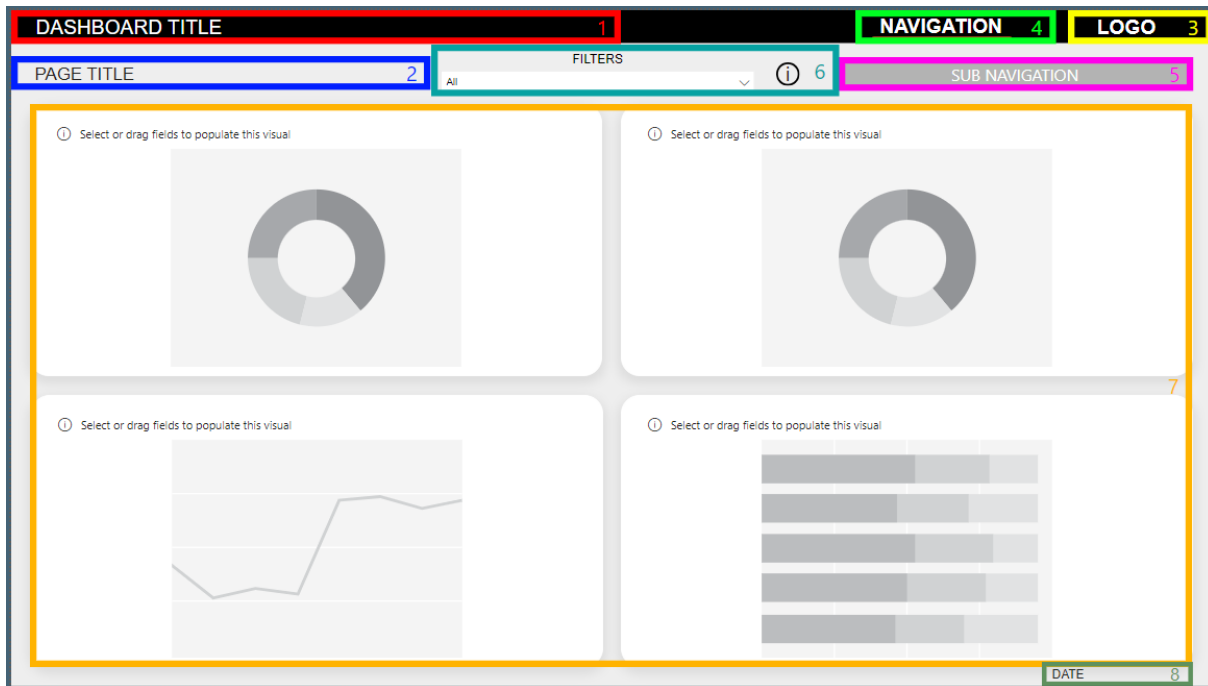
Figure 7 – Generic Dashboard Layout

The elements can be described below:

1. Dashboard Title – clear and concise title that repeats for the entire dashboard.
2. Page Title – explains what the page will demonstrate with the charts.
3. Logo – company logo
4. Navigation – bookmark navigations that allows navigation between main pages.
5. Sub-Navigation – bookmark navigation that allows navigation between sub-pages.
6. Filters – although Power BI has its own filter panel, the author found it easier for the users to have filters available on the page and it was more practical to control available filters.
7. Chart Area – area where all charts will lay. As visible it is good to give adequate spacing between elements to let them breath. The author kept a 20-point space between visuals and if there are more visuals inside the chart area, the spacing was upheld and evenly distributed based on the number of charts and the sizes required to emphasize them.
8. Date – date of data export, helps give insight into what the data represents.

3.5.2. Charts

Once the layout is set the rest of the elements can be added, specifically the charts. The charts are the most important elements in the dashboard since they carry all the data. Power BI carries 39 default visuals with the ability to search for many more online created by users and the Microsoft team. This chapter will discuss the charts most used for this research and the reason to use specific charts.

The first type of chart is the column/bar chart. There are six different variations that can be seen in Figure 8 which include the stacked, cluster, and 100% stacked charts. Bar charts are better suited for longer labels while column charts are useful for chronological or structured data in order. The stacked chart is useful for comparing totals of the categories while splitting them with the different variables,

while the clustered chart is more useful for comparing the variables since the categories have the split within them. The 100% stacked chart is useful to see progress based on percentage not on value.
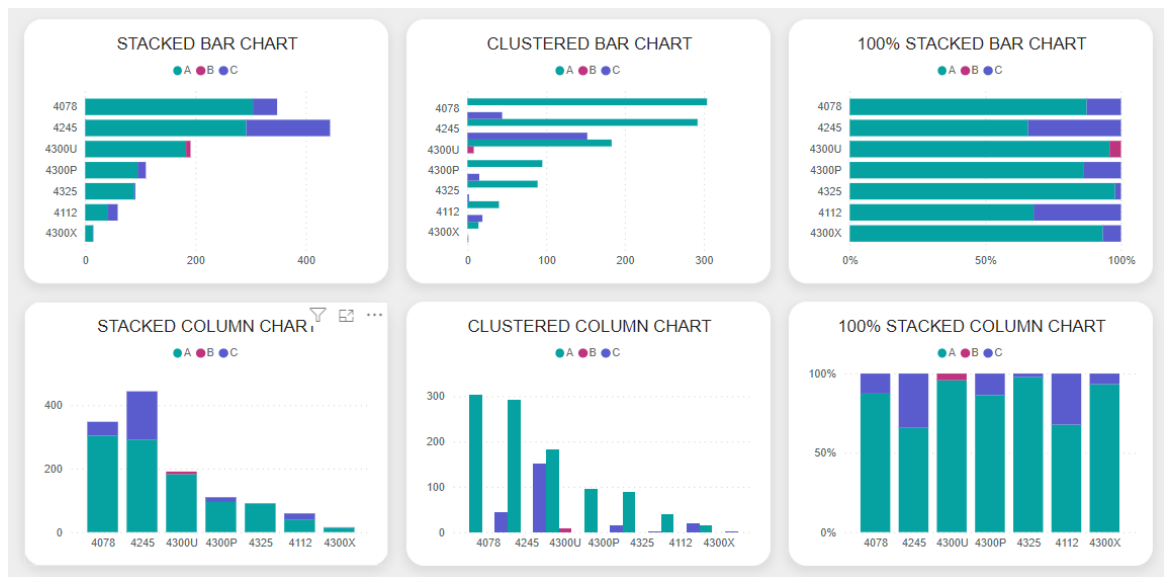
Figure 8 – Bar/Column Charts

Following the bar charts, the next most popular charts are the line charts as shown in Figure 9. Line charts are useful for comparing variables over time. The three most popular charts that follow lines are the line, area, and stacked area line charts. Line is simply used to compare values over time, the area chart is useful to see when the difference in density between the same variables, and finally the stacked chart allows for the addition of variables to similar to the stacked bar/column charts but over time.
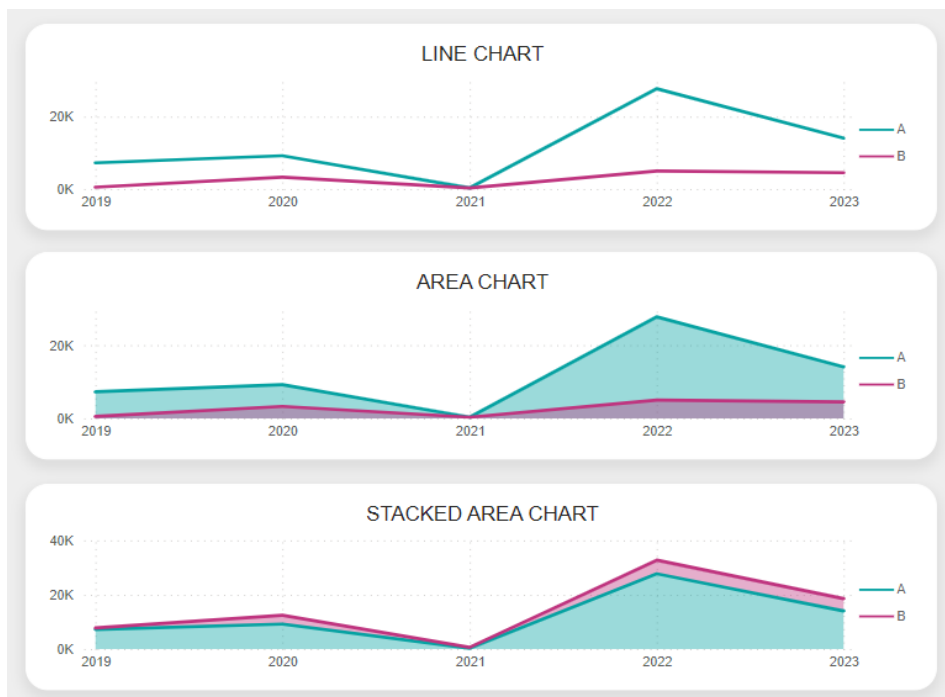
Figure 9 – Line Charts

Progress can be used in both bar and line charts, but the author found the gauge more useful. As shown in Figure 10, it allows to show the progress of work with a target value and a maximum value.
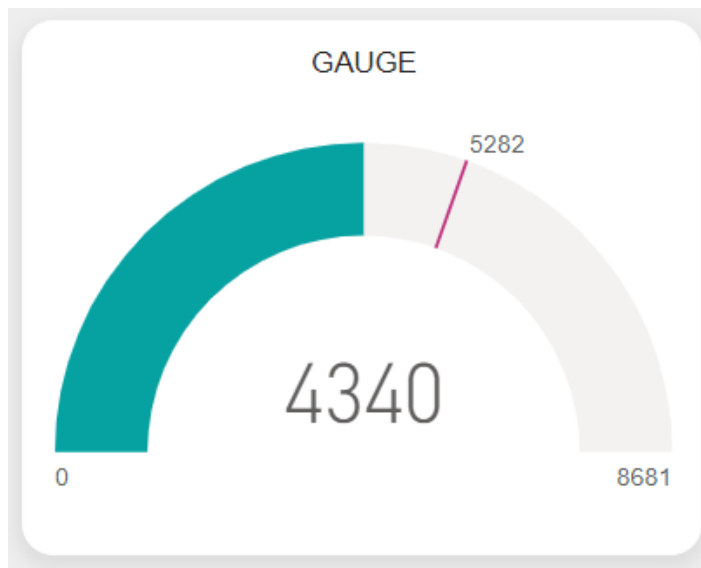
Figure 10 – Gauge

Another type of chart used are the "volume" charts, which includes the pie, donut and treemap charts as shown in Figure 11. These help users understand the quantity difference between categories. The author found the pie and donut charts to be very similar, but the donut chart was cleaner and clearer to read so it was used more often. The treemap on the other hand, helps show size differences visually better which can be used for project sizes while the other two can be used for completion percentages and other categories.

Figure 11 – Volume Charts

The final visualization used in the research was not a chart but rather two types of tables. A simple table to show databases and exact data points and a matrix which allows a user to investigate different columns

and values without opening an entire table and scrolling through it. Figure 12 shows both these examples.

TABLE

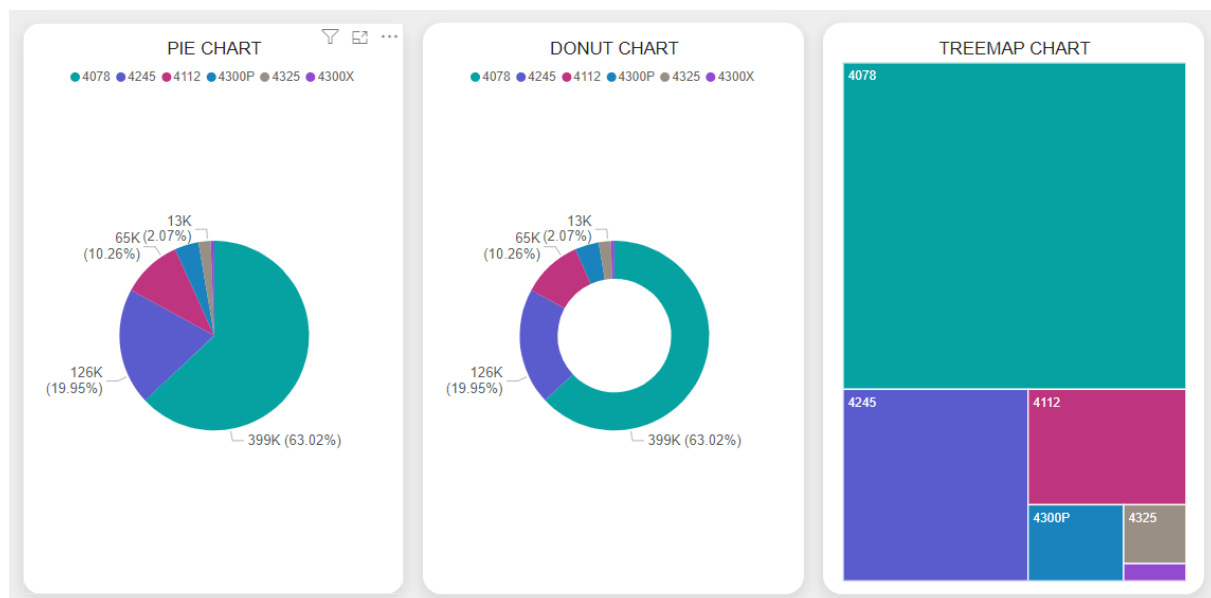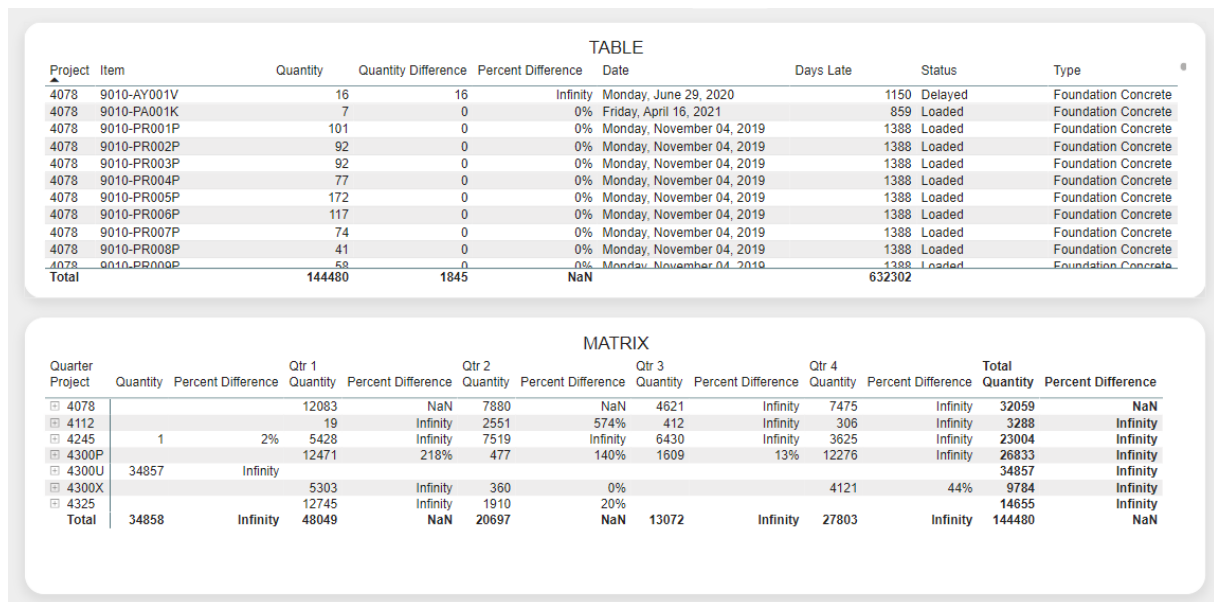| Project	Item	Quantity	Quantity Difference	Percent Difference	Date	Days Late	Status	Type
4078	9010-AY001V	16	16	Infinity	Monday, June 29, 2020	1150	Delayed	Foundation Concrete
4078	9010-PA001K	7	0	0%	Friday, April 16, 2021	859	Loaded	Foundation Concrete
4078	9010-PR001P	101	0	0%	Monday, November 04, 2019	1388	Loaded	Foundation Concrete
4078	9010-PR002P	92	0	0%	Monday, November 04, 2019	1388	Loaded	Foundation Concrete
4078	9010-PR003P	92	0	0%	Monday, November 04, 2019	1388	Loaded	Foundation Concrete
4078	9010-PR004P	77	0	0%	Monday, November 04, 2019	1388	Loaded	Foundation Concrete
4078	9010-PR005P	172	0	0%	Monday, November 04, 2019	1388	Loaded	Foundation Concrete
4078	9010-PR006P	117	0	0%	Monday, November 04, 2019	1388	Loaded	Foundation Concrete
4078	9010-PR007P	74	0	0%	Monday, November 04, 2019	1388	Loaded	Foundation Concrete
4078	9010-PR008P	41	0	0%	Monday, November 04, 2019	1388	Loaded	Foundation Concrete
4078	9010-PR009P	58	0	0%	Monday, November 04, 2019	1388	Loaded	Foundation Concrete
Total		144480	1845	NaN		632302		

MATRIX

Quarter / Project	Quantity	Percent Difference	Qtr 1 Quantity	Percent Difference	Qtr 2 Quantity	Percent Difference	Qtr 3 Quantity	Percent Difference	Qtr 4 Quantity	Percent Difference	Total Quantity	Percent Difference
⊞ 4078			12083	NaN	7880	NaN	4621	Infinity	7475	Infinity	32059	NaN
⊞ 4112			19	Infinity	2551	574%	412	Infinity	306	Infinity	3288	Infinity
⊞ 4245	1	2%	5428	Infinity	7519	Infinity	6430	Infinity	3625	Infinity	23004	Infinity
⊞ 4300P			12471	218%	477	140%	1609	13%	12276	Infinity	26833	Infinity
⊞ 4300U	34857	Infinity									34857	Infinity
⊞ 4300X			5303	Infinity	360	0%			4121	44%	9784	Infinity
⊞ 4325			12745	Infinity	1910	20%					14655	Infinity
Total	34858	Infinity	48049	NaN	20697	NaN	13072	Infinity	27803	Infinity	144480	NaN

Figure 12 – Tables

3.5.3. Filters

Power BI has a built-in filters panel that can be published with the dashboards and reports. Although this is useful for common users of Power BI, the author found it easier to use the "Slicer" visualization to create user-controlled filters. The benefit of having these slicers also allowed for filtering the slicer before giving the user control of it. For example, it allows the author to filter by specific projects first before allowing the user to pick a project. Since it was a visualization, it could also be controlled to filter only specific visuals which allows for more customization and control based on the clients' request.

It is recommended that the filters are simple and repeated across the entire dashboard if possible. If some pages might need additional filters, the author implores for the consistency of location of filters from page to page to ease navigation. Another recommendation will be the ability to understand how the slicer selected will be used; if, for example, the list of options is long and can be typed, it is recommended to allow a text search in the slicer. On the other hand, if it is a short list of letters, it may be worthwhile to use a tile setting so that the user can just click the option they want to see on their sheet. The final recommendation would be to limit which visuals are affected by the slicer since some will become blank or only a single value if the slicer affects it so. An example of this is if a visual shows the quantity by item type and there is a slicer of item type, once a specific item is selected, the entire chart will show just one thing. If this is not beneficial to the user, the author may edit the interactions between the slicer and the visuals that do not need to be affected by it.

3.5.4. Color Scheme

Although the author did not have much freedom in the color scheme due to client requirements, there are many options available to create a color scheme that compliments and makes the use of the dashboard

pleasant for the user. The author recommends using two sources for help in the design of colors and shapes provided by Google and Adobe.

