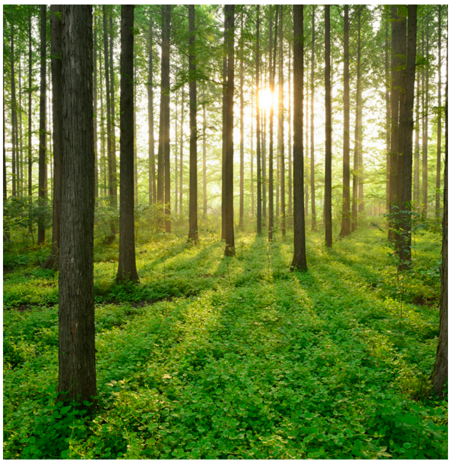
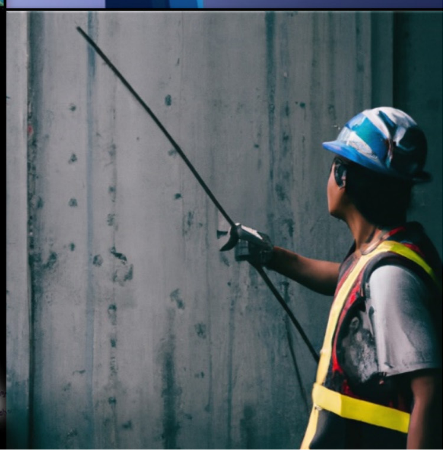
METHODS FOR COLLECTION AND EVALUATION OF CONCRETE DATA WITHIN THE NUCLEAR INDUSTRY

REPORT 2024:1032



NUCLEAR POWER CONCRETE TECHNOLOGY



Energiforsk

Methods for Collection and Evaluation of Concrete Data within the Nuclear Industry

FELIX KÅHRSTRÖM, JUDITH HAMBRAEUS,
MÅRTEN JANZ, PER JOHANNESSON, AFRY

ISBN 978-91-89919-32-7 | © Energiforsk July 2024
Energiforsk AB | Phone: 08-677 25 30 | E-mail: kontakt@energiforsk.se | www.energiforsk.se

Foreword

The Energiforsk Nuclear Power Concrete Program aims to increase the knowledge of aspects affecting safety, maintenance and development of concrete structures in the Nordic nuclear power plants. A part of this is to investigate possibilities to facilitate and simplify the work that is performed in the nuclear business.

The Nordic nuclear power plants now have some 40 years of operation behind them, and the amount of data from inspections and measurements of concrete structures is steadily increasing. Therefore, there is a great need for efficient ways to handle, analyse and store both older analogue data and newer digitalised data. The objective of this study was to investigate and present methods for data analysis and predictions that can be suitable for use within the nuclear industry.

The study was carried out by Felix Kåhrström, Judith Hambraeus, Mårten Janz and Per Johannesson, AFRY. The study was performed within the Energiforsk Nuclear Power Concrete Program, which is financed by Vattenfall, Uniper, Fortum, TVO, Skellefteå Kraft, Karlstads Energi, SSM and SKB.

These are the results and conclusions of a project, which is part of a research programme run by Energiforsk. The author/authors are responsible for the content.

Summary

As the service life of nuclear power plants in Sweden and Finland has been extended, the well-being of the facilities has become increasingly important. Consequently, there are frequent inspections of the concrete structures, resulting in substantial amounts of data being stored. Due to security regulations, data collection and analysis are performed manually. To facilitate future data analysis, it is important to review not only the analysis methods but also the data collection and storage methods.

The project has identified variations in the level of digitalization across the nuclear facilities as well as differences in data storage and analysis methods. Consequently, each nuclear facility possesses its own set of advantages and disadvantages, which the other nuclear facilities can draw insights from. The low-hanging fruits of the project encompass improvements in the current data collection methods and data storage. This would facilitate analysis in the future. The programs IBM Maximo and Screening Eagle Inspect are recommended to be further investigated to improve the logging from visual inspection and facilitate analysis forward. Implementing BIM together with 3D models could also be a viable solution to facilitate the analysis of visual inspections. Moreover, recommendations regarding handling historical and current measurement data are to establish a data repository and implement an ETL (extract, transform, load) process to make data more searchable, traceable, and available. There exists many different statistical methods for trending and forecasting such as various regression methods. These methods need to be applied to real data and evaluated before a suitable method can be chosen and implemented. If real data is not available, synthetic data could be generated to train and evaluate the models. Natural language processing (NLP) is also an interesting technology to make historical data from visual inspections more available and searchable.

An outlook towards nuclear facilities in Switzerland and the USA indicates that they have a similar way of working with concrete inspections and measurements. The best practice in the USA is to collect the data into a database to facilitate traceability, availability and analysis.

Keywords

Data collection procedures, NLP, AI, analysis of data, BIM, data storage, methods of data analysis, concrete evaluation, concrete data, analysis of concrete data, data trending, future trending of concrete data. Datainhämtning, artificiell intelligens, dataanalys, data lagring, metoder för dataanalys, betong utvärdering, betong data, analys av betongdata, datatrendning, framtida trendning av betongdata.

Sammanfattning

När livslängden på kärnkraftverk i Sverige och Finland har förlängts har anläggningarnas skick blivit allt viktigare. Av denna anledning bland annat görs frekventa inspektioner och mätningar av anläggningarnas betongkonstruktioner, vilket resulterar i att stora mängder data lagras. På grund av säkerhetsbestämmelser utförs datainsamling och analys manuellt. För att underlätta framtida dataanalys är det viktigt att granska inte bara analysmetoderna utan också metoderna för datainsamling och lagring.

Projektet har identifierat en varierande grad av digitalisering hos kärnkraftsanläggningarna samt skillnader i metoder för datalagring och analys. Varje kärnkraftsanläggnings arbetssätt har sina fördelar och nackdelar, vilket de andra kärnkraftsanläggningarna kan dra lärdomar från. De lågt hängande frukterna i detta projektet innefattar förbättringar i de nuvarande metoderna för datainsamling och datalagring. Detta skulle underlätta analys i framtiden. Programmen IBM Maximo och Screening Eagle Inspect rekommenderas att undersökas vidare för att förbättra datainsamling från visuella inspektioner och underlätta framtida analys. Implementering av BIM tillsammans med 3D-modeller kan också vara en lämplig lösning för att underlätta analysen av visuella inspektioner. För att hantera historiska och framtida mätdata rekommenderas att etablera en databas och implementera ETL-processer (extract, transform, load) för att göra data mer sökbart, spårbart och tillgängligt. Det finns många olika statistiska metoder för trendning och prediktering, såsom olika regressionsmetoder. Dessa metoder behöver tillämpas på riktiga data och utvärderas innan en lämplig metod kan väljas och implementeras. Om riktiga data inte är tillgängliga, kan syntetiska data genereras för att träna och utvärdera modellerna. Natural language processing (NLP) är också en intressant teknik för att göra historiska data från visuella inspektioner mer tillgängliga och sökbara.

En utblick mot kärnkraftsanläggningar i Schweiz och USA indikerar att de har ett liknande arbetssätt för betonginspektioner och mätningar jämfört med Sverige och Finland. Bästa praxis i USA är att samla in data i en databas för att underlätta spårbarhet, tillgänglighet och analys.