Material.io by Google helps website designers with colors, shapes and design for a website, but can be utilized the same way for a dashboard. The Adobe color wheel allows a user to select a color and generate a complementary color wheel for the design of anything. Figure 13 shows the Adobe color wheel which allows a user to select a color and obtains the complimentary color palette. (Adobe, 2023) (Google, 2023)
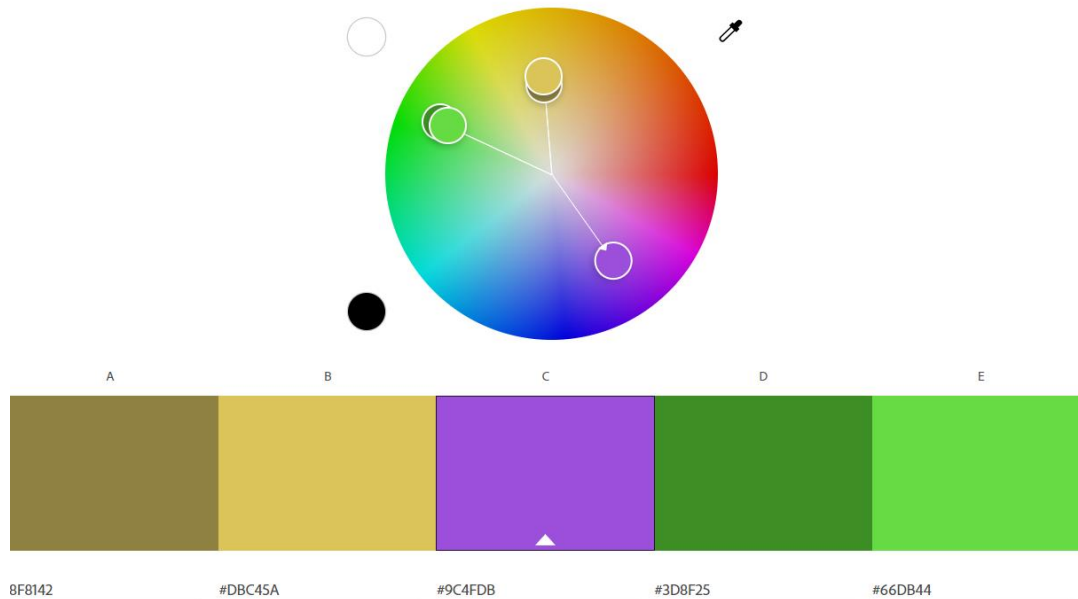

Figure 13 – Adobe Color Wheel

3.6. Automation

Automation with Power BI is simple and intuitive to most users. There can be three ways to automate a dashboard: manual refresh, semi-automatic refresh, and automatic refresh. Each one has its respective steps and availability based on client requirements and data location and availability. It is intuitive that the fully automatic refresh is preferred but it is not always available. There are three steps that take place that can be automated or not and this is where the different levels of automation differ. The three steps are:

1. Exporting the data
2. Refreshing the dashboard
3. Publishing the dashboard again

Manual refresh is rudimentary but might be required for most BIM applications since most data sources do not link directly to Power BI. Manually refreshing includes following the three steps above and an organization of folder location that Power BI reads from. The first step is exporting the data again from the source such as the Navisworks csv. Then refreshing the Power BI dashboard on the Power BI Desktop and finally publishing.

Semi-Automatic refreshing can be utilized when the folder used for refreshing is interoperable with Power BI such as a Microsoft SharePoint Folder. It is beneficial since it will elude the third step or re-publishing the dashboard. If the Power BI dashboard has a dataset that can be read from a SharePoint, the refresh can be done directly on the Power BI Service web-based platform by scheduling a recurring refresh of the dataset from the SharePoint folder. This bypasses the need to even open the Power BI Desktop application but requires a Gateway setup that can be limited by the client based on security levels. This method still requires the manual updating of exporting the data into the SharePoint folder.

Finally, the fully automatic refresh can bypass the manual exporting by creating a link between Power BI and the data source directly. In the BIM world, this can be done by the Autodesk Construction Cloud (ACC) Data Connector which allows a direct stream of data from ACC to Power BI. When this is combined with the scheduled refreshing on the Power BI Service and a correctly set Gateway, then every time the scheduled refresh takes place, the dashboard will read data directly from the ACC Data Connector and require no user intervention. This is a great way to completely rid of any manual application but requires several levels of security clearance and is still limited to Autodesk applications currently. The author believes there are other methods of creating this automation through the new Microsoft Fabric and Power Platform capabilities, but these exceed the limits of the authors knowledge and were not required for this research.

3.7. Testing & Reiteration

Once the dashboard is ready for use, it is advisable to allow a testing period of a month to allow the potential users to find bugs, problems, and inconsistencies that might occur from the work. It is helpful to then gather all this information and repeat the steps provided above starting with extraction. Remember that during the first run of the creation, it is easier to work with a small dataset so when reiterating the work new problems may arise with the addition of all the data available. In testing the extraction and formatting efforts might be more difficult if more data sources are being introduced, but the visualization effort should be complete from the first time (unless of course the client wants changes). Reiterating the work may be cumbersome, but it is necessary and may be a continual effort throughout the lifetime of the dashboard, especially with the consistent changes in the AEC industry. Finally, when the dashboard is ready for deployment on a company level, it is the client's responsibility to share it with the correct email lists, but the author must ensure that read-only access is the one that is mass published to avoid changes done by other users.

The above steps were not the first method the author followed when creating the dashboards discussed in Chapter 4, but they became the standard way to work through a period of trial and error. The author would like to reiterate that this methodology was helpful for this research, but it can be changed and adapted according to the work being undertaken. The next chapter will go through the five dashboards created by following the same methodology stated above with some deviations based on the requirements of the dashboard. For example, the layout was done only once and the repeated for all five, while some extraction methods were just repeated since the source data was similar. The purpose of this is to give a guideline that can be followed but is not strict to creating new dashboards in an EPC firm looking to manage the vast amount of data created throughout the project, especially to monitor the quality of data across different data sources.

This page is intentionally left blank

4. CASE STUDY

During this research, the author worked for MAIRE S.p.A daily in creating Power BI dashboards as part of the Civil Department. MAIRE leads a technology and engineering group that develops and implements innovative solutions to enable Energy Transition. MAIRE offers Sustainable Technology Solutions and Integrated E&C Solutions in nitrogen fertilizers, hydrogen, circular carbon, fuels, chemicals, and polymers. Tecnimont S.p.A is MAIRE's large-scale EPC contractor. In Tecnimont there are many departments responsible for the construction planning and drawings of the plants. Among these departments are the piping, civil, electrical, building, instrumentation, health and safety. The author was responsible for multiple dashboards, mostly concerning the quality of BIM models in the Civil department but one dashboard focused on communication delay on Squad Checks from the Autodesk Construction Cloud (ACC).

4.1.   Company Hierarchy and Model Creation Process

In order to understand how the dashboards were created, the hierarchy of how models and figures needs to be discussed first. Company hierarchy starts and ends with the Piping department since the projects focus on the piping mostly. Once the piping sets the location and sizes of equipment and pipes according to the process diagrams, the other departments working on a project place their required elements and then the Piping department again reviews the work of the other departments and finalizes the models by creating one federated model that combines all models.

Given the models are the communication factor between departments, it is imperative to know how data is transferred through the model creation process. Figure 14 shows the entire process of model creation and data sharing which will provide the data sources used by this research in the creation of the dashboards. The modeling is created using many modeling software, but in the Civil Department mostly Autodesk Revit and Tekla Structures are used, this is the starting point of most data.

Once the models are ready, they get updated to the Common Data Environment (CDE) of either Autodesk Construction Cloud (ACC) or BIM 360 to be shared with all parties of the project, as seen in the first branch of Figure 14. MAIRE Tecnimont's Civil Department created a web-based database, the Civil Item List (CIL), as a quantities controller to congregate all official project data from drawings schedules in one place. A new version of the CIL was created to not only have a list of items, but also have linked 3D models. This version is called the Civil Master Data (CMD) and was created used Autodesk Forge from the ACC / BIM 360 connection. Although the CMD now includes the 3D model, the quantities and properties of the main items is manually updated by Civil department discipline leaders based on official drawings and drawing schedules.

The second branch of Figure 14 is created by the Piping Department. To create a federated model from all models, the Industry Foundation Class (ifc) extraction file is used for each Revit or Tekla model and federated in Navisworks. This federated model then has all the same properties and all the models that are required to build the project. The third branch of the figure shows an export of all the model elements based on AVEVA engineering rules. AVEVA is a database that the Advanced Work Package (AWP) team uses for construction quantities and parameters. Since it requires specific importing rules, an

application programing interface (API) was built to export an excel workbook with the exact tagging conventions that the company requires and that can be imported into AVEVA Engineering.
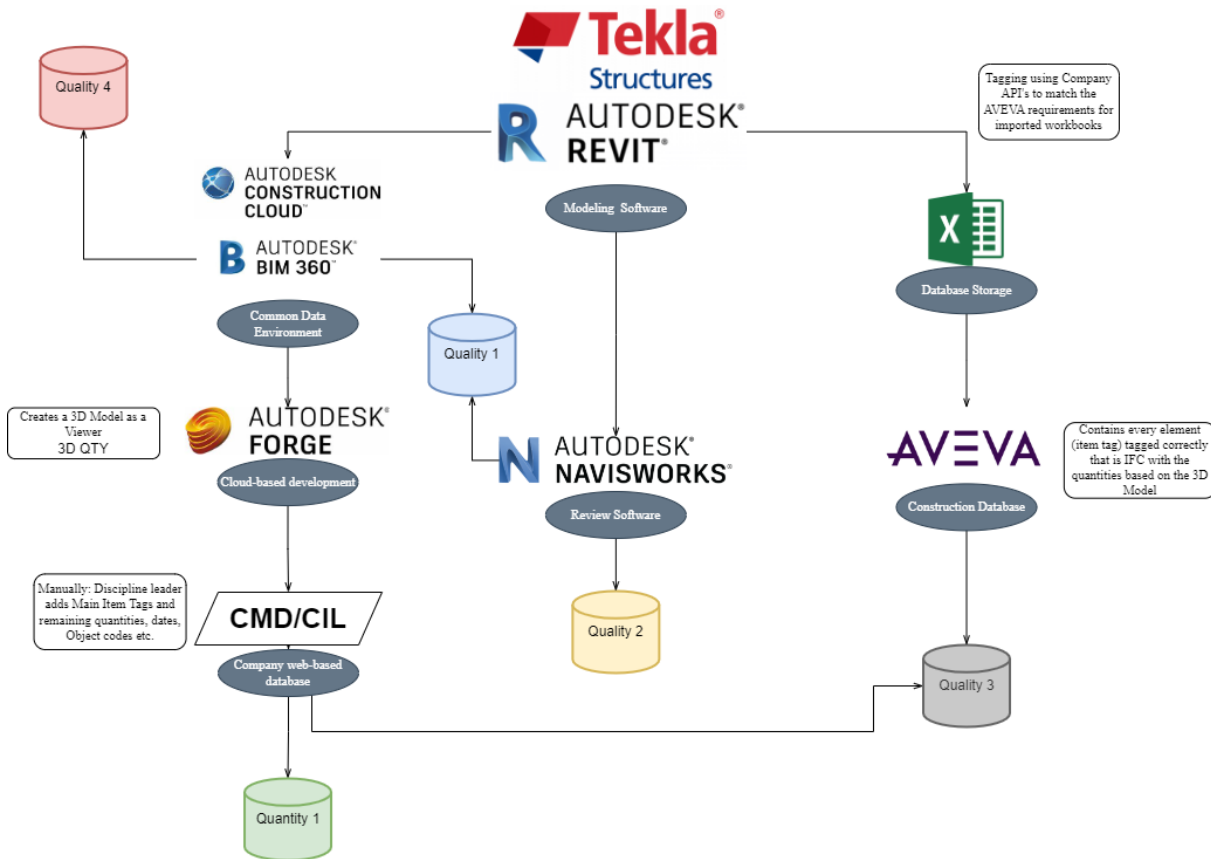

Figure 14 – Data Transfer

Once these branches of data transfer are understood a quick explanation of what each database for dashboard can be found below:

- Quality 1 (blue) – responsible for checking model file location in ACC / BIM 360 and the Federated Model gets model lists from both sources and compares between them as well as checking rules within them.
- Quality 2 (yellow) – check element attributes. From the federated model all elements are exported to check for properties and all tagging rules according to company standards.
- Quality 3 (grey) – checks alignment between the manually inputted main items in the CMD / CIL and the exported AVEVA elements. Data coming from both sources should align, not only in name of main items but also the quantities inputted.
- Quality 4 (red) – checks the squad check reviews. From ACC / BIM 360 some reviews are created to ensure model accuracy and require input from the many departments. To ensure departments are on time, this dashboard checks which users and departments are late and how late they are in their reviews. The data is directly imported from ACC / BIM 360 using the ACC Data Connector.

- Quantity 1 (green) – checks the overall quantities of projects across the entire company. Data from the CMD / CIL is used since these quantities are the official quantities from the signed drawings schedule.

Now that the data is transparent and known in different areas, a folder structure for each dashboard needs to be designed. This can be seen in Figure 15, as there are three folders the company can use once the research is completed.

1. Dashboards Folder: contains all the dashboards.
   a. Rules Folder: contains Excel Workbooks that have additional rules for the dashboards.
2. Data Sources: where data is stored and can be updated regularly.
   a. Models Lists: the lists of models from the CDE and the Federated Model.
   b. Model Attributes: files that include element properties from the Federated Model.
   c. CMD/CIL: Export of element properties from the CMD / CIL.
   d. Object Lists: AVEVA export of element properties.
   e. Current Estimates: Historical trend of estimates for projects.
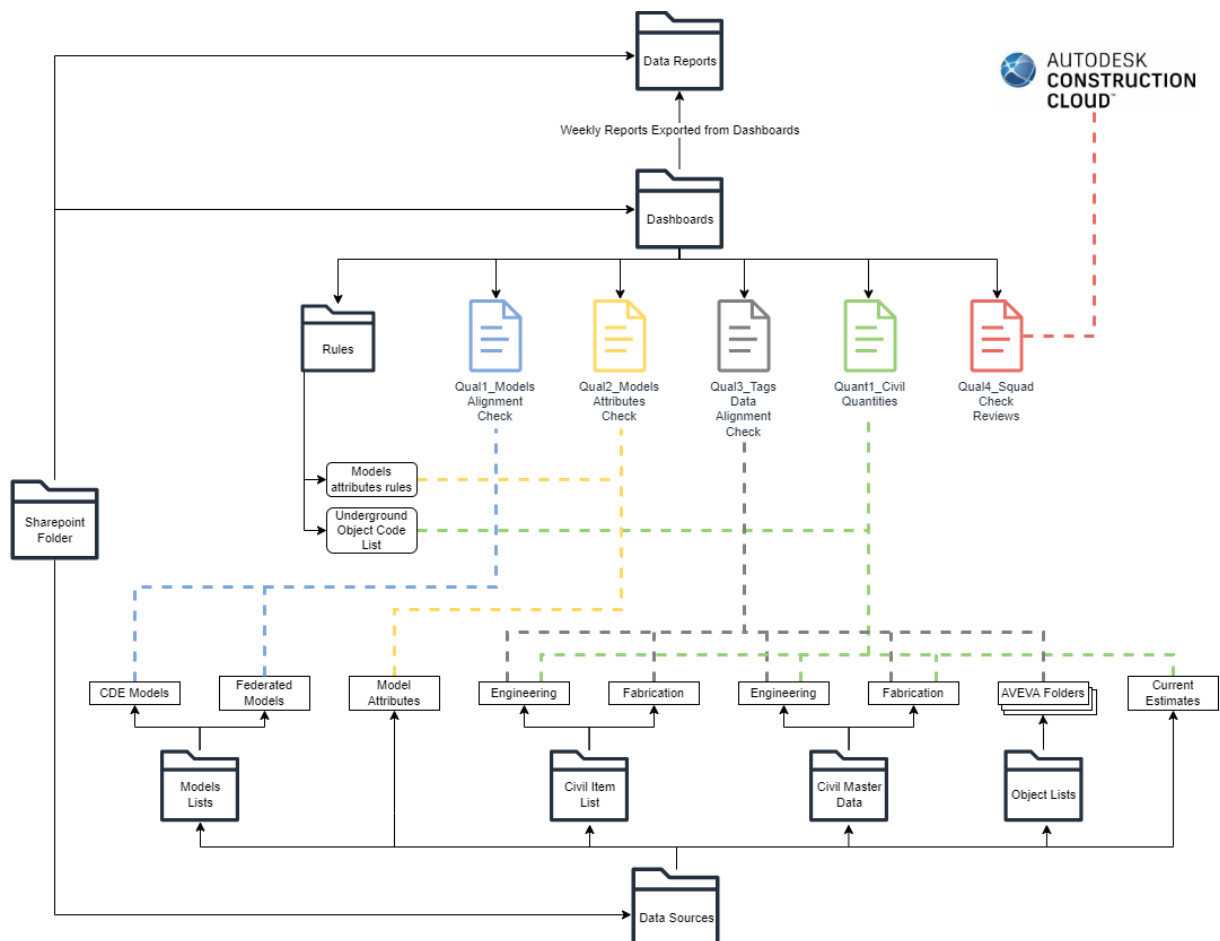3. Data Reports: where exported reports of the dashboards will be stored in the future.

Figure 15 – Folder Structure and Dashboards

The type of automation, as described in chapter 3, for most of the dashboards was semi-automatic, which required folders/SharePoint folders to store the data. While Quality 4 was fully automatic due to the

connection to the Autodesk Construction Cloud. With this explanation, the author is ready to delve deeper into the development of each of the dashboards and how they came to be from the original sources to the manipulation and visualization of the final dashboard.

4.2. Quality 1: Model Alignment Check

The first dashboard was created to check the ACC / BIM 360 model location and alignment between different folders. As shown in Figure 15, this check gets the data from names of models in the Common Data Environment (CDE) and the Federated model. During a project, there is a multi-step process of creating and sharing models with all disciplines, so this dashboard allows the checking of the models and their sequential progress to sharing and transferring.

4.2.1. Exploration and Mapping

The data for this dashboard was simple; it contained multiple lists of model names within the different folder structures in the CDE and the Federated Model. Each folder has a specific role in the progress of a model and the ability to transfer it, explained below:

- ACC / BIM 360 – shareable online environment to store and share models, with the following folders created for the phases a model might go through:
  - Work in Progress (WIP): Location of models that are currently being worked on or are finished, all models start here and get pushed to different folders as they develop. Access is restricted to the discipline.
  - Shared: Models that are ready to be shared within the discipline to be evaluated, reviewed and confirmed that they are accurate by team leaders. Check clashes in ACC / BIM 360.
  - Collaboration/Consumed: Models that can be used and linked by other disciplines in their cloud Revit model. Used for visual checks and clashes in Revit.
  - Transfer: Models that are ready to be shared with all project disciplines and should be included in the federated model of the plant, no matter the status of the model, which then includes every update to be shared with others. Must be industry foundation classes (ifc) models. Official clashes done from this folder using Navisworks.
- Federated Model: this model combines all the ifc files extracted from all models and appends them together in Navisworks both for clash detection and for model alignment. The Piping department is responsible for this as they are on the top of the command chain.

Once the data is known, the next thing to organize was the checks required for this dashboard. Figure 16 shows the mapping of the checks created from the two file sources.
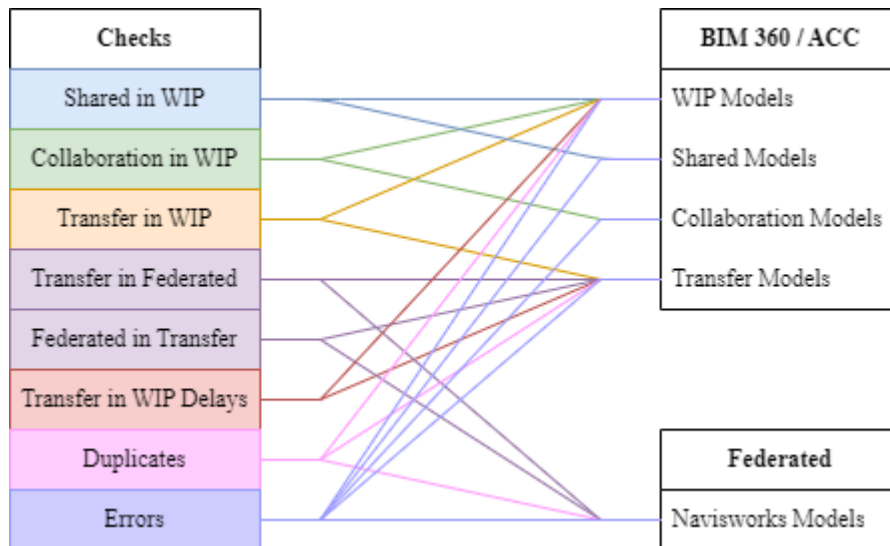
Figure 16 – Quality 1 Mapping

These checks are further explained below:

- Location
    - Shared files in WIP: all files in the Shared Folder need to be originally in the WIP, sometimes a model is put in the Shared Folder then updated with a new name in the WIP, must be re-updated in the Shared Folder and the previous version deleted from the Shared Folder.
    - Collaboration files in WIP: follows the same logic as above, all files in Collaboration must originate from the WIP and if changes occur, they must be reflected accordingly.
    - Transfer files in WIP: follows the same logic as above, all files in Transfer must originate from the WIP and if changes occur, they must be reflected accordingly.
    - Transfer files in Federated: Transfer files must all be existent in the aggregated Navisworks file and if a file is missing it must be checked whether it needs to be removed from Transfer since it is no longer relevant or it was missing from the aggregated Navisworks file.
    - Navisworks files in Federated: Files that may not be showing in Transfer that should be there based on the Navisworks. This also requires coordination on whether the models require to be in the Navisworks file and are missing in the Transfer or if they should not be added in the Navisworks to begin with.
- Delays
    - Transfer in WIP Delays: Transfer files should be regularly updated with relation to the WIP files. If a WIP file has been updated and the Transfer file has not been updated in more than 10 days since the WIP file has been changed, then the Transfer file is considered "Outdated."
- Duplicates
    - WIP, Transfer, Navisworks: only these three folders are checked for duplicates because the other folders are set in a way where there cannot be duplicates from their ACC / BIM 360 folder structure.

- Errors
  - This entails all errors that might arise from wrong files being uploaded to the ACC / BIM 360 folders, the errors are as follows:
    - All files must either be rvt or ifc files, any other formats are errors.
    - In steel files: an AIVD file becomes an AIVD-3D file when it is shared with a Vendor and published by them and therefore any AIVD-3D file cannot also be an AIVD file (considered a duplicated file or old file)

When the entire dashboard is organized, it is then time to focus on extraction, formatting and creating the visualizations to translate these checks into reality.

4.2.2. Extraction

Data extraction consists of two sources: the CDE and the Federated model. The CDE used is ACC / BIM 360 which allows for Project Files extraction which gives a report of all files in their respective folder location. This extraction is in the form of an Excel Workbook with column headers such as "Folder Name and Path," "Name," and "Last Updated" among others. Using these three columns was enough for the dashboard checks. The second list of file names was generated as a csv from Navisworks. This file included just one column, "Item Name" when filtering by models as required by the dashboard.

4.2.3. Formatting

Formatting includes multiple steps as explained in Chapter 3.4, for this chapter, the author will define the queries within the Power BI Power Query Editor. Then the author will show the data linking between the sources. Finally, any important steps or equations that are specific to this dashboard will be highlighted.

Table 5 shows the queries and their sources:

Table 5 – Quality 1 Queries

Group	Query	Source	Objective
Source Folders	*CDE Folder*	-	Allow for easy change in folder location which allows for PBI template to be shared with folder source to be changed.
	Federated Folder	-	
CDE Group	*CDE*	*CDE Folder*	Combine all files from CDE
	Projects	*CDE*	List of unique projects
	Delays	Merged: 01c & 04b	Find date difference between Transfer files and WIP files
Federated Group	05 Federated	*Federated Folder*	Create list of models by project name and find if there are any duplicates or not
WIP	*01a WIP errors*	*CDE (filter 'WIP')*	Finds errors based on file extension
	01ai WIP AIVD errors	*CDE (filter 'WIP')*	Find errors based on steel parameters
	01b WIP removed consumed & extensions	*CDE (filter 'WIP')*	List of files to be compared with Collaboration folder, leaves Building in because both are present in Collaboration
	01c WIP removed consumed, BB & extensions	*CDE (filter 'WIP')*	List of files to be compared with Shared Folder, Transfer folder, and check for duplicates
Shared	*02a Shared errors*	*CDE (filter 'Shared')*	Find errors based on file extension
	02ai Shared AIVD errors	*CDE (filter 'Shared')*	Find errors based on steel parameters
	02b Shared removed BB & extensions	*CDE (filter 'Shared')*	List of files to be compared with WIP folder
Collab	*03a Collab errors*	*CDE (filter 'Collab')*	Find errors based on file extension
	03ai Collab AIVD errors	*CDE (filter 'Collab')*	Find errors based on steel parameters
	03b Collab COL_AX folder no extensions	*CDE (filter 'Collab')*	List of files to be compared with WIP folder. Only checking the specific folder that is consumed. (COL_AX)
Transfer	*04a Transfer errors*	*CDE (filter 'Transfer')*	Find errors based on file extension
	04ai Transfer AIVD errors	*CDE (filter 'Transfer')*	Find errors based on steel parameters
	04b Transfer AA/AI/AQ no extensions	*CDE (filter 'Transfer')*	List of files to be compared with WIP folder, Navisworks folder, and check for duplicates
Errors	All Errors	Append of all: 0xa errors	One query that combines all errors
	AIVD Errors	Append of all: 0xai AIVD errors	One query that combines all AIVD errors

Italicized queries are not loaded into the Power BI Model

Once the query editor work is finished it is important to connect the resulting tables to be able to compare the file path names across different elements. Figure 17 shows such connections:
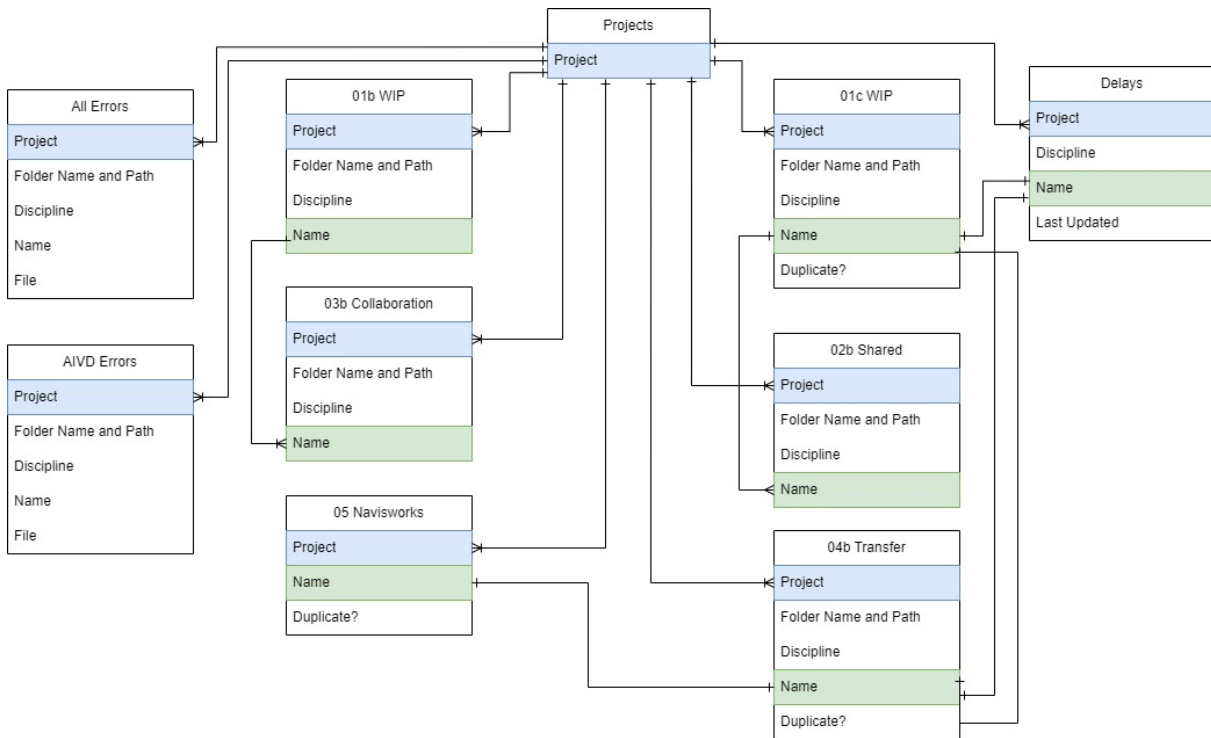

Figure 17 – Quality 1 Data Connections

With the connections made it is now possible to check whether a certain model "Name" is included in other queries through the relationships shown in Figure 17. A sample equation of this would be as follows:

```
SHARED IN WIP =
IF (
    CONTAINS ('01c WIP removed consumed, BB & extensions','01c WIP removed
    consumed, BB & extensions'[Name],'02b Shared removed BB &
    extensions'[Name]),
TRUE(),
FALSE())
```

This is then repeated for all other folder checks.

Finally, when all the checks are created, the visuals are ready to be created.

4.2.4. Visualization

The visualizations were split into four sheets, each with its own purpose. The first sheet, shown in Figure 18, is created for the model name alignment between the different folders. The left three columns compare the shared, collaboration and transfer folders with the Work in Progress folder, while the right two compare the transfer and federated model together.
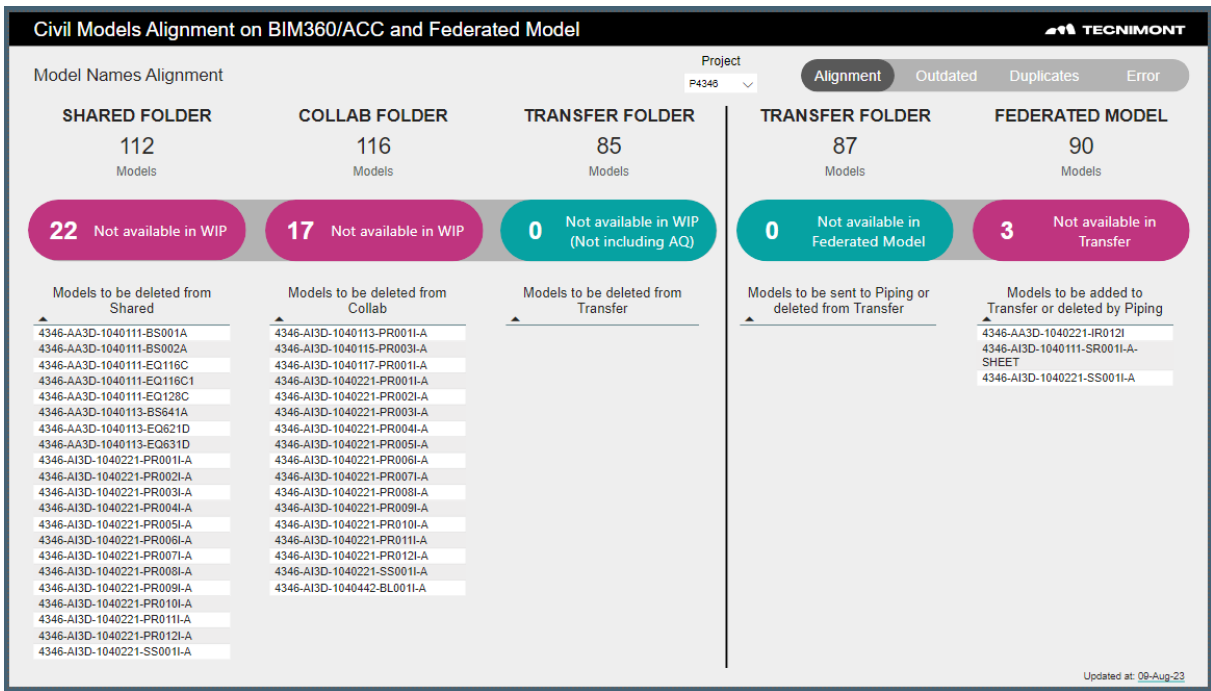
Figure 18 – Quality 1 Page 1

Figure 19 shows the second page of the visual, focusing on the update date of the transfer models in comparison with the WIP models.
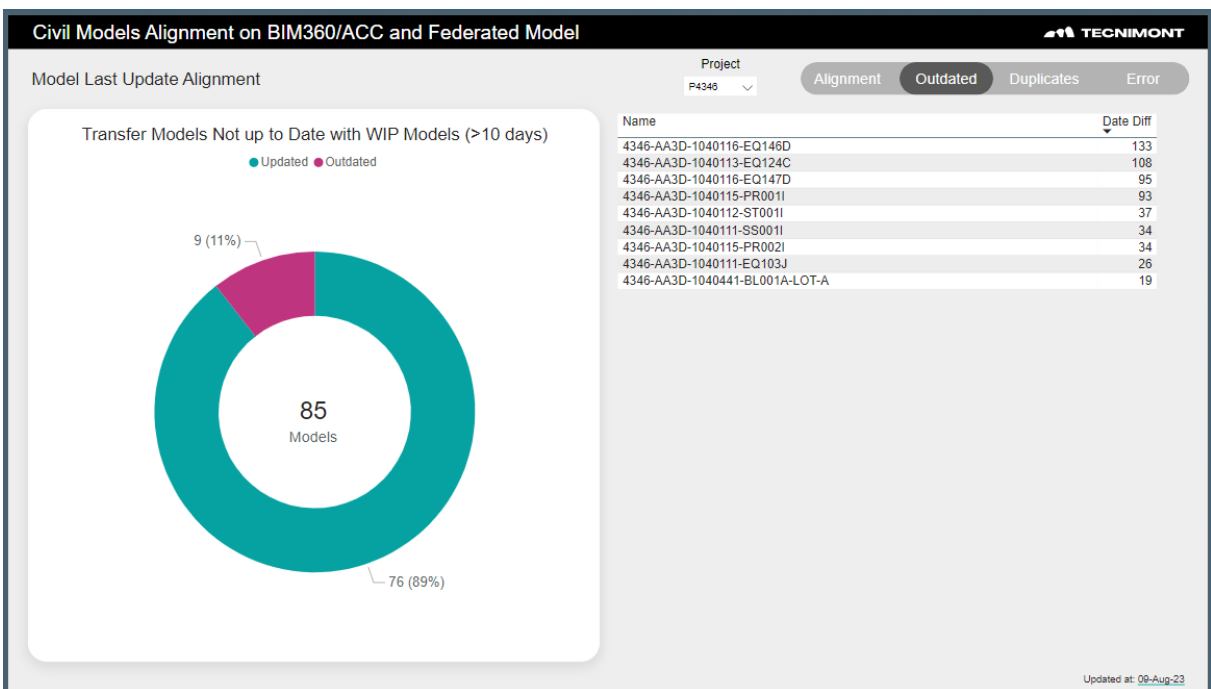

Figure 19 – Quality 1 Page 2

The third page of the dashboard focuses on the duplicates in the WIP, Transfer and Federated folders. It highlights which folders have duplicate files since those become errors and need to be removed. Figure 20 below shows this:
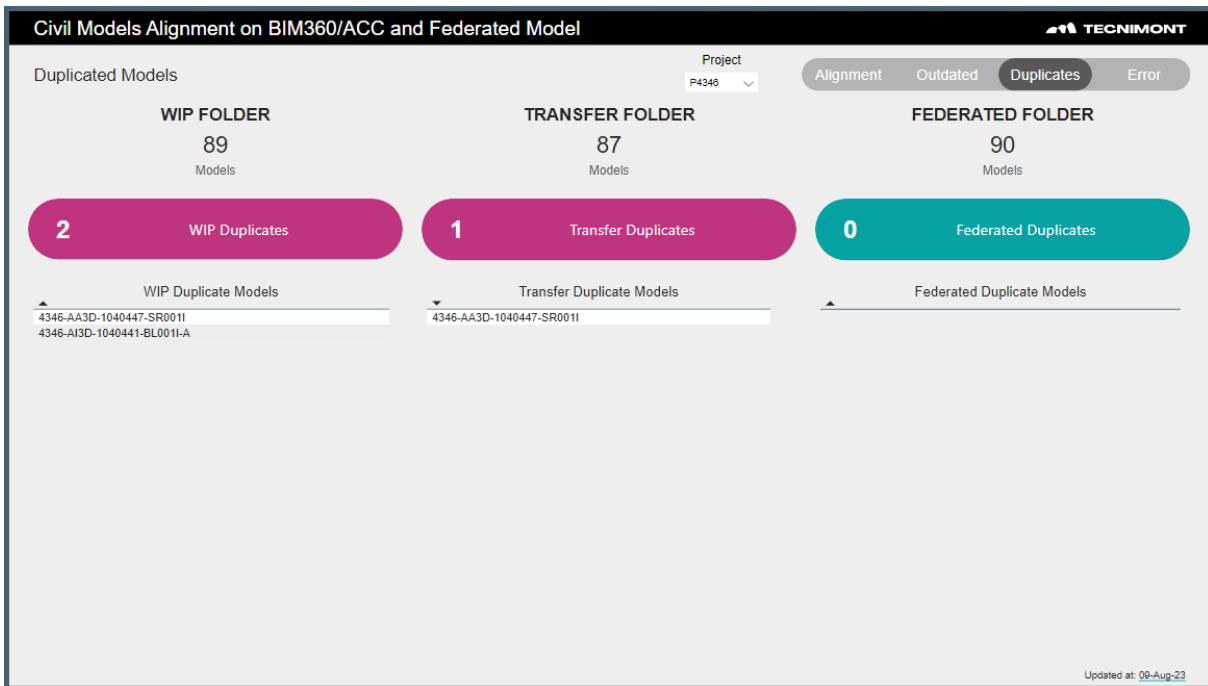
Figure 20 – Quality 1 Page 3

The final page shows all the error files in the folder, split by disciplines and steel errors. Figure 21 shows this database which allows a user to look through their project and pinpoint the errors within.
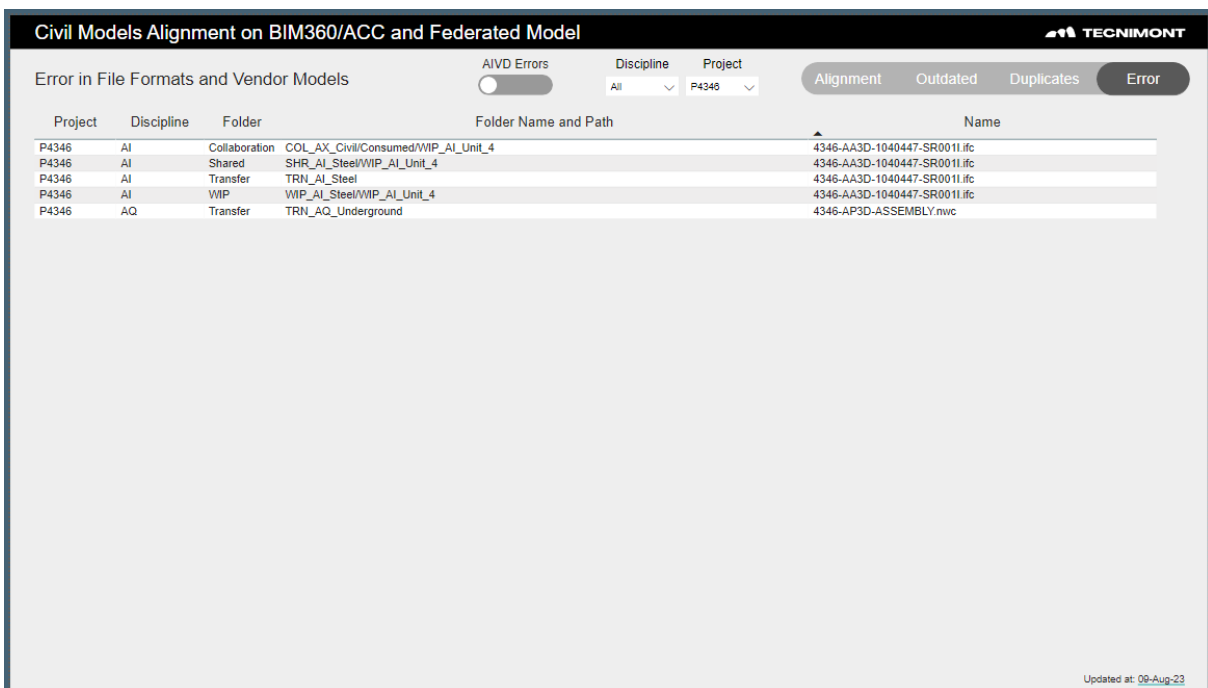
Figure 21 – Quality 1 Page 4

4.2.5. Automation and Testing

During the testing phase of this dashboard, the users gave some insight to the visuals that helped them visualize better. For example, the line to separate the WIP checks and Federated checks in Figure 18 was an idea by the users. It was helpful to understand which visual tricks help the users better focus on their different tasks and checks.

Automation in this dashboard can be classified as semi-automatic with new exports from the federated model and ACC / BIM 360 weekly and an automatic refresh on the Power BI Service web platform every week. It was then published and shared with the modelling teams across the company to ensure the model locations were corrected.

The resulting dashboard is now in use by BIM coordinators to ensure model alignment, this saved more than one hour per week for each BIM coordinator. Before this dashboard, the BIM team would export the file names and upload it to an excel sheet formatted to check some of the locations but would need to first manually filter out the models that are errors or duplicates which is time consuming. It would also require the team to do this for each project instead of all of them at once.

4.3. Quality 2: Models Attributes Check

The second quality dashboard created checked the attributes of each model element in all projects. It truly was a data heavy dashboard given the sheet number of modelled elements within not only one project, but eight active projects. The number of elements checked in this dashboard near 500,000 elements and keep growing. This dashboard is very useful since it allows for sharing with all modelers in the project and gives the errors in element attributes according to the companies' standards. Although they do not necessarily follow any international standards, the rules in this dashboard are comparable to the rules of tagging created by Uniformat. These rules are important for the Client as attributes required for Advanced Work Package (AWP) information for the signing off of the construction drawings which in turn will be used during the construction of the project.