List of content

1 About the project

Nuclear facilities contain a vast number of concrete structures that are vital both from a security perspective and from a structural perspective, as concrete is used for e.g., reactor containments, reactor structures as well as the storage basin and cooling basin to name a few. Inspections of the concrete structures are frequently conducted, to control the status of the structures and ensure the fulfilment of the safety requirements. With the aging of the facilities the amount of available data from inspections and measurements increases. Thus, the need for analysis and evaluation of the available data also increases, to conduct a continuous monitoring of the concrete condition.

The focus of this project is to find and describe potential methods to be used at the Nordic nuclear facilities when analysing measurement data and inspection data from the sites. The focus of the methods is to be able to find normal variations, deviations and trends in the data as well as to be able to predict future values and behaviours together with establishing alert limits and threshold values. That the data is mostly qualitative, and text-based, together with a random sampling of which structure is being inspected, results in some difficulties when analysing the data. This results in big data variations, which is important to keep in mind. The facilities included in this project are OKG, Ringhals, Forsmark and Svensk Kärnbränslehantering (SKB) in Sweden and Olkiluoto and Loviisa in Finland.

As a good data analysis depends on the quality of the data, a big part of this project has also been to map out what type of data is collected and how. The current way of analysis at the nuclear facilities has also been vital to map out and summarize, to understand what methods could be possible to use. Therefore, the first part of the report mainly focuses on data collection and data storage, thereafter, possible methods of analysis of historical data as well as facilitating future analysis of data are presented.

2 Abbreviations, terminology and acronyms

AI - Artificial Intelligence

ARIMA model- Autoregressive Integrated Moving Average model

BIM – Building Information Modelling, a type of digital model

CMMS – Computerized Maintenance Management System

Digital Model – A 3D-copy of a building or any other object where it is possible to digitally see the whole structure.

Digitalization - The process of converting information into a digital (i.e. computer-readable) format

ERP – Enterprise Resource Planning

ETL – Extract Transform Load, a process transforming data to information

FEM – Finite Element Method

IAEA – International Atomic Energy Agency

IoT – Internet of things

NLP – Natural Language Processing

NPP – Nuclear Power Plant

SCI – Screening Eagle Inspect

SKB – Svenska Kärnbränslehantering

SMP – Structures Monitoring Program

THAG-report – Joint document from the NPPs presenting the inspection procedure in more detail and the grading system of each NPP

3 Introduction

The study conducted is a qualitative study, therefore data has mainly been collected through semi-structured interviews. In addition to this, general information about inspection procedures of the concrete structures, together with information about types of data collected, has been provided by Energiforsk's steering group and other interviewees. This data together with the interviews have then been analysed and compiled into this report. The nuclear facilities this project concerns are OKG, Ringhals, Forsmark and Svensk Kärnbränslehantering (SKB) in Sweden and Olkiluoto and Loviisa in Finland.

3.1 BACKGROUND

This project was conducted from November 2023 until April 2024, where a project team at AFRY has investigated methods for evaluation of concrete data at nuclear facilities. There exist different reasons as to why an inspection is conducted, e.g. routine or project-based, as well as different types and levels of inspections. An inspection can be only visual, or it can include different types of measurements depending on the depth of the inspection. Given that this study is primarily concerned with the evaluation of data methods, it will not provide an in-depth analysis of various inspection details. Rather, the reader needs to be aware of the existence of various inspection methods that are used.

3.1.1 Previous studies by Energiforsk related to concrete analysis

This pre-study partly aims to close the gap in-between two other studies conducted at Energiforsk. The first report from 2020, "Förvaltning av kärnkraftverkens betongkonstruktioner"[1], mainly focuses on the technical aspects of concrete inspection methodology with a focus on the water-cooling path. Furthermore, the definition of structures elements and their inspection procedures in BaTMan is investigated regarding the potential usefulness in nuclear facilities. The methodology presented in the report is inspired by the methodology procedure used at Trafikverket, documented in BaTMan. The current report will also examine BaTMan. However, the focus is rather on the analysis tools and data storage within the system. Therefore, the current report and the report from 2020 will be a complement to each other regarding how the nuclear industry could improve its data analysis and data storage systems, as the focus of the two reports differs.

The second report, "Making inspections of concrete structures in nuclear powerplants more efficient"[2], published in 2021 mainly focuses on inspection procedures and how to make the data collection methods more efficient. As the current report has a focus on the analysis of data, there is naturally also a focus on data collection methods as it is something that affects future data analysis.

---

[1] Manouchehr Hassanzadeh. Energiforsk, *Förvaltning av Kärnkraftens Betongkonstruktioner*, Energiforsk Rapport 2020:684.

[2] Peter Cvitanovic, Marit Täpp and Finn Midböe. 2021. *Making inspections of concrete structures in nuclear powerplants more efficient.* Energiforsk report.

Therefore, the overlap is bigger between these two reports, however, the scope and objective differ and therefore they also work as a complement to each other. The report from 2021 briefly highlights BaTMan, BIM and laser scanning, but the main result and focus is rather on how the documentation method during the inspection can be more efficient. The current pre-study will investigate specific functions of management systems and software tools together with potential ideas as to how they could be implemented. Therefore, it could partially be seen as a continuation of the report from 2021.

3.2 SCOPE AND OBJECTIVES

According to the tender document from Energiforsk, the goal of the project is to recommend methods that can be used in the analysis and evaluation of data from e.g. inspections or measurements of concrete structures.

The goal of the project was to:

- Collect and compile information on the current situation in the nuclear power plants.
- Explore and describe different appropriate methods for the analysis of measurement data and inspection data from concrete structures in nuclear facilities.
- Recommend methods that can be used in analyses and evaluations of data from inspections and measurements.

This project is a pre-study. There has not been any possibility of testing and implementing any of the recommended methods, even though they are well-rooted in the data and existing knowledge of the project.

3.3 PROJECT EXECUTION

For this project, AFRY has used a core team consisting of experienced personnel within data analysis and concrete technology. In addition to the core team, experts in other areas such as AI and cybersecurity have been at their disposal as needed. This team has worked in collaboration with nuclear facility representatives from the Energiforsk concrete steering group.

The work has been divided into collection and analysis of data and information, internal and external experience retrospective, and method evaluation and customisation.

3.3.1 Collection and analysis of data and information

The main method for the project was to collect relevant information such as data and data format, as well as obtain an understanding of the problem at hand and the current situation. This was accomplished by conducting semi-structured interviews with people in the reference group as well as other interesting interviewees. The interviews were conducted between December–February. In total 30 people were interviewed though video calls. Two to three people conducted each interview, one interviewer and one to two writers. Afterwards the notes were refined and clarified.

Thereafter, solution proposals were collected from external sources together with the expert competencies in the project group and at AFRY, which is what is presented in this report.

Semi-structured interview format

The semi-structured format of the interviews facilitates rich data collection as there is a potential for the interviewees to express opinions, which then can lead to themes that have not been thought of beforehand. The potential discovery of valuable insights is allowed if the interview questions have been constructed in a manner conducive to sufficient openness, that allows the interviewees to express their own opinions. To create questions for the interviews, the project group conducted a small workshop to discuss the purpose of the interviews and general themes important to the study. The information collected during the workshop was subsequently organized and synthesized into a set of agreed-upon interview questions.