4.3.1. Exploration and Mapping

The data for this dashboard may be the most basic data of the dashboards since it only has one source, but the size of it make it more complex than it seems from first glance. The data was exported from the Federated Models of each project, split by discipline to lessen the size of the file being opened. To understand the mapping, Table 6 shows the exported properties from the model elements:

Table 6 – Quality 2 Exported Properties

Category	Property	Description
Item	Source File	Model
	Name	Element Name
ITEMTAG	CWA	Construction work area
	Main Item Tag	Maire naming convention for different structures
	Client Tag	Tag requested by client
	Item Tag	Element tag based on tagging convention
	Mark	Vendor tag of element
	Lot	Sub-division of item if there is any
ENGINEERING DATA	Engineering Status	Status of design level
	Contractor Identification Code	Document number for contractors' drawings
	Employer Identification Code	Document number for employers' drawings
	Document Revision	Number of revisions for element
	Hold Status (Y/N)	Whether the element has a hold required to be resolved by other disciplines or not
	Hold Type	Type of the hold if status is "Yes"
	CMD Checker	Element name repeated if element is tagged correctly
MATERIAL	Material	Material of element for Concrete and Steel
CONCRETE DATA	Volume	Volume of Concrete Elements
STEEL DATA	Weight	Weight of Steel Elements
ITEMTAG	LV02_Material Work Group	Material Work Group of Underground Elements

With the properties now available, the author needed to create the rule checks based on the different errors possible in accordance with the properties above. The checks can range from naming conventions to missing properties in the element. Table 7 details every check created in this dashboard and the error message a user will receive if an element has an error.

Table 7 – Quality 2 Checked Errors

Checked Column	Error Messages	Checks
CMD Checker	EXPORT ERROR	If it's present or not, meaning the element was exported incorrectly
CWA	CWA MISSING	If it's missing
	CWA AND MAIN ITEM TAG UNALIGNED	If Main Item Tag contains the CWA
	CWA LENGTH ERROR	Project specified CWA characters length is wrong
Main Item Tag	MAIN ITEM TAG MISSING	If it's missing
	MAIN ITEM TAG AND ITEM TAG UNALINGED	If Item Tag contains the Main Item Tag
	ERROR IN MAIN ITEM TAG	If it contains " " or "." Or if the Main Item Tag length is shorter than the CWA + 5
Model Name	MAIN ITEM TAG AND MODEL NAME UNALIGNED	If Model contains the Main Item Tag or the Model allows for specific Main Item Tags (exception AQ)
Client Tag	CLIENT TAG MISSING	Project specific, client requires a specified tag
	CLIENT TAG NOT REQUIRED BUT PRESENT	Client tag is present but was not required, if dummy needs to be a "-"
Item Tag	ITEM TAG MISSING	If it's missing
	ERROR IN ITEM TAG	If it contains " " or "."
	MARK BY TEKLA MISSING	Tekla maps Mark in Item Tag column, so for AI missing
Mark	MARK MISSING	If model is in AIVD or AAVD stage (vendor model) need a mark
Lot	LOT MISSING	If it's missing
Eng Status	ENG STATUS MISSING	If it's missing
	ENG STATUS CODE NOT STANDARD	If it is not standard or part of a project specific standard
Contractor Identification Code	CONTRACTOR DOC MISSING	If Eng Status is IFC or IFR and the contractor code is missing
	CONTRACTOR CODE WRONG	If Eng Status is IFC or IFR and the contractor code starts with "0000" (dummy code)
Employer Identification Code	EMPLOYER CODE MISSING	If Eng Status is IFC or IFR and the project specific employer code is required, yet it is missing
	EMPLOYER CODE WRONG	If Eng Status is IFC or IFR and the project specific employer code is required, yet it starts with "0000" (dummy code)

Checked Column	Error Messages	Checks
	EMPLOYER CODE NOT REQUIRED BUT PRESENT	If Eng Status is IFC or IFR and the project specific employer code is not required, yet it is there
Doc Revision	DOC REVISION MISSING	If Eng Status is IFC or IFR and the doc revision is missing
Material	MATERIAL MISSING	If material for AA or AI is there or not
	MATERIAL NOT STANDARD	For Concrete only, make sure it's a standard material
Name	FAMILY CODE NOT STANDARD	For Concrete and Underground, Family Code of Revit element needs to be from a standard list
	STEEL NAME NOT STANDARD	For steel, check if name is from standard list
Hold Type	HOLD TYPE MISSING	If Hold is "Yes" and Hold Type is missing
	HOLD TYPE NOT STANDARD	If Hold is "Yes" there is a standard list of Holt Types available
Volume	AA VOLUME MISSING	For concrete, if it's missing
Weight	AI WEIGHT MISSING	For steel, if it's missing
LV02_Material Work Group	AQ MATERIAL WORK GROUP MISSING	For underground, if it's missing
CMD Checker	EXPORT ERROR	If the export does not follow the template created by the company from the CMD

4.3.2. Extraction

The research team created a Navisworks template to export the exact columns to be able to be consistent when importing into Power BI. The users can then open updated federated models and export the model attributes and properties into a .csv and place that file in the Power BI folder as explained in the beginning of this chapter.

The other source of information is an excel workbook with multiple sheets that includes the rules that can be cross referenced for different projects. For instance, each project requires a different length of the CWA which can be referenced from this "rules sheet" and then updated per project for this dashboard.

4.3.3. Formatting

Formatting includes multiple steps as explained in Chapter 3.4, for this chapter, the author will define the queries within the Power BI Power Query Editor. Then the author will show the data linking between the sources. Finally, any important steps or equations that are specific to this dashboard will be highlighted.

Table 8 shows the queries and their sources:

Table 8 – Quality 2 Queries

Group	Query	Source	Objective
Sources	*Elements Folder*	-	Allow for seamless change in folder location which allows for PBI template to be shared with required parameters to be changed.
	Rules File	-	Allow for seamless change in file location which allows for PBI template to be shared with required parameters to be changed.
Elements	Elements	*Elements Folder*	Navisworks export with all the rules that do not need checking standards or lists
	Source	Elements	Used to get unique list of source, discipline, and project to be connected in the Power BI Model View
Rules	CWA Check	*Rules File*	CWA length check by Project
	Main Item Tags Allowed	*Rules File*	Main Item Tags allowed in Models that do not share the name
	Client Tag Required	*Rules File*	Client Tag requirements by Project
	Allowed Eng Status Code	*Rules File*	Allowed Eng Status Codes by Project
	Employer Code Required	*Rules File*	Employer Code requirements by Project
	Materials Allowed	*Rules File*	List of standard Materials by Project
	Family Codes Allowed	*Rules File*	List of standard Family Codes across company
	Steel Names Allowed	*Rules File*	List of standard Steel Names across company
	Hold Type Allowed	*Rules File*	List of standard Hold Types across company

*Italicized queries are not loaded into the Power BI Model

Once the query editor work is finished it is important to connect the rules to the columns they will be checking. Figure 22 shows how each rule is connected and linked to the respective column:
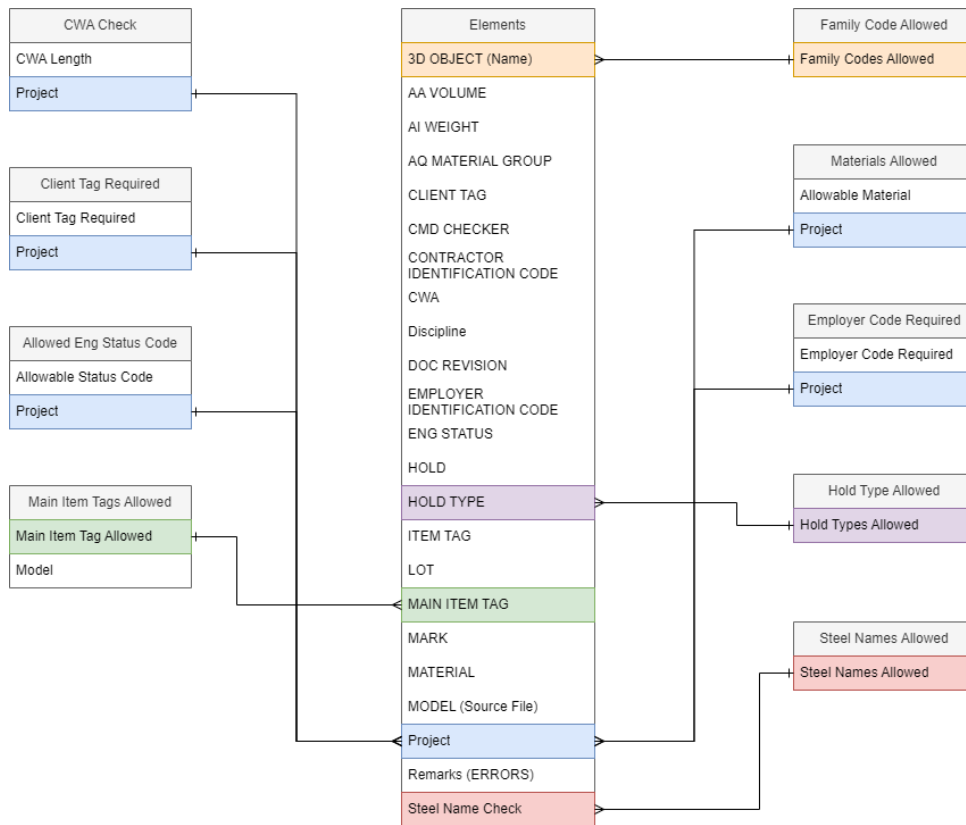
Figure 22 - Quality 2 Data Connections

With the connections in place, it is now easy to check whether each column follows the rules stated in Table 7. The checks that do not require references were done in the query editor while the rest were written in DAX within the Power BI Desktop Model view. For these referenced checks, they can be summarized with the following code and repeated for the rest of the checks:

```
Check Employer Doc =
IF(
    Elements[Errors Employer Doc] = "Check Employer Doc Required,",
    IF(
        RELATED('Employer Code Required'[Employer Code Required]) = "Yes",
        IF(
            Elements[EMPLOYER IDENTIFICATION CODE] = BLANK() || Elements[EMPLOYER
            IDENTIFICATION CODE] = "",
            "EMPLOYER DOC MISSING,",
            IF(
                LEFT(Elements[EMPLOYER IDENTIFICATION CODE],4) = "0000",
                "EMPLOYER DOC WRONG, ",
                BLANK())),
        IF(
            RELATED('Employer Code Required'[Employer Code Required]) = "Yes",
            IF(
                Elements[EMPLOYER IDENTIFICATION CODE] = "-",
                BLANK(),
                "EMPLOYER DOC NOT REQUIRED BUT PRESENT, "),
BLANK())))
```

The next step was to combine the remarks for the errors as shown below, making sure that if there is an export error to not show any other errors:

```
Remarks =
IF(
    [Query Remarks] = "EXPORT ERROR",
    "EXPORT ERROR",
    ([Query Remarks] & " " & Elements[Check CWA Length] & Elements[Check Model
    Name] & Elements[Check Client Tag] & Elements[Check Eng Status] &
    Elements[Check Employer Doc] & Elements[Check Material] & Elements[Check
    Family Code] & Elements[Check Steel Name] & Elements[Check Hold Type]))
```

Finally, a check to count errors and number was created in this query. After this a list of all the models that need to be updated was taken into a unique list. This list was also responsible for creating a new table with total number of each category as shown in Figure 23:

Number of Models	Number of Incorrect Models	Number of Correct Models	Source.Name	Project	Discipline
7	5	2	4245_AQ_EXPORT_230707.xlsx	4245	AQ
342	146	196	4274_AA_EXPORT_230717_checker.xlsx	4274	AA
2	2	0	4274_AQ_EXPORT_230707.xlsx	4274	AQ
380	62	318	4300P_AA_EXPORT_230717_checker.xlsx	4300P	AA
1	1	0	4300P_PO_AQ_EXPORT_230710.xlsx	4300P	AQ
491	175	316	4300U_AA_EXPORT_230717_checker.xlsx	4300U	AA
1	1	0	4300U_UO_AQ_EXPORT_230710.xlsx	4300U	AQ
41	23	18	4300X_AA_EXPORT_230717_checker.xlsx	4300X	AA
1	1	0	4300X_XLPE_AQ_EXPORT_230710.xlsx	4300X	AQ
152	40	112	4325_AA_EXPORT_230710_checker.xlsx	4325	AA
4	3	1	4325_AQ_EXPORT_230707.xlsx	4325	AQ
48	12	36	4346_AA_EXPORT_230710_checker.xlsx	4346	AA
159	35	124	4355_AA_EXPORT_230717_checker.xlsx	4355	AA
1	1	0	4355_AQ_EXPORT_230707.xlsx	4355	AQ

Figure 23 - Quality 2 Total Model Numbers per Project

4.3.4. Visualizations

This dashboard included four pages: one for overall model and element status, one for all errors by project, another for AWP errors by project, and the final one as a database that shows all errors and element names. The difference between the one that shows errors by project and the one that shows AWP errors by project is that AWP errors focuses on errors that contain "CWA," "ITEM TAG," "MARK," "VOLUME," or "WEIGHT."

Figure 24 shows the first page, which can be used to understand overall quality of models and errors based on project and discipline in the company. Middle management can use this sheet to get an understanding of how far the company is going and which projects need more focus.
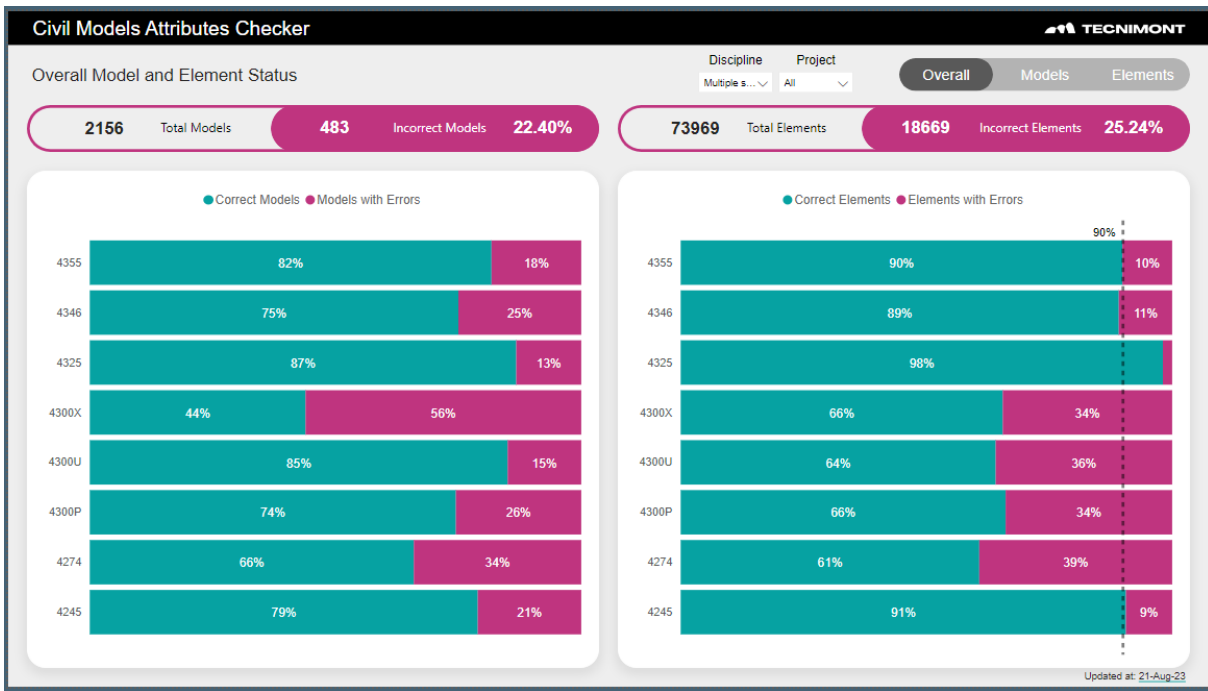
Figure 24 – Quality 2 Page 1

The next page focuses on specific errors by models within a project. Once filtered for a specific project, a project leader can understand which models are underperforming and what are the top five errors within the project. Corrective action is easy and accessible with the ability to see the number of incorrect elements vs the total number of elements and which errors are combined in which models. Figure 25 shows this page:
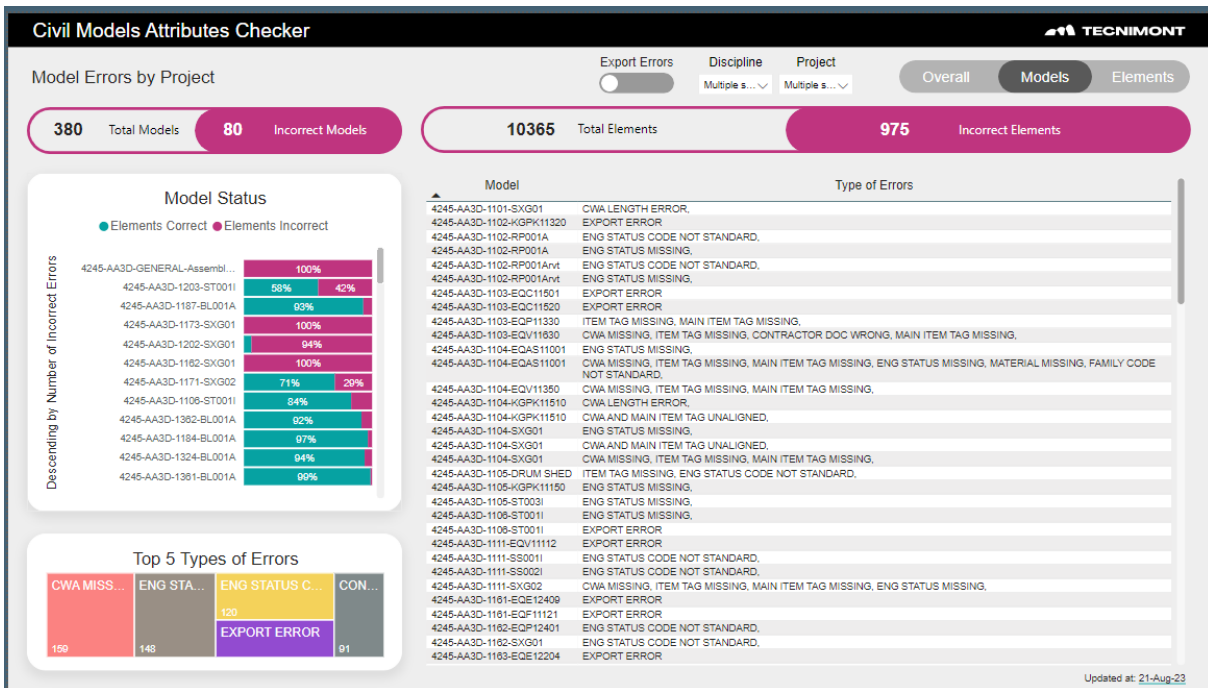

Figure 25 – Quality 2 Page 2

The next page follows the same format but focuses only on the AWP errors. AWP errors are more important because they will hinder construction progress, in comparison to some errors which can be rectified after the construction takes place. Figure 26 below shows this:
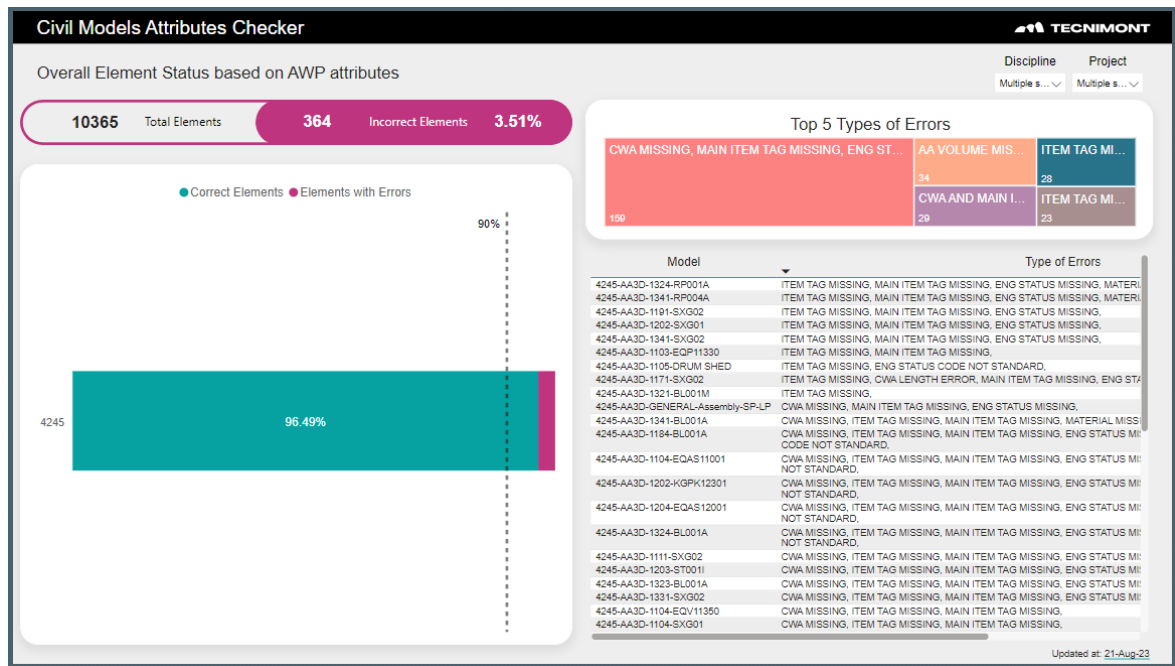
Figure 26 – Quality 2 Page 3