Transcription and analysis of data

When the interviews were transcribed, before commencing the analysis, the responses to each question were shared with the interviewees for review. This precautionary step aimed to ensure accuracy, mitigate the possibility of misunderstandings, enhance transparency, and reduce the risk of bias in the project. A qualitative analysis is only as good as the collected data; therefore, the project group have done their best to make this part of the process as transparent and correct as possible.

3.3.2 Internal and external experience retrospective

AFRY, as a company, has an extensive project portfolio and a large pool of knowledge, some of which can be used to suggest effective solutions for the challenges presented by Energiforsk. When investigating internal and external sources, potential solutions and/or technologies have been identified to address the problem. In chapter 6 these will be discussed more thoroughly as to how they work and the project group's reasoning behind the proposition. Moreover, reference projects at AFRY where similar problems have been handled will also be presented.

3.3.3 Method evaluation and customisation

The project group has identified several possible methods to address the problems identified during the interviews and in the initial request. These methods will be presented and evaluated to demonstrate how they could be a help in tackling this complex challenge. Due to the problems highlighted by the interviewees, the methods presented do not only focus on how to analyse and evaluate data but also on how to collect data in the future. Changes in the method of collecting and the reporting process could facilitate future analysis of data, as well as result in more efficient and structured ways of working.

3.4 REPORT STRUCTURE OVERVIEW

In Chapter 5, *Current metod of data collection and analysis at the facilities*, the findings from the interviews of the Swedish and Finnish Nuclear facilities, as well as international facilities are summarized. This includes methods for collections of data, analysis of data, challenges and proposed solutions, as well as desired long and short term work approaches. In chapter 6, *Identified solutions and Technologies*, Solutions and technologies identified by the project group are presented. In chapter 7, *Opportunity analysis and evaluation of technologies and solutuons*, the solutions and technologies in chapter 6 are further evaluated as to their applicability to the Nordic Nuclear facilities. In chapter 8, *accounting of main risks and/or potential difficulties*, risks in implementing the technologies are presented. Furthermore, chapter 9, *recommended methods and next steps presents recommendations based on the previous information in the report* and chapter 10 concludes the report.

4 Current method of data collection and analysis at the facilities

In total 12 people from Energiforsk's steering group and some of their colleagues from the different nuclear facilities have been interviewed. From these interviews, a general understanding has slowly formed regarding the current method of analysis and data collection that seems to be used at the facilities and possible solutions to the current problem.

Disparities were observed in both data collection and storage methods between Swedish and Finnish power plants. Which parts of a plant that are considered critical also differ as the structure might be different, e.g., the reactor containment is made of steel at Loviisa, Finland. This means that there is a slight difference in the inspections carried out at each place. However, pressure tests are important both at Finnish and Swedish plants which also was highlighted during the interviews.

The Swedish power plants have a joint THAG-report[3], which presents the grading system in use at each power plant together with a specification of the inspection interval for each element. Moreover, there is a general description of the approximate extent of each inspection. In the following paragraphs, each plant's general method of conducting the inspections and compiling the data will be presented.

4.1 OKG INSPECTION AND DATA COLLECTION PROCEDURES

At **OKG**, in Sweden, the interviewees explained that there are thousands of rooms that all are connected to various inspection rounds, which are typically divided by floor or accessibility, with one floor being inspected at a time. The inspections are mainly visual, where identified deviations will be documented with photographs which in turn is complemented by a grade representing the severity of the deviations. In the report, there are also prior pictures acting as a comparison. To mitigate the risk of subjectivity, inspections are typically conducted by two individuals together.

The inspection interval is three years, and the system (ODU) will keep track of when an inspection is due and send out an automatic work order. Initially, inspections were conducted annually, but this was later modified to a three-year cycle, allowing for a more meticulous examination during each inspection. An exception is locations that are more challenging to access, such as water canals, which have longer inspection cycles, but each inspection is also more thorough.

Smaller things identified in the day-to-day work will be noted as an observation, if this observation is not fixed within a reasonable time a deviation report might be created. As observations are the foundation for preventing major accidents at the plant, campaigns have been launched to encourage all plant personnel to complete

[3] The THAG report presents the inspection procedure in more detail and the grading system of each NPP

them more frequently. It is preferable to have an abundance of observations and deviation reports of potential safety hazards rather than a scarcity of them. The observations are stored in a different system than the identified deviations from the inspections, where the affected section also is tagged.

Maintenance engineers are generally responsible for conducting inspections, and while there is a desire to use the same person for each inspection, there are no formal requirements in place according to an interviewee. A document outlining inspection procedures is currently being created, to ensure consistency in knowledge and procedure.

4.2 LOVIISA INSPECTION AND DATA COLLECTION PROCEDURES

At **Loviisa**, in Finland, there is a document which includes all the buildings and information about the annual inspections. The report contains pictures that indicate the locations to be inspected, along with instructions on how to conduct the inspections and the documentation procedure. Upon completion of an inspection, the document will be marked as completed. Utilizing an external company for inspections, the same consultant typically conducts the annual inspections, despite not being an internal employee at the plant.

Analysis and trending of data

At Loviisa there is a digitalized system/program in use, in a similar manner to OKG, that is called Lomax. A protocol with specified requirements will be filled out and sent to the Finnish Radiation and Nuclear Safety Authority (STUK) each year, where all deviations will be brought to attention. Analyses are conducted only when a deviation is noted or when requested by personnel. However, some parts are continuously monitored to promptly detect deviations, e.g., seawater tunnels and reactor containments.

4.3 OLKILUOTO INSPECTION AND DATA COLLECTION PROCEDURES

Olkiluoto, in Finland, differs from the other plants regarding their usage of technology. There are live systems installed at the plants to be able to monitor how the whole structure is moving. One of the installed online monitoring systems is the same type of system in use at some bridges, however, it is hard to understand the results, according to the interviewee. The data from the different systems and sensors are collected locally at the plant, and the sensors are wired, which means Wi-Fi cannot be used. Many different types of systems are used where data is stored at the plant, therefore, it might be time-consuming to find the correct data. Due to the amount of data that continuously is measured, it is complex to analyse and understand all the information gathered, according to the interviewee.

In addition to the continuous monitoring, regular visual inspections are also performed. Every four years, a thorough inspection is conducted, which may involve both real-world assessments and non-destructive testing (NDT). Each room within the facility undergoes visual inspections, scheduled either every second year or every fourth year based on its criticality. The digital system (Enko) currently in use for visual inspections bears a resemblance to the systems in use at

Loviisa and OKG. It stores inspection documents and facilitates automated reminders.

Analysis and trending of data

The inspection reports include specific data from the prior two years, but there are also trends which display the entire historical evolution. Every second year STUK inspections are done, which includes both visual inspections and survey studies. All inspections except these are managed internally. Moreover, the STUK inspection reports are stored in a different system.

Furthermore, concrete experts from AFRY conducting different types of measurements and analysis of the concrete data at the plant, have been interviewed to learn more about important aspects of concrete data. An important aspect when analysing concrete data is that it is not enough to only have measurements, one also needs to know why different deviations have arisen. Therefore, there is complexity associated with trend analysis that needs to be considered. Mechanical data, such as tensile strength and compression strength data that is numerical and easy to measure is usually more fitting for conducting trend analysis. However, to only conduct a trend analysis of the mechanical data will not be enough to gain a comprehensive understanding of the concrete structure.