Finally, the complete database shown in Figure 27 helps the modeler fix the errors by searching for the elements or their model and finding the errors by element highlighted. This helps in finding the element that needs correction to be fixed quickly and efficiently.
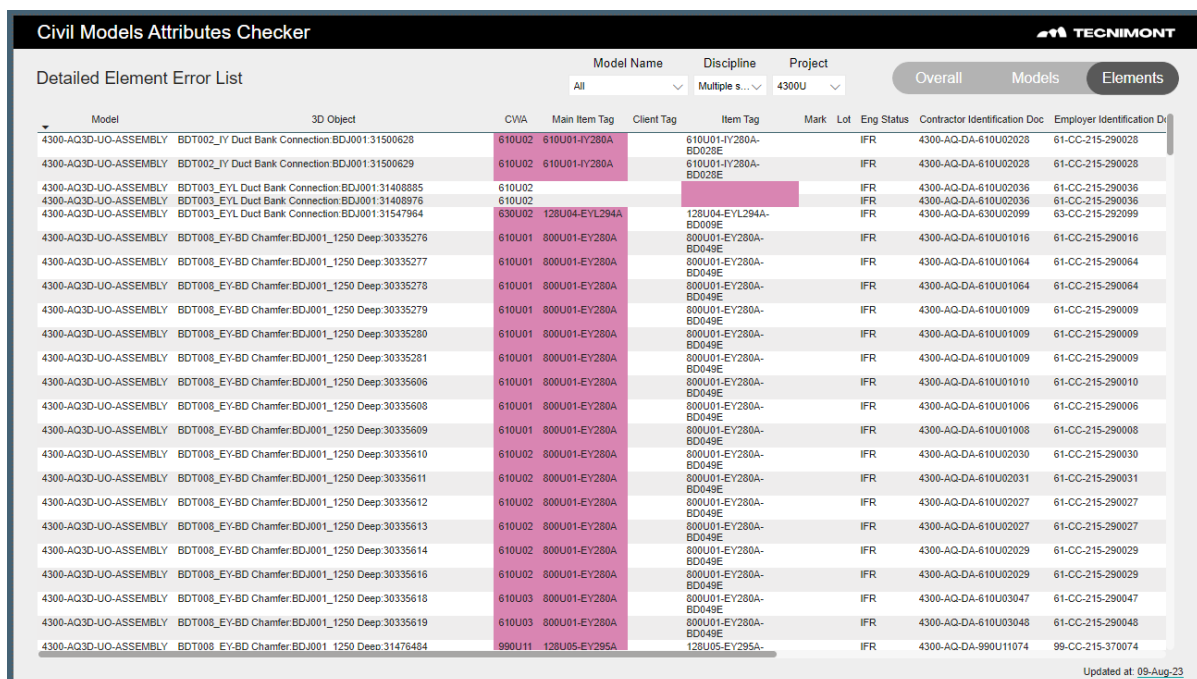
Figure 27 – Quality 2 Page 4

4.3.5. Automation and Testing

Testing took more time in this dashboard given the requirement to be accurate so all users can utilize it effectively. The dashboard had three main users: the modelers, BIM coordinator and management; therefore, it was crucial to get feedback from each group and ensure that it can be understood by all three. Due to some complexity and the modelers' familiarity with excel workbooks, an excel workbook was created to recreate the final page by copying the data into a formatted sheet and sharing the excel file with them. This goes to prove that sometimes a team is not ready for advancement in technology and a combination of automation and old methods might be more useful until they are prepared to try new methods, in a new project for example.

Automation in this dashboard can be classified as semi-automatic with exports coming from the federated model into the shared folder. The dashboard then had the ability to be refreshed periodically depending on what the client required for this dashboard.

This quality check used to be a complex excel sheet that required a download of all the elements and copying and pasting the data required for check into this excel sheet. Now this dashboard has created a workflow that does not require copying and pasting or doing this multiple times for multiple projects. It is used by the BIM coordinators and the lead BIM managers to check the big picture problems in the models. The BIM modelers and lead modelers use it to fix some of the elements, but due to their familiarity with excel and the old standard, the author created a formatted excel sheet that mimics Figure 27 to be shared with the modelling team.

4.4. Quality 3: Tags Data Alignment Check

This dashboard checks the quality between two sources of data for Main Item Tags. Main Item Tags are a collection of elements that share a code, this is relevant to an assembly of columns or assembly of pipes, they all get saved under a singular Main Item Tag with the same unit of measure and combined quantity. The first source shows the items that are exported from the models into AVEVA engineering and the second is those same items filled in by discipline leaders in the companies Civil Master Data. The purpose of this is to ensure that all quantities are accounted for and all elements that are modelled are available for the construction and estimate groups as well as the Civil department.

4.4.1. Exploration and Mapping

Data for this dashboard comes from two sources, so that was the first thing to check is the data sources that are exported. From AVEVA there are multiple folders for each discipline which exports excel workbooks that have multiple columns for each model element and its quantities and attributes. From the Civil Master Data (CMD) and the Civil Item List (CIL) it is possible to export excel workbooks that have the same data but for entire Main Item Tags instead of for each element individually. Right away it is clear that the AVEVA file needs to combine elements by main item tag so that they can be compared with the elements from the item lists which in turn need to be merged into one list.

Again, the CMD is the company's method of having all data centralized on a company hosted website. The Civil Item List (both within the CMD and the older CIL) includes metadata of each element in a project inputted manually to match other data.

Now that the data is known, some the checks can be mapped as shown in Figure 28.
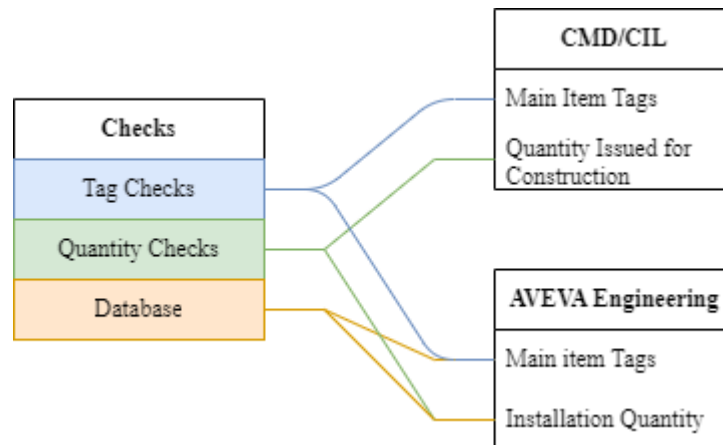

Figure 28 - Quality 3 Mapping

The checks can be further explained below:

- Tags
    - Loaded – the Main Item tag from the CMD / CIL is in the IFC (issued for construction) phase and is available in the AVEVA Engineering database.
    - Pending – the Main Item Tag from the CMD / CIL is not available in the AVEVA Engineering database but has been uploaded to the CMD / CIL and is in the IFC phase for less than 20 days from today.
    - Delayed – the Main Item Tag from the CMD / CIL is not available in the AVEVA Engineering database and has been uploaded to the CMD / CIL and is in the IFC phase for more than 20 days from today.
- Quantity
    - Matching
        - Compares the quantities loaded from AVEVA (Installation Quantity) and the CMD / CIL (Quantity Issued for Construction) that have matching Main Item Tags.
    - All
        - Compares all available quantities from the AVEVA (Installation Quantity) and the CMD / CIL (Quantity Issued for Construction) no matter if they are in the other or not.
- Database
    - All elements in AVEVA in case a user would like to search easily for a Main Item Tag.

4.4.2. Extraction

Extraction of data in this dashboard was not as simple as the ones before given the data location is stored in two sources that require their own methodology. For the CMD / CIL, further introduced as "Engineering," the data is manually inputted by discipline leaders to the web-based company database, so to get the data an authorized user must navigate the folder and export the entire Civil Item List. A goal of the project that did not come to fruition was to convert all CIL project (old CMD) into the CMD

but due to time constraints and some limitation of the imports it was unachievable. This causes some additional formatting to merge both sources together.

For the AVEVA export, the data originates from the modelling source such as Revit or Tekla. Once exported as an ifc file the tagging is the done using an internal Application Program Interface (API) to complete the properties required into an excel sheet that matches with the AVEVA Engineering upload requirements. Once the excel files for each model are uploaded to AVEVA, the approved users can export the entire projects as another excel workbook that will be used for this dashboard. Obviously, the repetition of these steps is laborious and need multiple users to gain access to different platforms which becomes a bottleneck in the updating of data.

4.4.3. Formatting

Formatting includes multiple steps as explained in Chapter 3.4, for this chapter, the author will define the queries within the Power BI Power Query Editor. Then the author will show the data linking between the sources. Finally, any important steps or equations that are specific to this dashboard will be highlighted.

The queries in this chapter are much longer and complex for two reasons. First the merging of the CMD and CIL folders into an "Engineering" table requires some formatting to align columns and at the same time there is a company list and a vendor list that also need to be merged since most steel attributes are taken from a vendor company. The second problem is that the AVEVA files are discipline based so they need to be done individually and fixed individually before combining the queries. This second problem also requires the merged Engineering table to be split by discipline. All this can all be seen in Table 9 below:

Table 9 - Quality 3 Queries

Group	Query	Source	Objective
Sources	*CMD Folder*	-	Allow for seamless change in folder location which allows for PBI template to be shared with required parameters to be changed.
	CMD Vendor Folder	-	
	CIL Folder	-	
	CIL Vendor Folder		
	Concrete Folder	-	
	Paving Folder	-	
	Piles Folder	-	
	Steel Folder	-	
	Underground Folder		
Engineering	*CMD Engineering*	*CMD Folder*	Clean original CMD / CIL exported from the web and filter for IFC elements only.
	CIL Engineering	*CIL Folder*	
Fabrication	*CMD Vendor*	*CMD Vendor Folder*	Steel has a different requirement for IFC elements based on Eng Status from steel vendors, so different file for steel elements
	CIL Vendor	*CIL Vendor Folder*	

Group	Query	Source	Objective
Concrete	AVEVA Elevation Concrete	*Concrete Folder*	Clean and filter data for Elevation concrete elements.
	AVEVA Foundation Concrete	*Concrete Folder*	Clean and filter data for Foundation concrete elements.
	Engineering Elevation Concrete	*CMD & CIL Engineering*	Filter for Elevation Concrete and remove duplicates of Main Item Tags.
	Engineering Foundation Concrete	*CMD & CIL Engineering*	Filter for Foundation Concrete and remove duplicates of Main Item Tags.
Piles	AVEVA Piles	*Piles Folder*	Clean and filter data for Piles elements.
	Engineering Piles	*CMD & CIL Engineering*	Filter for Piles and remove duplicates of Main Item Tags.
Steel	AVEVA Steel	*Steel Folder*	Clean and filter data for Steel elements.
	Fabrication Vendor Steel	*CMD & CIL Fabrication*	Remove duplicates of Main Item Tags.
Underground	AVEVA Underground	*Underground Folder*	Clean and filter data for Underground elements.
	Engineering Underground	*CMD & CIL Engineering*	Filter for Underground and remove duplicates of Main Item Tags.
Paving	AVEVA Paving	*Paving Folder*	Clean and filter data for Paving.
	Engineering Paving	*CMD & CIL Engineering*	Filter for Paving and remove duplicates of Main Item Tags.
Databases	*AVEVA Elevation Concrete DB*	*Concrete Folder*	Used for database combination.
	AVEVA Foundation Concrete DB	*Concrete Folder*	
	AVEVA Piles DB	*Piles Folder*	
	AVEVA Steel DB	*Steel Folder*	
	AVEVA Underground DB	*Underground Folder*	
	AVEVA Paving DB	*Paving Folder*	
	Overall Database	*All DB Queries*	Appended all DB's to be able to show database by discipline.
Calculations	Projects	*CMD*	List of unique projects.
	Tag Type	*CMD*	List of unique Tag Types and their disciplines.
	AVEVA Append	All AVEVA Queries	Used for total quantities unrelated to Engineering or Fabrication
	Engineering / Fabrication Append	*CMD and CMD Vendor*	Used for total quantities unrelated to AVEVA

Italicized queries are not loaded into the Power BI Model

Now that the query is finished and formatted to match all tables, the connection of tables is the next step. Again, due to the many queries, each sub-discipline required meticulous focus to connect the correct columns to each other and create a database that encapsulates all elements and matches the tagging from the AVEVA to the Engineering with all disciplines. Figure 29, Figure 30, and Figure 31 show the connections for concrete, steel and underground respectively.
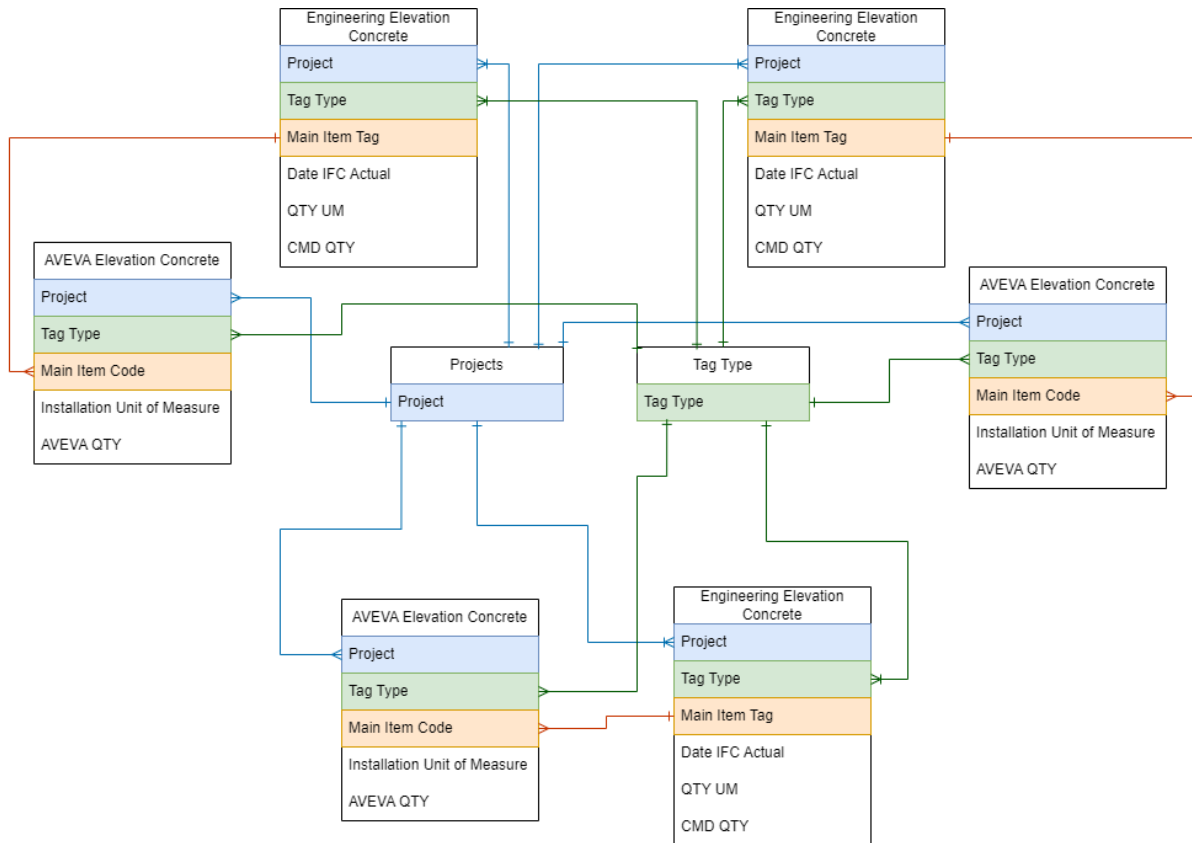
Figure 29 – Quality 3 Concrete Data Connections
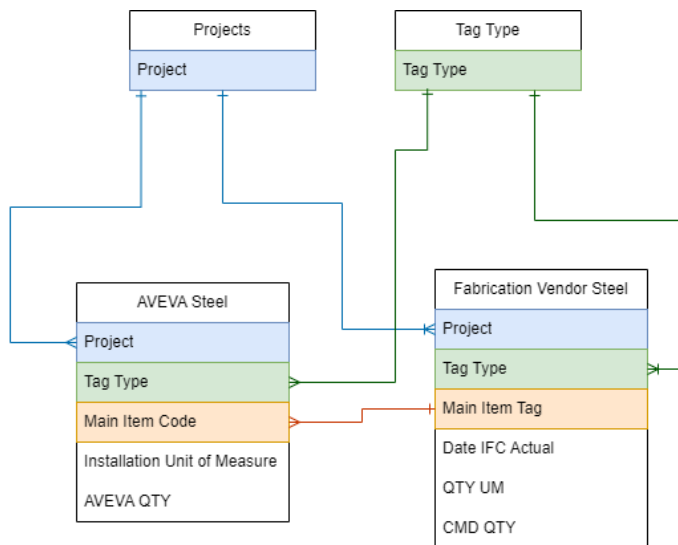
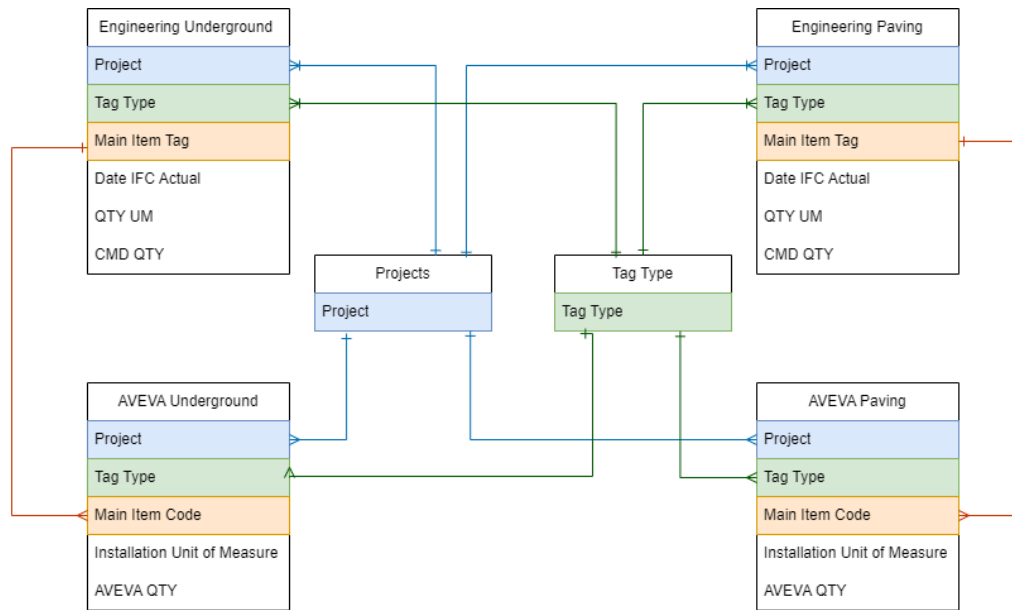**Figure 30 – Quality 3 Steel Data Connections**

Figure 31 – Quality 3 Underground Data Connections

Now that the AVEVA and Engineering are connected by the Main Item Tag it was imperative to the several checks described in sub-chapter 4.4.1. This was created for each Engineering table since all of them had the same column names and formatting coming out of the power query. The first step was to find the status of the Main Item Tag in AVEVA by using the equation below:

```
STATUS =
IF(
    NOT ISBLANK('CMD Piles'[MAIN ITEM TAG]),
    IF(
        'CMD Piles'[MAIN ITEM TAG] IN SELECTCOLUMNS (RELATEDTABLE ('AVEVA Piles'),
        'AVEVA Piles'[Main Item Code]),
        "Loaded",
        IF(
            'CMD Piles'[Days Late] > 20,
            "Delayed",
            "Pending")))
```

Then to find the days late, the following equation was utilized:

```
DAYS LATE = DATEDIFF('CMD Piles'[DATE IFC ACTUAL], TODAY(), DAY)
```

The next step was to find the difference in quantity between the AVEVA and CMD quantities that match, the equation below shows the usage:

```
QTY DELTA = CALCULATE(AVERAGE('CMD Piles'[QTY CMD]) - ROUND(SUM('AVEVA Piles'[AVEVA
QTY]),0), ALLSELECTED('AVEVA Piles'[Main Item Code]))
```

The average of the 'CMD Piles' [QTY CMD] was used because for each row there was only one number to be used, and in order to use the calculate function, a function needed to be used. In comparison, the

AVEVA had multiple rows for one Main Item Tag therefore the sum was required. The three equations above were then used for each tag type.

For the final analysis of the CMD, the CMD append needed to find totals for each project and tag type. This allows for a complete picture of the tags loaded for entire projects by tag types. Repeating this for all tables was done and at this point the visualizations were ready to be executed.

4.4.4. Visualizations

Given the different checks and disciplines, this visualization had multiple pages and sub-pages as seen in the navigation throughout the following figures. The first "chapter" includes a summary of all sub-disciplines as shown by the overview tab in Figure 32.
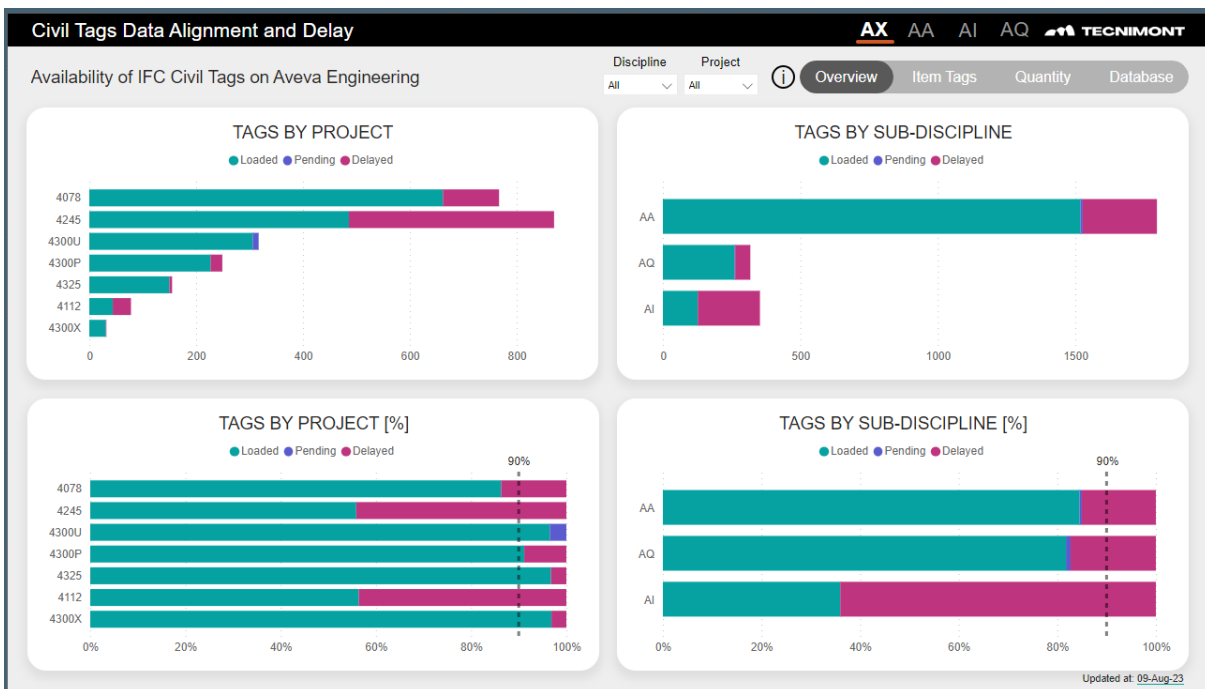
Figure 32 – Quality 3 Page 1

Moving to the next page, the item tags which are in the IFC phase can be seen whether they are loaded, pending or delayed as shown in Figure 33.
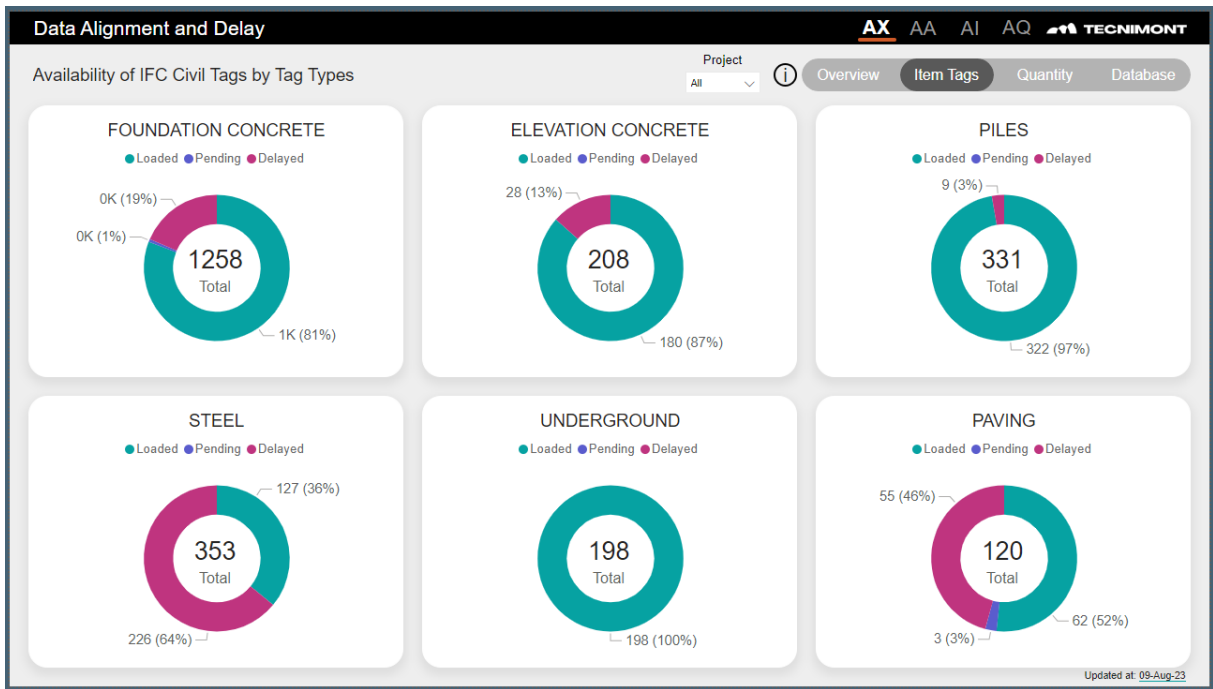
Figure 33 – Quality 3 Page 2

Now to make sure the quantities are matching up for the elements that are loaded, the third page shows the comparison of quantities shown in Figure 34.
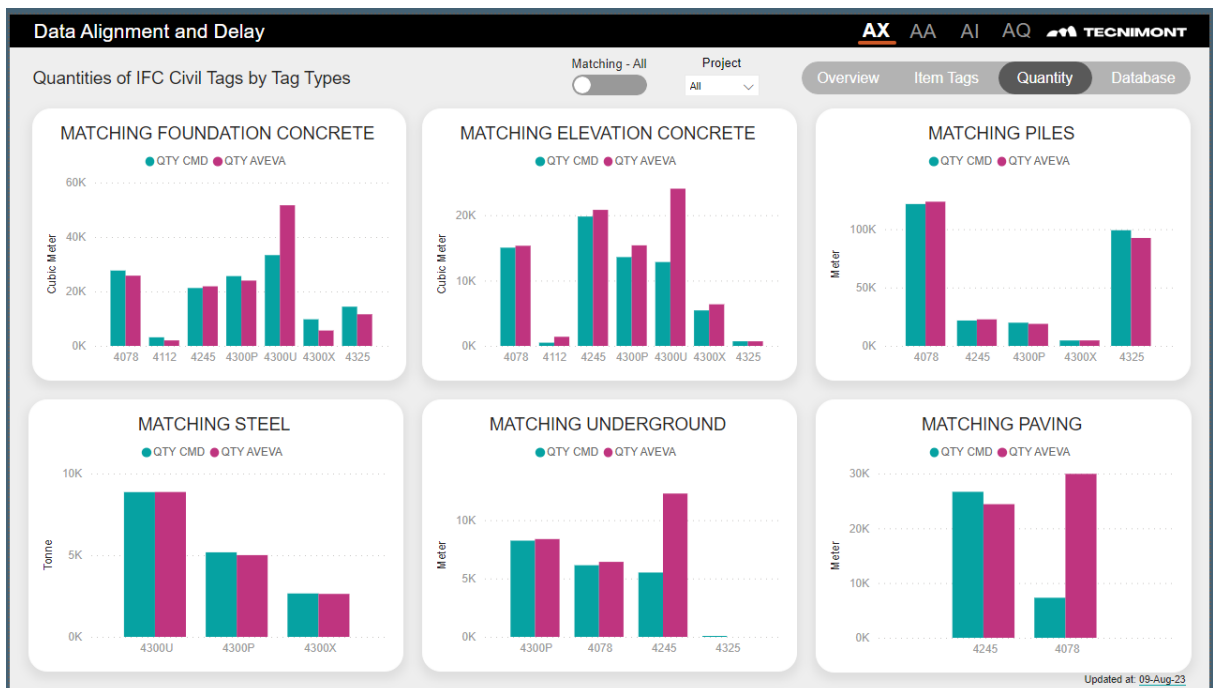
Figure 34 - Quality 3 Page 3

Finally, a database of all loaded elements in AVEVA will be required in case someone needs to check the data manually. Figure 35 shows this:

Figure 35 - Quality 3 Page 4

Once these were finished, discipline-based pages were created. These pages had the delay and quantity difference by Main Item Tag to highlight the incorrect Main Items. Concrete pages can be seen in Figure 36, steel pages in Figure 37 , and underground pages in Figure 38. The database pages were omitted to give space for the visuals.
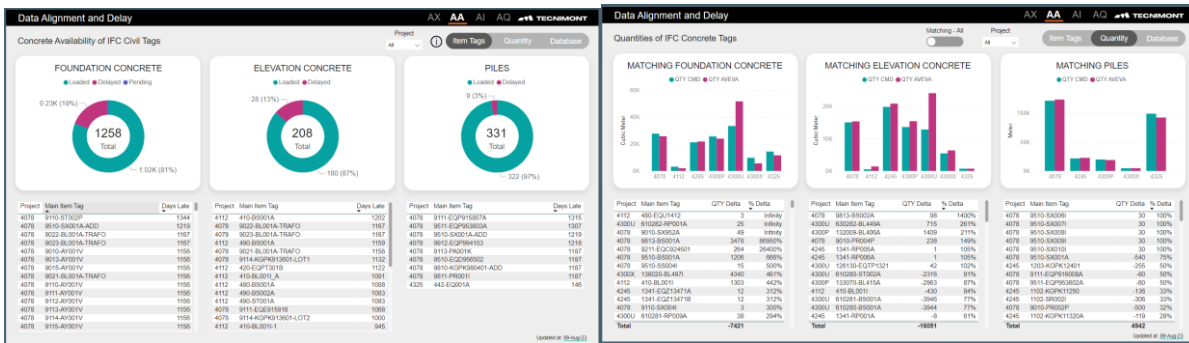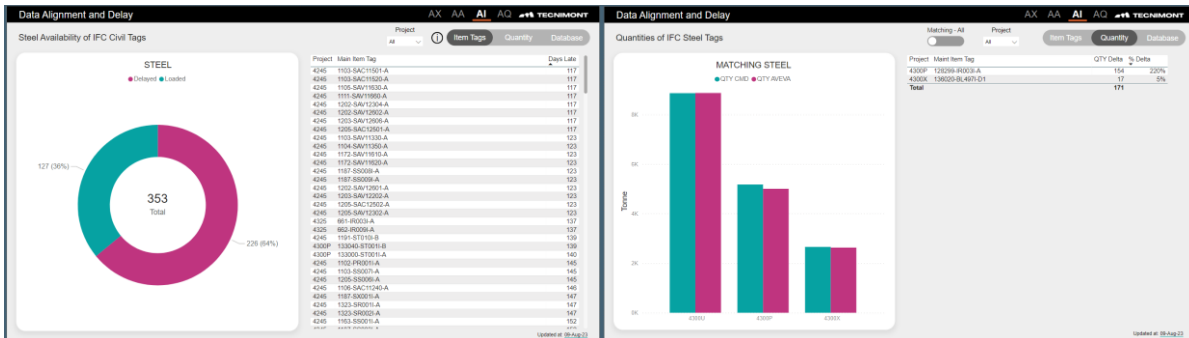

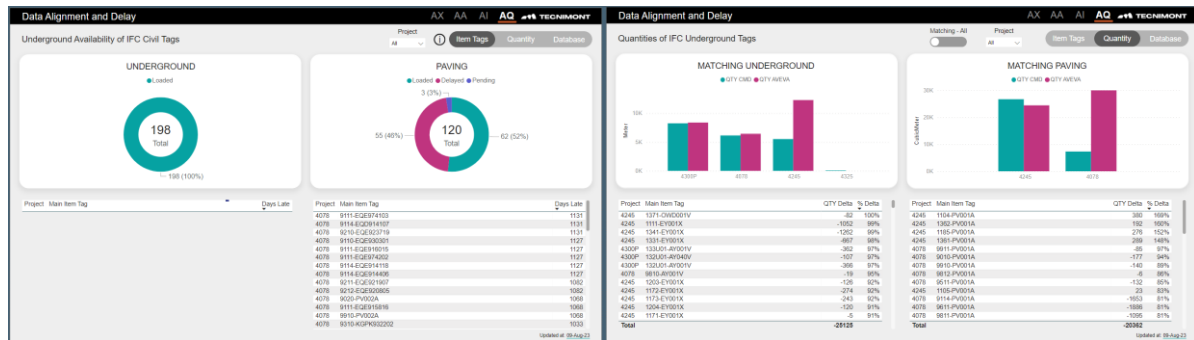**Figure 36 – Quality 3 Pages 5-6**

Figure 37 – Quality 3 Pages 8-9

Figure 38 – Quality 3 Pages 11-12

4.4.5. Automation and Testing

The testing of this dashboard brought out many problems at first. This was due to the incorrect method in linking the tables as explained earlier in Figure 29, Figure 30, and Figure 31. Testing was extensive and very useful in creating a more comprehensive database. With AVEVA and the Engineering tables connected, the rest of the testing was straightforward given the dashboard was already useful before the completion of the visualizations.

This dashboard was the first to be visualized with the company colors and the template, so it was a tedious effort to create something compatible with all dashboards and presentable to the users. This also required multiple iterations of visualizations that took more time than expected, but resulted in a product that both the author and the users were proud of.

Automation for this dashboard was unfortunately manual since the extraction process must be manually done and imported into the dashboard. The automatic refresh can be implemented, but whenever a new excel workbook is added from AVEVA, the sheet name needs to be manually updated to match the rest of the workbooks which requires a manual step.

This dashboard allows some saving in time worked by the BIM coordinators because it allowed for all projects and disciplines to be viewed simultaneously. The process preceding this effort required each discipline BIM coordinator to work individually which was not as time effective as now. Due to the manual automation of this dashboard, there is still some improvement that can be made to make it even more efficient.

4.5. Quality 4: Squad Check Reviews

During the engineering phase of a project, the modeler's sometimes need the discipline leaders to review their work to make sure they can carry forward with the modelling of the project. So, the modelers open what is called a "Squad Check" on ACC / BIM 360 as a review task and assigns it to different disciplines as required. The Squad Check then needs to be reviewed by the required disciplines and then closed by the Civil BIM department. Before this work, when there were delays in the disciplines, the data of reviews was downloaded and uploaded to a comprehensive excel sheet that was sent to the required disciplines for each project. The purpose of this dashboard was to mitigate the manual work of downloading each project and uploading it to different excel workbooks and show the errors directly

through an online dashboard. This was also supplemented with an "error" workbook that allows the BIM coordinators to extract the data from the Power BI dashboard and send it as an excel directly to the users who do not have access to the dashboard.

4.5.1. Exploration and Mapping

The data was different this time than the other dashboards in that it was directly imported from Autodesk Construction Cloud Data Connector. Looking at the data required looking at the schema documentation provided by Autodesk (Autodesk, Inc., 2022). At first, it looks like an overwhelming amount of data, but soon it becomes clear that it gives the exact data you might need in different schema to separate some of the data and keep it organized. For this dashboard the author wanted to examine the reviews only and more specifically the reviews that were part of a Squad Check. Given the data comes pre-set from Autodesk, there was no need to map which sources are required for the check, but rather which checks are required. This creates a bit of a change in the process since the mapping becomes a parallel effort with the linking discussed subsequently. The checks that were focused on are the following:

- Squad Check Status: whether the squad check review is open, closed, or pending closing by the Civil Department. Also checked by disciplines.
- Detailed information about each Squad Check filterable by discipline, project and user. Allows for a department head to check which reviews are missing from their team and a user to check which reviews they need to do.

4.5.2. Extraction

This was the one project that utilized the ability of ACC / BIM 360 to export all projects information for an organization directly into Power BI using the Autodesk Construction Cloud Data Connector. Therefore, the extraction was simple, but formatting the data required some advanced work. The only limitation in this is the user who connects the data needed to have special access to all projects, this limited the author but was supported by the research team.

4.5.3. Formatting

Formatting includes multiple steps as explained in Chapter 3.4, for this chapter, the author will define the queries within the Power BI Power Query Editor. Then the author will show the data linking between the sources. Finally, any important steps or equations that are specific to this dashboard will be highlighted.

Only the queries used for the dashboard will be used. There are plenty more available for use from the Connector. An additional query was imported by the author to help define some peoples' roles in different projects, since the data connector only had one role per person and did not define which project it was related to. This can all be seen in Table 10:

Table 10 – Quality 4 Queries

Group	Query	Source	Objective
Admin	admin_users	-	All users on ACC / BIM 360 for the enterprise
	admin_roles	-	All the types of roles available in the enterprise
	admin_projects	-	The projects in the enterprise
	admin_project_user_roles	-	The connection of user to their role in a specific project
Reviews	reviews_reviews	-	General status of different reviews across all projects and different types of reviews
	reviews_review_tasks	-	Gives each step taken in open and closed reviews
Others	squad_check_workflows	Excel	Gives the workflow to define which departments are required to answer based on project standards

Italicized queries are not loaded into the Power BI Model

Once the queries required for the checks are available, the connection of these queries were created. This was not easy at first since it required searching and understanding the Autodesk schema for the database, but after several iterations, the useful connections were made as shown in Figure 39, again, there are many other tables omitted for clarity.
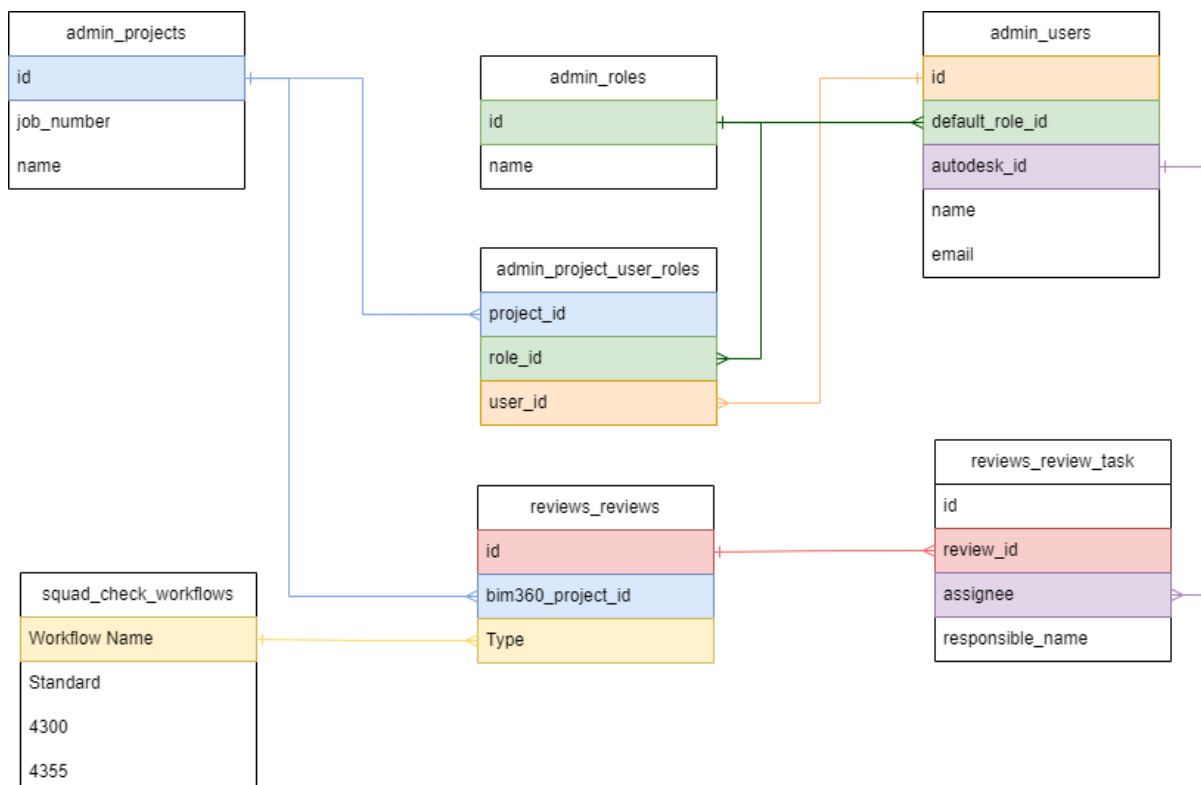
Figure 39 – Quality 4 Data Connections

This project required more work in DAX than the others since the user was not sure whether updating the connector from ACC / BIM 360 would cause the queries to reset. It was also required because only

the lead BIM coordinator had access to the Connector due to security concerns which did not allow the author to work on the queries directly.