4.4 RINGHALS INSPECTION AND DATA COLLECTION PROCEDURES

Depending on the system, environment, and risk of degradation the inspection interval spans from annually up to once every three or four years. **Ringhals**, in Sweden, uses SAP as a maintenance tool, which in turn is connected to the document system Darwin from **Vattenfall**, which is where the documents are stored with a specific ID. According to one of the interviewees, there are two main reasons for the inspections: reactor safety and the safety of personnel. Like other nuclear sites, photographs are put into the report together with a short assessment, which is both written in text and graded in an assessment matrix. The photographs lack date annotations which makes it impossible to identify them without the text data. One of the interviewees mentioned a control report of concrete deviations where the photo from the prior report was reused, instead of taking a new picture.

Analysis and trending of data

When conducting the inspection, information from the prior inspection will be examined and followed up, but eventual trends are not shown. The inspection protocol template is always the same, however, it is up to each inspector how to take the pictures and decide how things are written. The size of each inspection also differs, but according to one interviewee, there have been inspections where the protocol in the end had 169 different IDs written. An inspection that big takes three to four days to finish. Deviations can also be identified during day-to-day work; these will be added to SAP where they will be handled by the affected department. However, approximately 70-80 % of the identified deviations will be found during screening and preventive maintenance.

4.5 FORSMARK INSPECTION AND DATA COLLECTION PROCEDURES

At Forsmark, in Sweden, the inspection interval spans between yearly to once every six years, depending on the importance of the structure. Visio[4] is a new concept where the visual inspections are supposed to be uploaded, this program enables plans to be uploaded into layers which facilitates comparison between the years. Due to the plant not having wi-fi or 4G the inspectors use pen and paper. During the inspection paper plans are filled out which then need to be uploaded to Visio, once back at the office. According to the interviewees, this presents a risk of forgetting to upload the information into Visio. Moreover, the document is stored at G; which is a hard drive specific for each group, therefore it is not accessible to all personnel at the plant. This presents a risk as the knowledge of the Visio document becomes bound to specific people which could lead to it getting forgotten if these people would quit without clearly passing on the knowledge. According to the interviewees, there are several similar problems at Forsmark, where procedures fall into oblivion due to knowledge being tied to specific people. To ensure knowledge is kept and passed on it should be saved onto the system Arken, according to the interviewees.

SAP is used as a maintenance tool; however, this is also a rather new system and according to the interviewees, the reason behind the change might be that Ringhals and Forsmark should use the same system. Before a system named Fenix was in use, where the main difference was that Fenix could not handle images, however, the protocol and information in Fenix had a more mandatory procedure. Therefore, the new system on the one hand is seen as an improvement due to the possibility of linking images of the defects straight away into the system, but on the other hand, the new system is seen as a deterioration due to inspection protocols getting more subjective and varying depending on the inspector.

Analysis and trending of data

Trending of data is done to some extent by putting data into Excel and using acceptance criteria to compare it. The data that currently is being trended using graphs is test data from e.g., concrete, rubber, and pre-stressing cables. The interviewees emphasized that using Excel facilitates easier comparison of testing data. Depending on the types of data the length of trends varies, the only thing having been tested from the start, and therefore already having long trends, is testing of pre-stressing cables. As personnel is changing things are forgotten, therefore, there are some data where the ambition from the start was to conduct trend analysis but with time this has been forgotten.

4.6 SKB INSPECTION AND DATA COLLECTION PROCEDURES

In addition to nuclear power plants, there are also nuclear storages where the concrete structures are equally important. At **SKB** (Svensk Kärnbränslehantering), in Sweden, major parts of the structures are built within the bedrock, which means there are both rock- and concrete inspections to be conducted. There are two types of inspections required from authorities: one small and one big inspection. The

[4] By Microsoft

small inspection is to be conducted each year, and the bigger inspection is every fourth year in which the Swedish Radiation Safety Authority (SSM) also will take part. Internally, every room is required to be inspected every second year. The information is stored in a database (GDU) which sends out reminders when an inspection is due, this system is specifically used by the maintenance team. Moreover, there is a document handling system (SKBdoc) where all other types of reports and documents will be stored.

At SKB there is an openness to technology, however, there are not as many systems in use yet. However according to the interviewee there already exists 3D models of all the new buildings. There are currently discussions regarding the possibility of digitally conducting the investigations.

4.7 SOLUTIONS AT NUCLEAR FACILITIES IN OTHER COUNTRIES

To gain more insight regarding potential improvements within the Nordic nuclear facilities, the nuclear industries in Switzerland and the USA have briefly been investigated.

4.7.1 Swiss inspection and data collection procedures

Based on information from an interview with a concrete expert, who has worked in the nuclear industry in Switzerland, the process of data collection and data analysis does not differ significantly from Sweden and Finland. The inspections are a mix of visual and hands-on, the environment is very controlled. The interviewee recommends using different kinds of non-destructive testing methods combined with destructive testing and visual inspection. If one wants to be able to follow up on the work, it is important to have it well organized; possibilities of analysis depend on the quality of the collected data.

When analysing data and doing follow-ups on time-dependent data the nuclear plants in Switzerland usually hire engineering companies, according to the interviewee. Every room is numbered, therefore, to be able to know where each inspection has been made one needs a room book. Tools such as a portable crack meter, and sometimes a stationary crack meter, is used to measure deviations. Cracks smaller than 0.3 mm were not considered in the analysis/investigation, as these are not assumed to be critical.

There is a specific project handbook that describes everything that is needed to know; therefore, information is not person-bound. The data is stored in Excel or other such document with room, structure, and other information.

4.7.2 American Inspection and Data Collection Procedures

In a report from EPRI[5], best practices related to inspection recording and record retention within supporting a structures monitoring program (SMP) are described. The report consists of collated example SMP procedures from eight US utilities.

[5] EPRI, "*Structures Monitoring Program Inspection records - Best practices for Preparation, utilization and retention*". Published August 2021, https://www.epri.com/research/products/000000003002020988

There exist differences between the plants in how the inspection records were prepared and retained, for example in the levels of detail.

The report describes the procedure of inspections briefly and what is best practice to record during an inspection. These are:

- Locations inspected
- The structure or structural element inspected
- Inspectors and applicable date
- General description of observed conditions
- Photographs and sketches of observed condition
- Inspection criteria categorization - reference to separate evaluation reports or similar

It is recommended utilizing a standardized form or template during the inspection process. This approach ensures consistency and facilitates thorough documentation.

It is not described whether the inspections are performed with pen and paper, but it is described that notes are taken during the inspection and a "walkdown report" is later written and stored in a system. When deviations are found, the evaluator should compare the photographs with previous inspections to determine the rate of deterioration. Work orders and conditions reports created as a consequence of the need for corrective action should be referenced in the walk down report or inspection record.

The inspection records and reports should be stored in the SMP database. The SMP database is described as follows in the report from EPRI:

"The SMP database contains all relevant information regarding the monitoring of conditions of structures under the SMP. The goal of the SMP database is the following:

- Accessing all conditions identified during the implementation of the SMP
- Tracking the status of the current condition
- Trending of conditions to inform evaluations (both on an observation level and a plant level, as discussed in the local evaluations section."

There is an emphasis that the data is recommended to be stored in a simple and organised manner and that the most important content should be summarised in a table structure in the SMP database, see Table 1, Table 2. This allows it to be more searchable and filterable. Using this table gives a better overview of the inspection history and can also be used for trending.

Table 1: Recommended table to summarize important information from a visual inspection.

#	Building id	Room/ area	Elevation	Location	Observation	Current evaluation level	CR	WO

Table 2: Continuation of Table 1.

CR/WO status?	Walkdown report ID	Current inspection frequency	Prepared/reviewer

4.8 GENERAL/SUMMARIZED METHOD (DATA COLLECTION)

Based on the information gathered from the interviews, it has been identified that each plant employs at least two distinct digital systems. One is mainly dedicated to document storage, whereas the other is a maintenance system utilized, amongst other things, for tracking inspection schedules. At each power plant, deviations are assessed using a customized graded scale, at some plants it spans from 0-4 whereas at others from 1-5. However, it appears that data collection methods and the overall execution of inspections remain consistent across all power plants in both Sweden and Finland.[6]

At Swedish power plants, inspections are carried out by internal personnel, with variations in competency requirements across different plants. However, in recent years, authorities have imposed more rigorous training requirements for individuals involved in conducting investigations. Swedish sites also hire specialized competencies when analysing data if it is considered necessary. In Finland, the knowledge seems to be both internal and external, where the external competencies will be hired from specialized companies.

As the inspections are mainly carried out using pen and paper, interviewees spoke about the nuclear industry "*still being stuck in the 1800th*". Various methods exist to locate a previously identified deviation in a room. Some approaches involve the use of coordinates, while others rely on maps or cardinal points. However, it is important to note that the position of the deviation will be approximate, as it is up to the investigators to estimate its location. This means that the next investigator needs to spend time searching for the exact position of the earlier deviation. Moreover, a lot of time will be spent on writing the report after the inspection round is finished. All the information collected with pen and paper together with the photos needs to be put into a report.

The inspections are overall subjective, however, during the interviews it was highlighted that each level on the grading system is connected to a definition. The definition exists to minimize the subjectiveness of the inspection. But even so, each inspector decides what kind of deviation is considered severe. According to an interviewee, data from early reports differs slightly as the inspections were not as meticulous then. Therefore, manual work is required if an analysis of the trend in a specific room were to be made. However, some sites continuously monitor temperature, leakage, and other such operational data.

[6] The THAG report presents the inspection procedure in more detail and the grading system of each NPP

4.9 CURRENT DIGITALIZATION AT THE PLANTS

At **OKG** there are coordinates for the entire plant that are intended to relate to scans of the entire plant. An interviewee mentioned that scans are currently being conducted, but they will not be merged into a complete 3D model at this time, even though it would be beneficial for day-to-day operations. There exists 4G wireless network at the entire plant, which facilitates potential installations of more digitized operation procedures.

Loviisa has a digital system where inspection reports are stored, similar to the other plants investigated in this study. But apart from this digital storage, there appears to be no current digitalization underway.

Olkiluoto appears to use technology to a bigger extent than other plants investigated in this study. According to the interviewee cathodic systems over the civic structures have been in use for over 20 years already. There is also an extensive usage of different sensors within the concrete that enables trend analysis. Even though there is not a digital model in use today, some smaller parts of the plant already have digital 3D models. Wireless networks such as e.g. Wi-Fi is something that does not exist at the plant, and according to the interviewee will probably never exist, due to security reasons and technical reasons.

Ringhals has 3D images of the plant, however, these are not used when carrying out inspections. Moreover, the images are not connected to form a model, instead, there are single images stored at various places. According to the interviewees, there are tools within SAP that might be useful for trend analysis. However, the knowledge regarding how to use these tools seems to be lacking. At locations that are difficult to access, due to the height of the walls, drones have been used. Except for 3D images and, to some extent, the utilization of drones, Ringhals exhibits limited digitalization.

SKB also has digital systems where documents and inspection scheduling are stored. Every building is divided into different systems, and every room has specific numbers; These are in turn connected to plans. Therefore, there is an efficient storage system, however, it is not digitalized and hence it is not possible to easily find all reports for a specific location. SKB is currently creating 3D images of all the buildings, which are to be connected to the documents in the future. Even though there are digitalization plans underway, the current digitalization is limited similar to the other nuclear plants being interviewed in this study.

Forsmark has recently updated the data systems in use to store data. It is a similar way of working as OKG and Ringhals, however, the interviewees did not mention 3D models. There are currently plans/ideas at Forsmark to try taking tablets into use when conducting the inspections. Almost all data is collected manually, there are no sensors or other types of digitalization at the plant.

4.10 IDENTIFIED PROBLEMS BASED ON THE INTERVIEWS

From the conducted interviews a few specific problems have been identified, that have been highlighted by interviewees from four of the six sites. As a major part of data analysis is dependent on available data to analyse, different problems not only specifically connected to the method of analysis have been mentioned.

Forsmark and Olkiluoto differ a bit from the other four plants that all have similar strengths and weaknesses. Olkiluoto already collects a wide array of different types of data, some of it seems to also be used in trend analysis. Therefore, the main problem identified from the interview rather seems to be that there is such a vast amount of data that all are stored in different places and therefore difficult to find. Olkiluoto is thus a plant worth investigating further and for the other plants to take after regarding data collection. Forsmark also appears to have a broader data collection than the other plants, and also a bigger openness for trying technology. The main problem identified there is rather that a lot of information is tied to specific people which results in a lot of information and ways of work getting lost.

Manual work and the need for digitalization

The problems highlighted mainly concern the current manual workload, which is time-consuming and ineffective. At present, analysing the existing data is cumbersome, as the information must first be extracted from each report. This in turn also leads to reduced searchability of the data. Another problem highlighted is that there is a need to make information more available and standardised. Standardisations as to how trending should be conducted and on what data, as well as where these trends and analyses should be stored, are lacking.

The interviewees have all in different ways spoken about the difficulties with the nuclear industry not being digitalized. When examining a power plant, the formerly conducted report will serve as a base regarding which parts of the plant to focus the inspection on. However, as it is only possible to show approximations of where a deviation might be in a room, it is not unusual that a lot of time is spent trying to find the correct spot. Moreover, the administrative work takes a lot of time for the inspectors as everything is first written down using pen and paper, to afterwards be summed into a digital report.

Key competence holders and the risk of important information getting lost

One of the interviewees highlighted another problem, regarding competencies within the company being bound to specific people. Currently, many issues are solved with manual work, that is not defined or written down. Some people have worked within the industry for a long time and know what to do and solve the problem. Furthermore, the knowledge on what analysis to do when and on what data was also mentioned to have been connected to specific people. Therefore, the data and trends have been stored on that specific person's work computer, and not in the documentation system where everyone can access the data.

This presents a risk for the future when these people are no longer working at the plant, or if something were to happen to any of the key competence holders before their knowledge is documented. One interviewee mentioned important knowledge and data getting lost due to this, where a standardised way of working might come and go over the years due to employees quitting and new ones coming in without the information getting adequately shared and transferred.