Most of the DAX work was done on creating the correct columns to identify roles and missing comments by disciplines in the reviews_reviews query. The following are the additional columns made for the query to be successful, keep in mind that only one department is shown, this would be repeated for the remainder of the departments required.

First, the disciplines needed to be created based on the workflow name to make sure the review was mapped to the correct discipline:

```
Discipline =
IF (
    CONTAINSSTRING (reviews_reviews[workflow_name], "AA"),
    "AA",
    IF (
        CONTAINSSTRING (reviews_reviews[workflow_name], "AI"),
        "AI",
        IF (
            CONTAINSSTRING (reviews_reviews[workflow_name], "AQ"),
            "AQ",
            IF (
                CONTAINSSTRING (reviews_reviews[workflow_name], "AP"),
                "AP",
                IF (
                    CONTAINSSTRING (reviews_reviews[workflow_name], "BB"),
                    "BB",
                    BLANK()))))))
```

Then the type needed to be created from the workflow name to be aligned with the squad_check_workflows from the external query to map the correct department required to the review:

```
Type =
IF (
    CONTAINSSTRING (reviews_reviews[workflow_name], "SQUAD CHECK"),
        MID(reviews_reviews[workflow_name],
        FIND("CHECK",reviews_reviews[workflow_name],,1)+8,100),
    BLANK())

Squad Checker =
IF (
    RELATED(admin_projects[name]) = "P4300",
    LOOKUPVALUE(squad_check_workflows[4300],squad_check_workflows[Workflow
    Name],reviews_reviews[Type]),
    IF (
        RELATED(admin_projects[name]) = "P4355",
        LOOKUPVALUE(squad_check_workflows[4355],squad_check_workflows[Workflow
        Name],reviews_reviews[Type]),
        LOOKUPVALUE(squad_check_workflows[Standard],squad_check_workflows[Workflow
        Name],reviews_reviews[Type])))
```

The next three steps are the final ones to be able to identify the status for each department and how delayed the response time was compared to when it was created. As stated earlier these steps were re-created for different departments but will be omitted from here:

```
Squad PIP =
IF (
    CONTAINSSTRING(reviews_reviews[Squad Checker], "PIP"),
        CONVERT(
            FORMAT(
                LOOKUPVALUE(reviews_review_tasks[updated_at],reviews_review_tasks[D
                epartment],"PIP",reviews_review_tasks[review_id],
                CONVERT([id],STRING),"MISSING"),
            "d-mmm-yy"),
        STRING),
    "-")


PIP =
IF(
    reviews_reviews[Squad PIP] = "MISSING" && reviews_reviews[status] = "closed",
    "NO DATE",
        IF(
            reviews_reviews[Squad PIP] = "MISSING",
            "MISSING",
            reviews_reviews[Squad PIP]))


PIP Response Time =
VAR miss = NETWORKDAYS(reviews_reviews[created_at].[Date],TODAY())
VAR late = IF(
            reviews_reviews[PIP] = "-" || reviews_reviews[PIP] = "NO DATE" ||
            reviews_reviews[PIP] = "MISSING",
            BLANK(),
            NETWORKDAYS(reviews_reviews[created_at].[Date],CONVERT(reviews_revi
            ews[PIP],DATETIME)))
RETURN IF(
            reviews_reviews[PIP] = "-" || reviews_reviews[PIP] = "NO DATE",
            BLANK(),
            IF(
                reviews_reviews[PIP] = "MISSING",
                IF(miss > 45,
                    BLANK(),
                    miss),
                IF(late > 45,
                    BLANK(),
                    late )))
```

Once these new columns are created, they can be utilized in the visualizations as per the use case of the client.

4.5.4. Visualizations

Given the detailed work in the background and the singular aim of this dashboard, only two pages were required: an overview of the company and a detailed page for user to check their personal status.

The overview page seen in Figure 40 can be used for the project leaders and gives the squad check status by disciplines and the total number of squad checks. It also gives a summary of how each department is performing. Finally, it gives the user a chance to pick a date range of how the company is performing.
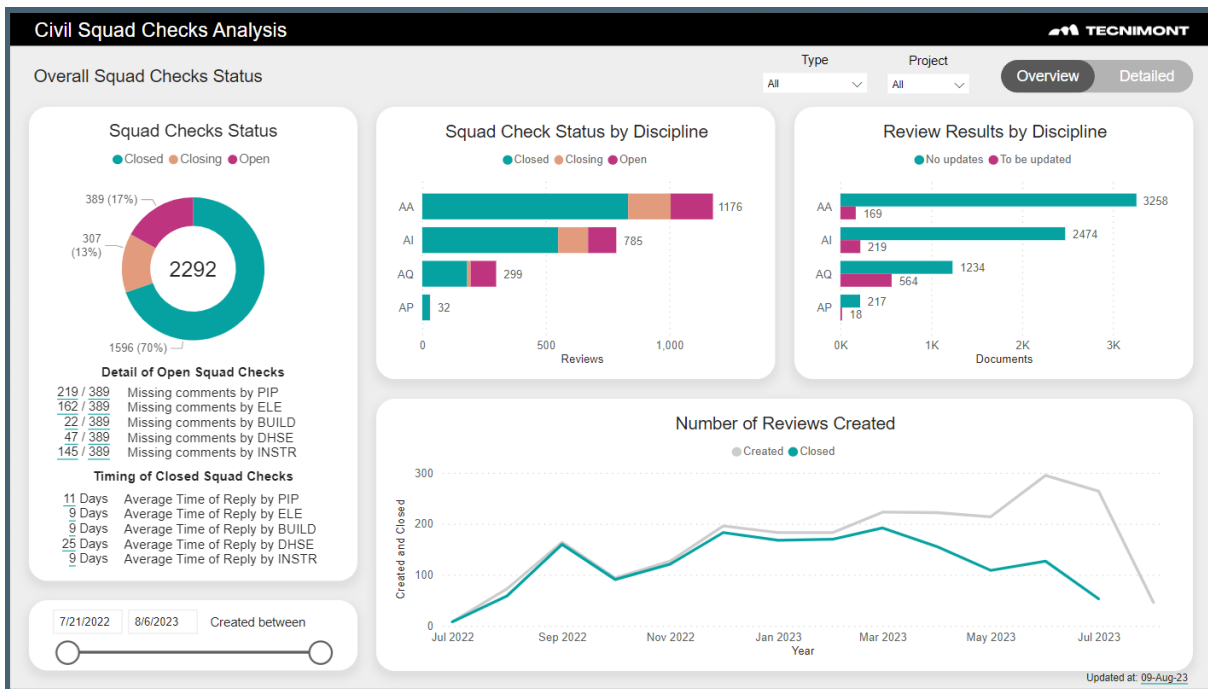

Figure 40 – Quality 4 Page 1

The detailed page, shown in Figure 41, can be used for users that have squad check reviews. It has the same summary as the first page but allows users to get into detail of which reviews' they are responsible for. A user can filter by their department to see which squad checks are still pending due to them, or they can search by specific reviews to see which users have already finished their work and who the squad check is waiting on. This allows both individual users and group leaders to understand who is responsible and who is late in submitting their reviews.
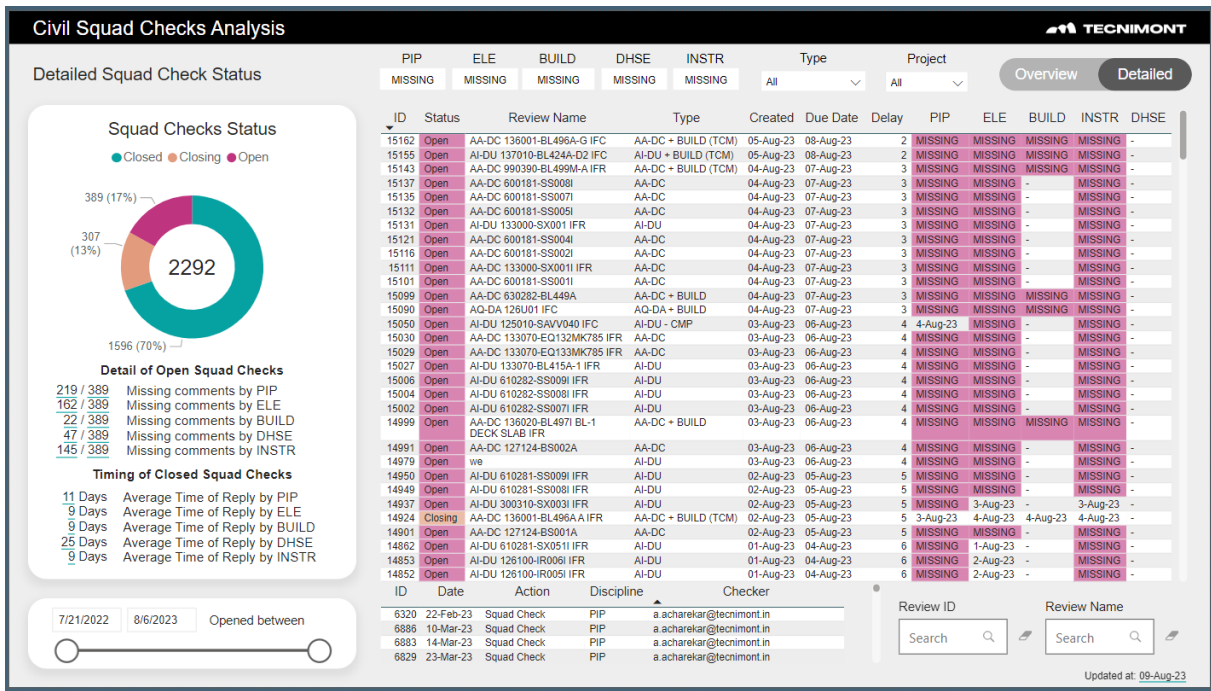
Figure 41 – Quality 4 Page 2

4.5.5. Automation and Testing

Testing was not as necessary for this dashboard since the output was exactly what the client was looking for. Since this data was previously manually exported and recreated into an excel workbook, this was easily accepted and consumed by the users.

This is the only dashboard that is fully automatic due to the data being updated regularly in ACC / BIM 360 and can be imported directly into the dashboard. With an automatic refresh on the Power BI Web Service, this dashboard has the ability to search for data and refresh automatically without any interference from the author. This requires the authorization of ACC / BIM 360 account to be logged in which would be further explored by the client.

The previous version of this check was not as effective due to the large quantity of data that is not related to Squad Checks that is exported from the ACC / BIM 360 platform. This dashboard allowed all BIM coordinators, managers, and involved departments to easily check the performance of their Squad Check and which ones still required action. The amount of work hours saved cannot be quantified due to each individual having their own method for checking Squad Checks, but it must be very beneficial due to the ability to search for the person, review, or department and know exactly what is missing.

4.6. Quantities 1: Civil Quantities

For the final dashboard, the company had wanted to understand the landscape of project quantities by disciplines with visuals that can show the past and broadcast the future of the company. The users of this dashboard are more middle management and higher to understand performance, differences and trends of the project quantities.

4.6.1. Exploration and Mapping

This dashboard takes data from the same CMD / CIL as Quality 3 as shown in Figure 15. This made the mapping relatively simple given the singularity of data from one source. The only thing that required some diligence was the re-merging of the CMD / CIL to create just one table due to some projects living in the CMD and other being in the CIL. An additional excel workbook of historical data of Current Estimates was provided by the company to incorporate a historical analysis of the data. Figure 42 shows the mapping of said elements:
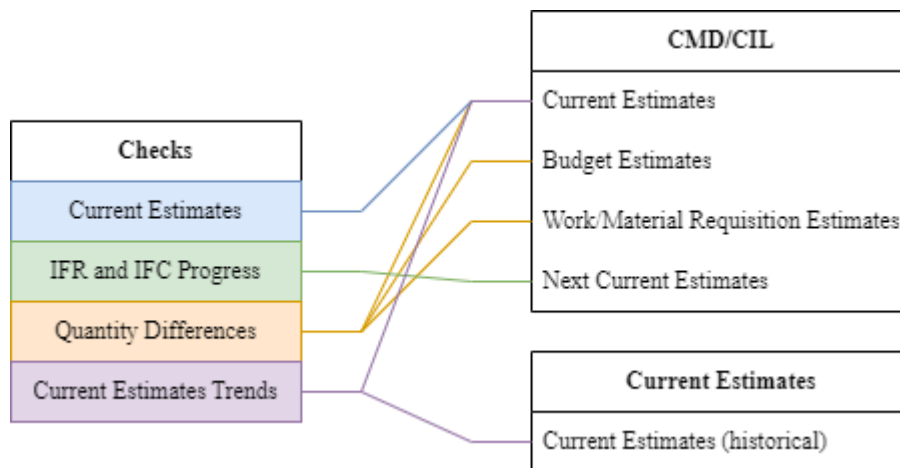

Figure 42 – Quantity 1 Mapping

The checks further explained can be found below:

- Current Estimates Quantities for: Foundation Concrete, Elevation Concrete, Piles, Steel, Underground Pipe, Underground Pit, Paving, and Trench's by their Object Code
- Issue for Review (IFR) and Issue for Construction (IFC) progress for each element in a project
- Quantity differences between QTY Budget, QTY WR/MR, QTY CE, and 3D QTY
- Current estimates trends by time from old projects and future feature to capture data of current estimate inputted with their dates.

With the checks mapped, the extraction method can now be finalized.

4.6.2. Extraction

For brevity, the extraction method for the CMD / CIL is similar to that explained in Chapter 4.4.2. In short, the CMD and CIL are company-based databases that can easily be exported into excel workbooks and contain similar data. The CIL being the older version of the CMD, and for the remainder of this chapter will be considered "Engineering."

The Current Estimates workbook was also provided as an excel workbook of historical data. This was given to the author by the discipline leaders and is a one-time data source given the older projects will not be updated again.

4.6.3. Formatting

Formatting includes multiple steps as explained in Chapter 3.4, for this chapter, the author will define the queries within the Power BI Power Query Editor. Then the author will show the data linking between the sources. Finally, any important steps or equations that are specific to this dashboard will be highlighted.

Table 11 shows the queries and their sources:

Table 11 – Quantity 1 Queries

Group	Query	Source	Objective
Sources	*SourceFolderCMD*	-	Folder location where all possible CMD / CIL's get downloaded
	SourceFolderCIL	-	
	SourceFolderCE	-	Folder location where the historical data of the Current Estimate is located
	SourceFileTCM	-	File with conversion of Underground Object Codes from previous versions to newer ones
Engineering	*CMD*	*SourceFolderCMD*	Combining and cleaning the CMD source files
	CIL	*SourceFolderCIL*	Combining and cleaning the CIL source files and matching the CMD format
	Engineering	*CMD & CIL*	Combining the two sources into one to be used for visualizations
CE	CETrend	*SourceFolderCE*	Combining all historical CE files
TCM	*TCM*	*SourceFileTCM*	Relates old Object Codes to new ones to be used in the AQ discipline

Italicized queries are not loaded into the Power BI Model

Due to the simplicity of this data source, there was not any linking required between the Engineering and CETrend tables. Also due to this dashboard being a visual representation of data and not a logical testing of data, there was no need for many DAX columns to be added. The focus of this dashboard was on the visualizations.

4.6.4. Visualizations

Visualizations for this dashboard included five sheets focused on showing the quantities by filters such as project, discipline, object code, and tag type. Object code is a two to three letter code to identify the type of element such as beam or column, while tag type gives information on the type of sub-discipline such as elevation or foundation concrete.

The first sheet focuses on the summary of all elements and can be filtered by the Object Code. This gives top management an ability to see where the quantities lie by object code and how they can be categorized. The author also created a button to transfer from AA / AI (concrete and steel) to AP / AQ

(paving and underground) as shown in Figure 43 and Figure 44. It is worth noting that Underground Pit and Trench are not available yet but will be used in future projects.
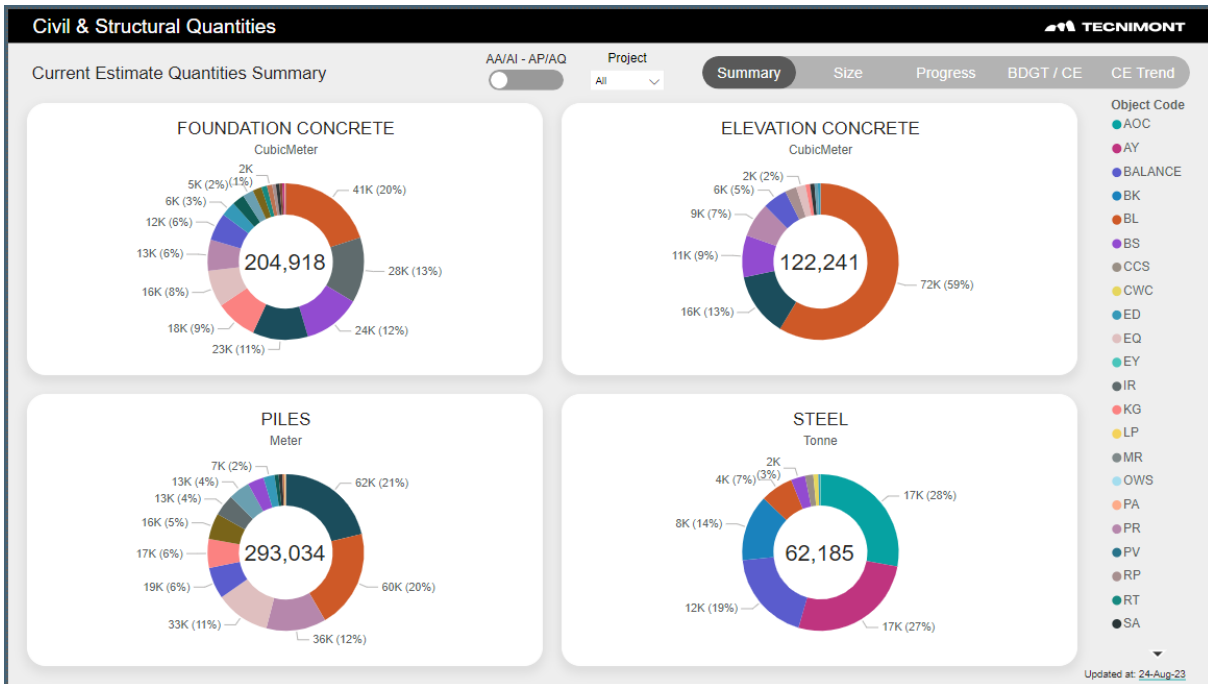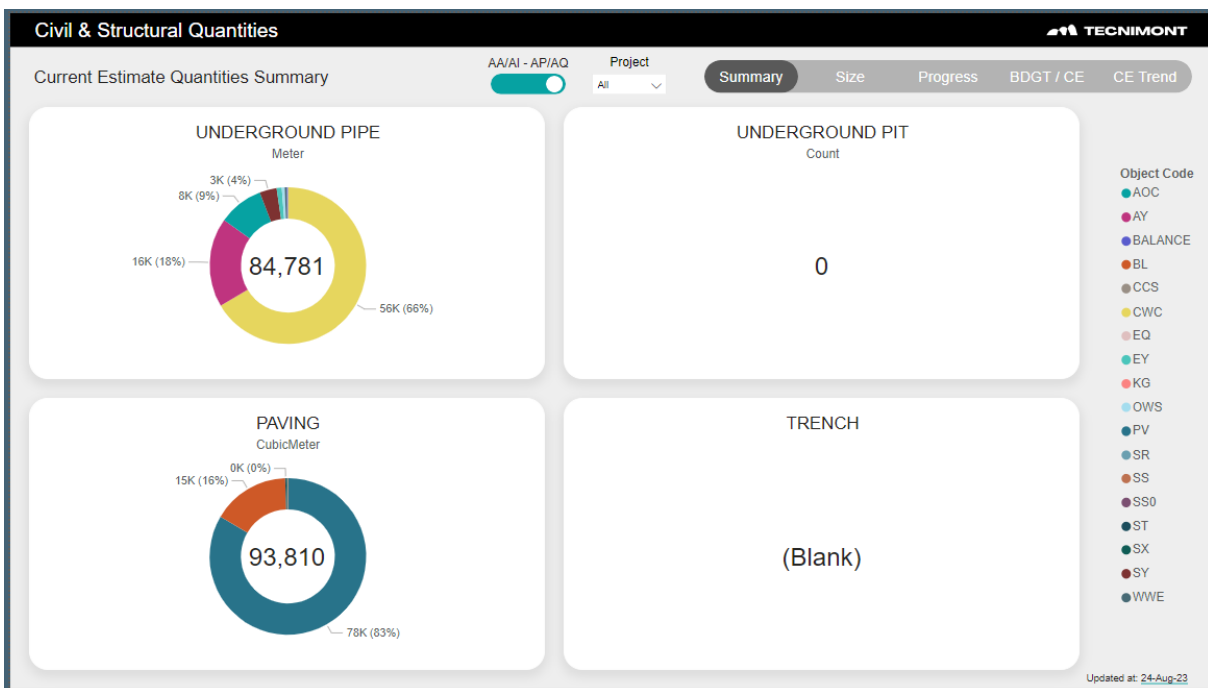