However, most of the power plants seem to be aware of this problem already, and therefore working on better documenting the knowledge. However, it was highlighted by another interviewee that there is such a large number of regulations and documents that new employees do not have the possibility of learning them all. Therefore, there is also a need for the knowledge documentation to be compressed into easily understandable standards, where the information is easy to find and not stored in a way where the employees explicitly need to seek that specific information.

4.10.1 Security

When discussing the need for digitalization, the associated concern of IT security was brought to attention. The main reason that digital tools are not used to a large extent at nuclear facilities, is because of IT security. It is not considered secure enough to have data stored and easily accessible in a cloud. Another difficulty with digitalization is that the security regulations hinder the usage of direct support from the suppliers, according to one of the interviewees. Therefore, sudden or unexpected issues with the system would be more difficult to solve than at companies with less strict policies. This means that digitalization would have to be customized to the nuclear industry, to comply with all regulations.

4.10.2 Documentation

The initial documentation during the inspection is conducted manually using pen and paper. When a deviation is identified it will be written down, the location will be marked on a map (or described) and lastly, a photo will be taken of the identified deviation. Thereafter, all this information will be collected and compiled into a report. One of the interviewees who has been conducting these investigations spoke about the difficulty of pinpointing the location of some of the deviations from the former report. This was due to a predecessor that only took close-up photos of the deviations in the concrete. A deviation with a small area can be almost impossible to find in a big room if the approximate location is not clear. This implies that there are no set guidelines at the plants for how a picture should be taken, nor knowledge about what kinds of pictures are needed in a report. Interviewees who have been involved with inspections or have had interactions with the inspection process at the power plants have highlighted issues with the current manual and time-consuming work methods. Moreover, because of the time spent to find the former deviations and be able to check whether they have been fixed, together with the time spent transmitting the information into a report, the time to thoroughly examine for any new deviations is limited.

4.10.3 Searchability

Based on the information gathered from the interviewees all the power plants seem to have a digital system where the finished reports are stored. However, the data is only to be found in the actual reports and some of the plants lack keywords or such connected to each report that would facilitate searchability. Therefore, to obtain specific information, it is necessary to manually search through each report. As every report is approximately 100 pages, based on information from the interviewees, it is safe to assume that searching for specific information is time-consuming. While there is a searchability feature within the data, it is complex and time-consuming, and an efficient searchability method is not present. Consequently, if one were to conduct a trend analysis, it would require a significant amount of time to gather sufficient data from previous reports.

4.10.4 Regulations

In Sweden, conducting inspections is mandatory, but there are no specific regulations governing the inspection process. Similarly, no specific regulations exist concerning the administration and interpretation of long-term or general data. The interviewees in Sweden referred to the law of planning and building (SE plan- och bygglagen) which requires that an inspection is conducted but does not define what should be included in the inspection. To ensure a high standard some of the plants are creating documents that define and clarify the inspection procedures. Additionally, these documents will act as a support to the person conducting the inspection.

Amendments to the inspection regulations are imminent, as SSM introduced new provisions last year (2022). There are recommendations to follow the praxis within IAEA, which the interviewees mentioned could be done to some extent. Though, IAEA is mainly based on the American way of working which differs from the needs in Europe and Sweden. Therefore, some plants incorporate internal knowledge and experience into inspections, in addition to following the guidelines set forth by IAEA.

Finland on the other hand seems to have formal regulations that also specify the inspection content and/or method. All inspections are therefore conducted by external experts, which at Fortum is hired from Renovatek according to an interviewee.

4.10.5 The challenges with analysing and trending concrete data

Trending and prediction based on measurement data of concrete structures is not a trivial task, and sometimes not even possible. Three concrete experts interviewed regarding data analysis were not sure of the amount of data that would be needed for trend analysis to be reliable. To do a trend is easy, but for the trend to have an accurate and reliable result is not as easy.

The three experts clarified that an important aspect when analysing concrete data is that it is not enough to only have measurements, one also needs to know why different deviations have arisen. Mechanical data, such as measurements of tensile strength and compression strength that are numerical and easy to measure is usually more fitting for conducting trend analysis. Crack widths should also be easy to trend. It is important to note that however numerical the data might be, only conducting a trend analysis on the specific data might not be enough to gain a comprehensive understanding of the concrete structure.

The experts said that the reason for concrete inspections and analysis to usually be put into extensive reports is that a lot of information is needed for the analysis to be reliable. Thus, it would not be helpful to only extract small parts of data from the reports and try to compare them with other reports as the analysis is more complex than that.

Cracks in concrete can appear for many different reasons, some are natural, and some are not. Therefore, not all cracks are dangerous or need analysis. A more important aspect of the cracks is to know why they appeared and thus how they will behave in the future. A natural, normal crack that is caused by shrinkage might be easier to predict than a crack caused by damage.

4.11 PROPOSED SOLUTIONS BY INTERVIEWEES

The interviewees highlighted potential solutions to the addressed problems, which would simplify a significant part of the work at the plants. The proposed solutions go hand in hand with the solutions the project group has identified in this project.

One interviewee answered, *"I would like a digital governance model which could be connected to each room, to which metadata and reports also could be connected"*. Other interviewees also highlighted that a digital model that allows for immediate input of information during inspections would be beneficial. Other interviewees were asked about the existence of a digital twin or 3D model or similar tool, which they denied but also highlighted its potential usefulness.

Interviewees from each plant all spoke about the wish for a digital hand computer or the possibility of using an APP or suchlike to fill in information during the inspection, instead of using pen and paper. In the Energiforsk report from 2021[7] the advantages of a digital hand computer were also highlighted. At Ringhals the possibility of using a tablet during inspections has been investigated, but up until now, the issue of IT security has hindered this development. At SKB there is an internal Wi-Fi in place, but not at the rest of the plant. At Olkiluoto Wi-Fi and 4G will never be acceptable due to security risks according to the interviewee. However, OKG differs as there is already 4G infrastructure in place throughout the entire plant according to an interviewee. This development implies that there is a possibility to digitize the inspection processes. Indeed, Wi-Fi and 4G might be a security issue, however, the usage of a tablet to conduct inspections does not necessarily need to be online.

The possibility of more meticulous trend analysis was also highlighted. An interviewee spoke about how it would be helpful to see trends regarding the number of deviations and fault reports. If more deviations than anticipated were to be registered, this would indicate a larger issue that could be addressed before it becomes too severe. Moreover, the visualisation that a digital model provides was highlighted as a potential benefit for structural engineers. Interviews from other industries with concrete structures also spoke about the usefulness of being able to easily track and visualize when a specific deviation is first observed[8].

Except for a digital twin and the possibilities with such a solution, the possibilities with drones and/or continuous monitoring were highlighted regarding the

[7] Peter Cvitanovic, Marit Täpp & Finn Midböe. 2021. *Making inspections of concrete structures in nuclear powerplants more efficient.*

[8] Interview with AFRY personnel working with large concrete structures.

potential to e.g., faster detect corrosion. Furthermore, the potential utilization of a drone that scans the room in conjunction with a digital twin could enhance inspection efficiency. This approach would allow the inspector to allocate more time to comprehensive documentation and reduce the effort spent locating specific deviations.