Figure 43 – Quantity 1 Page 1A

Figure 44 – Quantity 1 Page 1B

The next sheet focuses on project sizes to compare projects by tag type. This helps the estimating team understand how much each project was and if a new project needs to be estimated this can give a baseline based on size and location. Figure 45 shows the steel quantities.
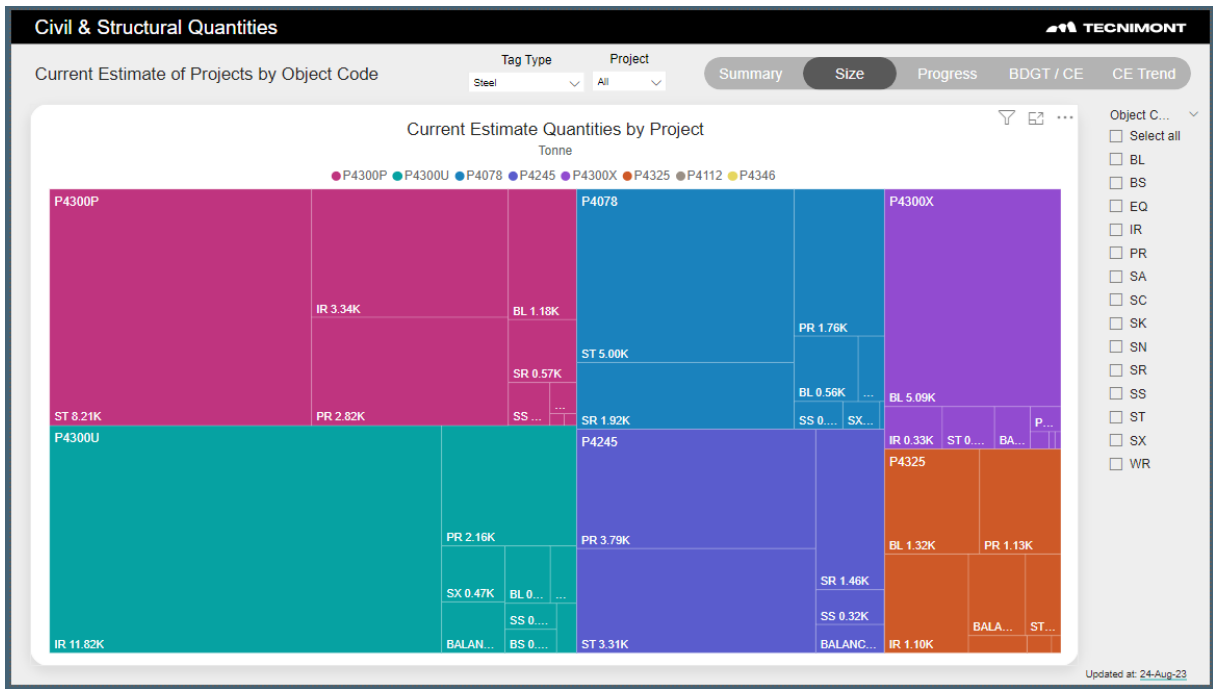
Figure 45 – Quantity 1 Page 2

After that, a project leader might want to understand how far the project is to design completion. The company has two statuses for the design progress of a modeled item: IFR (issue for review) and IFC (issue for construction). The next page, shown in Figure 46, gives the progress of each tag type to be ready for construction by project.
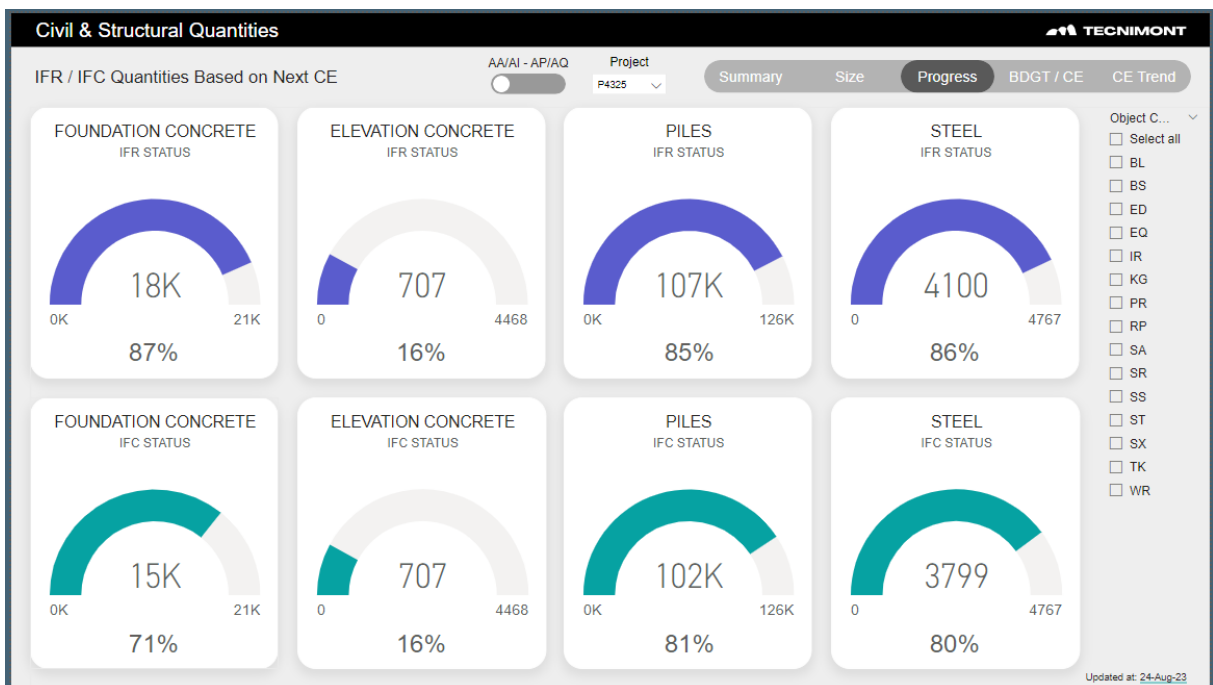

Figure 46 - Quantity 1 Page 3

The next page gives an insight of how projects were estimated and where they are currently. Some are older projects that are closed, and others are still open, but this page shows both to visually see a trend based on tag type. The three quantities used in this visual are: budget quantity which is the first estimate to be put on paper, work/material revision quantity which is the official estimate set before breaking ground on a project, and the current estimate as of today which can also be the final current estimate of a closed project. Figure 47 shows this page.
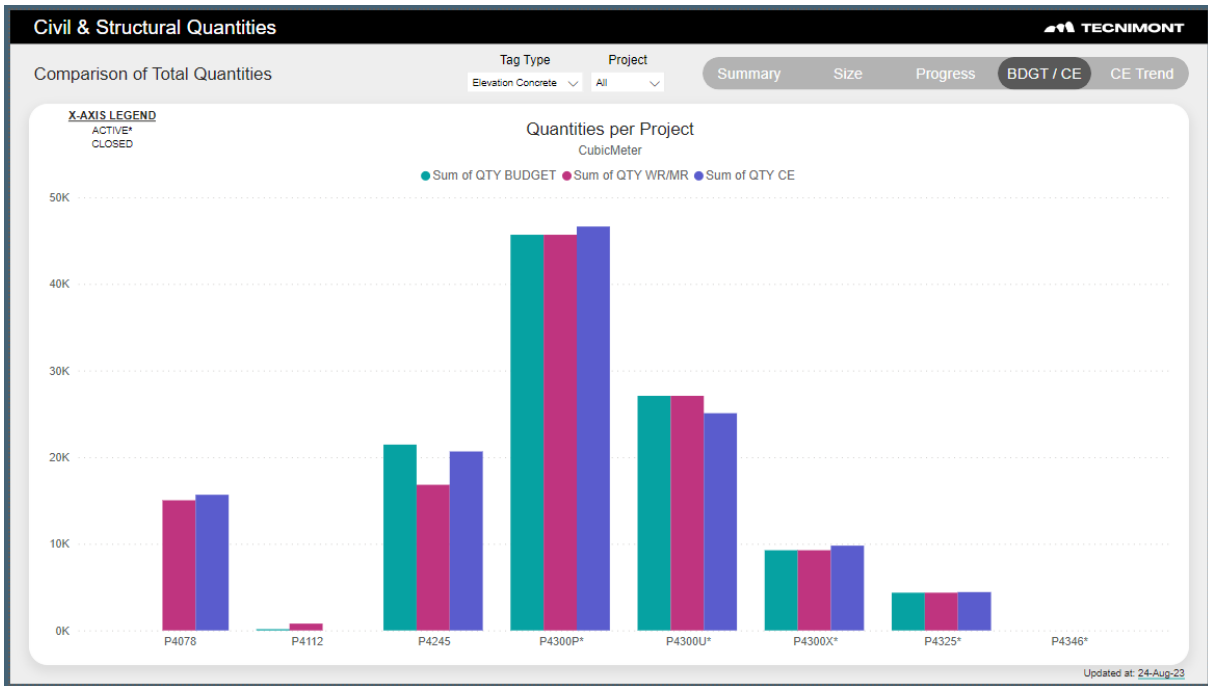
Figure 47 - Quantity 1 Page 4

The final page in Figure 48 shows the historical trend of estimates by project. This takes the budget and WR/MR estimates as well as a current estimate extracted every three months during the lifetime of the project. This helps management see trends in the projects quantities and estimates for the future as well as performance of past projects. It also gives an understanding of how far from the estimate the project was, which can give rise to conversations within the company.

Figure 48 – Quantity 1 Page 5

4.6.5. Automation and Testing

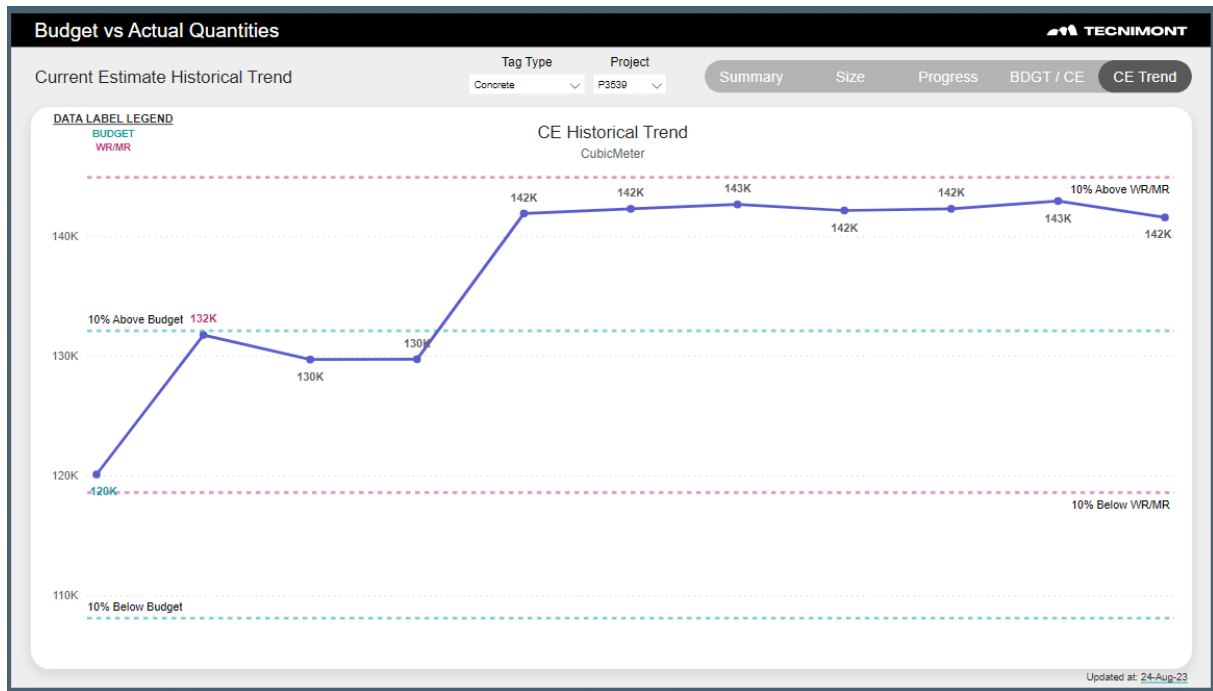The testing of this dashboard was straightforward with many iterations on what the discipline leaders wanted to see visually. Given the data management was not so difficult there was not too many revisions in the query editor or any DAX work.

This dashboard is considered semi-automatic given the Engineering data needs to be re-uploaded to a folder from the CMD / CIL. The option to automate the refresh through Power BI Service is what allows it to be semi-automatic.

This final dashboard is now being used by mid-level management to predict and control project quantities. It has become a favorite for estimating team as well to know the quantities used per project in comparison to other and in comparison to new bids.

This page is intentionally left blank

5. CONCLUSIONS

The goals set forth by MAIRE Tecnimont were achieved with the conclusion of this research. Given the entire goal was to create Power BI Dashboards to handle big data

In conclusion, the field of Building Information Modeling (BIM), has witnessed significant advancement in recent years. These advancements are characterized by increased software availability, enhanced interoperability, an increase in model elements carrying valuable data for construction projects, and broader accessibility to stakeholders across the industry. However, this surge in data availability and model complexity has created a pressing need for effective data management within the Architecture, Engineering, and Construction (AEC) sector.

The research was initiated in response to this growing challenge, commissioned by MAIRE Tecnimont's Civil Department striving to streamline big data checks and ensure data quality throughout their projects. The resulting work created a methodology that helps other Engineering-Procurement-Construction (EPC) or AEC firms in implementing quality checks using any data analytics and visualization software available.

The research also led to the creation of five Power BI Dashboards, four that serve as quality assurance and one that focused on a quantitative summary of project performance. The four quality dashboards focused on model location alignment, element data standardization, database alignment with model elements, and Squad Check Reviews progress. All dashboards acquired data from more than 15 projects that allowed for some automation in the updating of data and analysis. Only the squad check review (Quality 4) dashboard was completely automated due to the Autodesk Construction Cloud (ACC) Data Connector feature which allowed for a seamless connection to the Power BI Web-Based Service. Due to the manual exporting of data for the dashboard focused on data alignment between the BIM model and the Advanced Work Package (AWP) data (Quality 3), the automation was labeled as manual. The goal for this and the remainder of the semi-automatic dashboards is to eventually become automated like the squad check reviews dashboards. This is integral to the advancement of data analytics in the BIM sector.

Due to the iterative process of creating the dashboards, the author was able to identify a methodology that can be adapted to any AEC firm. The process included eight steps encompassing exploration, mapping, extraction, formatting, visualization, automation, testing and reiteration. Exploration focused on understanding data and the required data for the required checks. Mapping was used to compartmentalize the available data to know the workflow of achieving the checks from which data sources. Extraction is the method in which the data can be extracted, further research should focus on the ability to automate extracting data to data analytics software from BIM software. Formatting is the process of cleaning, creating or linking the data extracted. Visualization is an integral part of creating a report that is appeasing to the eye and useful for the intended users. Automation focuses on creating automatic refreshes to ensure the report is up to date and reflective of current conditions. Testing requires the potential users to give feedback to ensure the usability of the dashboards. And reiteration is the process of updating the data to include larger data pools and addressing any comments made during the testing phase. By following this process, an AEC professional can create a dashboard to check many of

the big data properties associated with BIM elements and create reports that can help guide their company to making more informed decisions.

The literature review, although done concurrently with the research, identified a significant gap between data analytics research and BIM research. While data analytics often assumes larger data with negligible anomalies, BIM data is unique given the importance of each data point and the inability to ignore as many anomalies. The landscape of BIM data requires more preprocessing, but it can adapt to the Big Data Management (BDM) principles laid out in the information technology sector. Finally, the literature utilized web-based graphic design principles to guide the visualization basics of this research.

Looking forward, the automation of data updates remains a priority. The vision is to enable the creation of fully automated dashboards that can be regularly updated. This will involve establishing workflows that ensure data is automatically refreshed on a platform compatible with Power BI.

In conclusion, the research has bridged the gap between the data analytics and BIM industries, offering a methodology to enhance data quality and streamline quality checks for AEC firms. As the AEC sector continues to evolve, embracing Big Data Management practices will be crucial for delivering successful construction projects and achieving higher levels of efficiency and quality. The research has also been successful in creating dashboards that are practically useful in MAIRE Tecnimont and save work hours for the BIM team to avoid repetitive work. The dashboards are being used by many employees, from modelers to management to ensure quality and create a checking system that is constantly updated. Each dashboard has mitigated a manual transfer of data to excel workbooks for individual projects which gives the BIM team time to truly understand the performance of projects and people in a wholistic way without losing the detailed information of each project.

The author recommends further research into automation techniques, data integration methods, and the development of comprehensive workflows to support the seamless utilization of big data in BIM. These efforts will not only benefit the industry but also contribute to the ongoing transformation of the AEC landscape.

REFERENCES

Adobe, 2023. Adobe Color. [Online] Available at: https://color.adobe.com/create/color-wheel (Accessed August 2023).

Afsari, K. & Eastman, C. M., 2016. A Comparison of Construction Classification Systems Used for Classifying Building Product Models. Atlanta, Georgia, Georgia Institute of Technology.

Ali, F. H. & Bandi, S., 2021. Associations Between Building Information Modelling (BIM) Data and Big Data Attributes. American Scientific Research Journal for Engineering, Technology, and Sciences, 76(1), pp. 1-11.

Autodesk, Inc., 2022. Autodesk Construction Cloud Connector. [Online] Autodesk.

Correa, F., 2015. Is BIM Big Enough to Take Advantage of Big Data Analytics?. The International Association for Automation and Robotics in Construction, Volume 32, pp. 1-8.

Farghaly, K., Abanda, H., Vidalakis, C. & Wood, G., 2017. BIM Big Data System Architecture for Asset Management: A Conceptual Framework. Heraklion, Greece, Lean and Computing in Construction Congress.

Google, 2023. Material Design. [Online] Available at: https://m3.material.io/ (Accessed August 2023).

Gordov, E. Y. & Gubarev, V. V., 2013. Analytical Review of Data Visualization Methods in Application to Big Data. Journal of Electrical and Computer Engineering, p. 7.

Liebich, T. et al., 2020. Industry Foundation Classes 2x Edition 3 Technical Corrigendum 1, s.l.: buildingSMART International Limited.

Hess, K., 2022. datylon. [Online] Available at: https://www.datylon.com/blog/types-of-charts-graphs-examples-data-visualization (Accessed 2023).

Microsoft, 2023. Data Analysis Expressions (DAX) Reference. [Online] Available at: https://learn.microsoft.com/en-us/dax/ (Accessed 2023).

Microsoft, 2023. Power Query documentation. [Online] Available at: https://learn.microsoft.com/en-us/power-query/ (Accessed 2023).

LIST OF ACRONYMS AND ABBREVIATIONS

Industry:

AEC	Architecture Engineering Construction field
ACC	Autodesk Construction Cloud
API	Application Program Interface
AWP	Advanced Work Package
BDM	Big Data Management
BIM	Building Information Modelling
CDE	Common Data Environment
CIM	City Information Modelling
DAX	Data Analysis Expressions
GIS	Geographical Information System
ifc	Industry Foundation Classes
IoT	Internet of Things
UI	User Interface
VDC	Virtual Design in Construction

MAIRE Tecnimont Specific:

AA	Concrete Discipline
AI	Steel Discipline
AQ	Underground Discipline
CE	Current Estimate
CIL	Civil Item List
CMD	Civil Master Data
CWA	Construction Work Area
IFC	Issue for Construction
IFR	Issue for Review
MR	Material Requisition
WIP	Work in Progress Folder
WR	Work Requisition