4.12 DESIRED SHORT-TERM WORK APPROACH

When asking the interviewees about short-term desires that would simplify work, many different ideas were highlighted. However, they were all in some way connected to digitalisation. Specifically, ways to improve the inspections were emphasised, e.g., with the usage of tablets where the pictures and information could be uploaded straight away. Another idea, partially connected to the usage of tablets, was to develop an app made specifically for the collection of this type of data. This would need to be more thoroughly investigated regarding the possibilities of offline usage, due to security issues, as most plants do not have Wi-fi.

Furthermore, there was a discussion about the possibilities of using drones, especially as there are already drones in use in other types of industries. The usage of drones to conduct parts of the inspection could be a more secure and efficient way of finding deviations in the concrete. At Ringhals drones have been used to investigate concrete within enclosure/containment structures, as the tall walls complicate access. Drones have also been used to some extent at OKG, with different success.

One of the problems with drones is the extensive amount of time needed for security reviews. An interviewee[9] mentioned that there might exist robots that are used in reactor pools for investigation, however, the person continued, that these are not part of the building segment. Usually, other parts of the organisation are more open-minded about using "new" technology as a tool than the construction department within the company, according to this interviewee.

4.13 DESIRED LONG-TERM WORK APPROACH

Due to the interviewees coming from different parts of the organisations, there are different types of long-term desires. Connected to digitalisation an interviewee spoke about the wish for a coordinate-connected camera system, that would automatically save the specific coordinates of an identified deviation at the plant. Another desire was to have a system where all cracks and deviations could be monitored live. The underlying rationale is that it could facilitate the identification of defective rooms and potentially reveal patterns among specific defects. Consequently, this approach would enhance the ability to earlier detect areas of the plants that require renovation or re-building. A continuous monitoring system already partly exists at Olkiluoto in Finland, therefore, it is safe to assume there is a possibility of implementation also in Swedish plants.

[9] From AFRY

The interviewees also mentioned a wish for access to available tools that can make the work more efficient in the long term. Another long-term work approach is the potential usage of trend analysis to facilitate earlier detection of deteriorations at the plants. The desired function is for a system to monitor trends, and if a deterioration trend is detected at a specific place the system would send out a warning. This proactive approach facilitates early detection and rectification of deviations, thereby minimizing renovation expenses. It allows for timely intervention before issues escalate. At Olkiluoto in Finland, the interviewee spoke about continuous measurements from sensors, including cathodic systems and moisture measurements spanning 20 years, this is an approach Swedish power plants should take after.

5 Identified Solutions and Technologies

In the following chapter, methods of analysis found in the external and internal experience retrospective are presented. In total 14 different people from AFRY, Trafikverket and Öresundsbron have been interviewed to get an insight into potential solutions. The presented solutions could be viable for the power plants and are evaluated in Section 6.

There are several challenges to be solved for this project as presented in chapter 5. One part is accessing old data, making it searchable, traceable and available. The other part concerns handling future data reducing manual labour and facilitating analysis. This section firstly discusses old data by talking about NLP and ETL and methods for analysis. Secondly it discusses handling of future data and manual labour by talking about different management tools or models such as BIM, 3D modelling, and tools currently in use in other industries.

5.1 NATURAL LANGUAGE PROCESSING

Natural language processing (NLP) is a subfield of AI that focuses on human language such as text data, human speech etc. NLP can be used for chatbots, sentiment analysis, language translation, and much more. Since 2017, NLP has taken a giant leap due to the development of attention-based methods such as transformers. Since the release of ChatGPT in November 2022, the interest and use of NLP-based solutions has increased dramatically. In November 2023, OPENAI, the owner of ChatGPT, had 2 million developers building on their API for a wide variety of use cases. Over 92% of Fortune 500 companies were building on their products and Chat GPT had about 100 million weekly users.[10]

Solutions including Natural Language Processing have become increasingly popular in recent years. This is due to the technological advances within NLP as well as the opportunity to reduce manual tasks such as searching, extracting and analysing information from text data such as reports. The time saved can be used to control the results of the models as well as for other tasks.

An AI team[11] has experience from several projects of creating NLP solutions using ChatGPT with similarities to this project. To implement the solution, the pre-trained model runs through the chosen data at the company to learn and can then be used like ChatGPT as an assistant, to gain insight and knowledge about the data. When asked, the model can for example point to the data origin, presenting document and page, making the data more searchable. The model can handle text as well as image data.

ChatGPT and Azure can and have been used in projects to collect and compile qualitative data efficiently from the reports. The companies included in the project that chose to implement AI in such a way saved a lot of time, earlier having been

[10] Alex Heath, "*All the news from OpenAI's first developer conference*", The Verge, last updated 2023-11-19, retrieved 2024-01-10, https://www.theverge.com/2023/11/6/23948619/openai-chatgpt-devday-developer-conference-news

[11] From AFRY

spent on handling the reports. Therefore, time was freed up to be used on other tasks instead. Moreover, trend analysis of priorly non-analysed data was facilitated by using AI to first compile the data.

The model stores data in Azure[12]. However, the level of security of Azure or any non-private cloud solutions is most likely not sufficient for the nuclear industry. Furthermore, it's possible to create an on-premises solution by implementing an open-source large language model. Such a solution is however a larger and more complex project and there is less experience of it in the industry since cloud solutions are more common and the technology is still developing fast.

5.1.1 Reference case 1 – Chatbot for reducing time

For a company's[13] geotechnical department, extracting required data and generating reports based on the project's data and documentation was a tedious and inaccurate process. It was a hurdle to navigate through lengthy documents in different formats, making the process challenging and inefficient to extract the testing area's geological data and generate reports. A team with relevant domain expertise, conducted a preliminary study, and developed a tailored proof-of-concept solution for a chatbot tool, based on ChatGPT, aimed at reducing the time needed for generating highly accurate reports and extracting data. The customer has indicated that the solution is promising and has shown great enthusiasm.

5.1.2 Reference case 2 – Chatbot in MS Teams for generating accurate offers

For a customer within interior design, finding information and generating offers based on the supplier's documents was a tedious and inaccurate process. The customer needed enhanced accuracy in both product data extraction and offer generation. A team with appropriate domain expertise performed a pre-study followed by a specified solution for a chatbot tool in Teams, targeting to minimize the time required for generating highly accurate offers. A ChatGPT-based solution was implemented, currently under validation period by the customer.

5.1.3 Reference case 3 – Chatbot for chatting with documents

In a project regarding environmental impact assessments on Natura 2000 areas, it was a time-consuming task to read through all related documentation describing the Natura 2000 site and the infrastructure project, thereafter, extract text and then populate a report. The team provided a proof-of-concept solution with a co-pilot web interface where questions can be asked against the documents. The customer can use the co-pilot to assess the quality of the answers for analysis and further development. The customer has indicated that the solution is promising and has shown great enthusiasm.

[12] By Microsoft
[13] AFRY

5.1.4 Reference Case 4 – Virtual assistant "AtomAssist"[14] -

A company[15] has created the world's first private, nuclear-specific generative AI model. The customers of the solutions can choose whether to use deployment from the company, deploy the model on-premise or in a private cloud. The product AtomAssist is a virtual assistant trained on 500 000 000 nuclear training texts specifically to aid workers at nuclear facilities in various day-to-day tasks. It can for example be used for question and answering on specific documents from the nuclear facilities, i.e. chatting with your documents. Moreover, AtomAssist can help with streamlining and automating mundane and routine tasks. The products from the company can furthermore be integrated with PowerBI.

5.2 THE PROCESS OF TRANSFORMING DATA INTO INFORMATION

The inspections and measurements of concrete structures has resulted in vast amounts of data. For the business to make data-driven decisions and to analyse and discover trends and patterns, the data needs to be transformed into information. The information should be gathered, collected and structured in a data repository. This is enabled by the process of ETL (Extract Transform Load), which will be presented more thoroughly in the following section.

5.2.1 ETL – Extract – Transform – Load[16]

Extract, transform, and load (ETL) is the process of combining data from multiple sources into a large, central repository called a data repository, thus transforming data into information. ETL uses a set of business rules to clean and organize raw data and prepare it for storage, data analytics, and machine learning (ML).

Businesses have information and decision support requirements, which in this case is data about aging concrete, generated from various inspections and analyses. The data is usually embedded in reports or attached to specific spreadsheets or data files. To combine the data and transform it into analytic models, the data first needs to be retrieved. In order to conduct and produce a qualitative analysis the ETL process is inevitable. The process is to gather and collect the data into a repository i.e., a data warehouse or data hub. The data is thereafter transformed, prepared, and cleaned to be available for business use, including analysis and report production. Giving the ability to combine, cross-analyse and trend the data in multiple dimensions.

The three steps of ETL must be performed to produce a useful analysis of the data. In the following section, these three steps will be explained.

[14] NuclearN, "AtomAssist", *NuclearN,* retrieved 2023-03-19, https://nuclearn.ai/our-product/nuclearn-atomassist/

[15] Nuclearn.ai

[16] Joe Caserta and Ralph Kimball, "*The Data Warehouse ETL Toolkit, Practical techniques for extracting, cleaning, conforming and delivering data",* (Indianapolis: Wiley Publishing Inc, 2004), https://nibmehub.com/opac-service/pdf/read/The%20Data%20Warehouse%20ETL%20Toolkit%20_%20Practical%20Techniques%20for%20Extracting-%20Cleaning-.pdf (retrieved 2024-04-09)

Extract

The first phase of ETL, the *Extract* phase, is when data is collected from one or more data sources. Examples of data sources include an enterprise resource planning (ERP) platform; maintenance system (CMMS), social media platform, Internet of Things (IoT) data, spreadsheets, text files and more.

During extraction, validation rules are applied. This tests whether the data meets the requirements of its destination, the test criterias could be concerning value, format, length, type and duplicates. The data that fails validation is rejected and does not proceed to the subsequent step. The rejected data is marked and available for analysis. After the extraction, the data is held in a temporary storage, where the next two steps of ETL are executed.

Transform

In the second phase, the *Transform* phase, data undergoes processing to ensure that the values and structure are adapted to suit the desired purpose of usage. The goal of transformation is to ensure all data is fit within a uniform schema before progressing to the ultimate step.

Typical transformations include a variety of operations to prepare the data for further use. Some of the typical transformations are aggregators, data masking, expression, joiner, filter, lookup, rank, router, union, XML, Normalizer, H2R (Hierarchical to Relational), R2H (Relational to Hierarchical) and web service.

Examples of transformation:

- Aggregators - add an age category to a single age value.
- Rank - adding the rank of the data, enabling statistical analysis.
- Normalisation - breaking data into smaller tables, minimizing duplicate information, and establishing relationships between tables for more effective analysis.

These transformations assist in normalizing, standardizing, and filtering data. This makes the data suitable for consumption in analytics, business functions, and other downstream activities.

Load

The final phase, the *Load* phase, transfers the transformed data into a permanent target system. This system could be a target database, data warehouse, data store, data hub or data lake, depending on the organization's infrastructure it could either be on-premises or cloud-based. The data integration process is finalized when all data has been fully loaded into the designated target system.

Many organizations regularly perform this process (i.e., ETL process) to keep the data warehouse up to date. Depending on the business requirements the process can be performed daily or monthly.

Tools

Various tools in the market are specialized for the three parts of the ETL process.

Gartner[17], a renowned consulting firm, occupies a central position in assessing and evaluating system vendors. Within the segment of data warehouse ETL integration, several prominent players are identified as market leaders. These are SQL server[18], PowerCenter[19], and lastly Oracle[20] with its leading position of data integration tools. Each of the vendors has databases and integration tools that are well combined and cater to diverse organizational needs. Additionally, smaller vendors[21] have stood out to also be well-prepared for data integration processes.

5.2.2 The data collection at nuclear facilities

The data collection at the nuclear facilities can be divided into two processes. The extraction of historical data and the process of collecting current and future data.

Historical data

It is important to assess the needs of historical data, as it is a time-consuming and costly process to collect the information. Although there are techniques to automate the process of retrieving and collecting data from source systems, it is still a manual process to assess, categorize and clean the raw data in order to set up validation rules and import criteria. When this manual process is completed, the historical data model can finally be built up. For relevant and critical parts of the premise, it is crucial to have access to historical data to be able to conduct time series analysis.

When the business requirement has been assessed and the decision to retrieve data is taken, the ETL process can be executed. Firstly, the extract process, where relevant measures from appended and embedded files from inspections and other analyses are retrieved. Secondly, the Transform process applies business rules and adds metadata to each data point and finally the load process where data is placed into a database or repository to form a single source of data for further visualization and analysis.

Metadata

To perform useful analysis there is also a need for metadata, data that better explains and enables categorization of the data in different aspects. Examples of metadata are location, room, coordinates (x y z), time, date, inspection type, asset ID, asset type, inspector, contractor, related work order, related inspection report tools, measurements, instruments, measurement type, Unit of measure, etc.

Future data

With a digitalization of the inspection process, the measurements and inspection data can be stored in databases that are well integrated with the analytics platform.

---

[17] Gartner, “Gartner Magic Quadrant & Critical Capabilities”, *Gartner* https://www.gartner.com/en/research/magic-quadrant, retrieved 2024-06-10

[18] By Microsoft

[19] By informatica

[20] By IBM

[21] Like Qlik

The CMMS system provides metadata of the inspection object together with detailed information on the maintenance activities performed at the plant.

Inspection data, both structured and unstructured, is stored in a database or repository. This data has a strong link to inspection objects. An ETL process transforms the data into a format more suitable for the repository, and further on layers the information for trending and analysis. Thus, the information becomes accessible for business intelligence and analytics platforms, enabling the organization to create data-driven decision support.

5.2.3 Managing data uncertainty[22]

Data uncertainty arises from various sources such as measurement errors, sampling variability, and inherent randomness in data generation processes. To effectively handle data uncertainty, it is recommended to implement a structured approach that includes data quality assurance, sensitivity analysis, Bayesian inference, and methods to identify, quantify, and reduce uncertainty.

Identifying the sources

The initial step in addressing data uncertainty is to identify the sources and types of uncertainty impacting your data.

This can be achieved through:

- Descriptive statistics
- Exploratory data analysis
- Data quality assessment
- Metadata examination
- Sensitivity analysis

A sensitivity analysis involves varying the values of certain variables in the analysis to observe how these changes affect the outcomes.

Common sources of uncertainty include random noise, systematic bias, human error, and other unknown factors. Identifying these sources allows the assessor to address specific issues impacting data accuracy, precision, completeness, or consistency.

Quantify the uncertainty

Quantifying uncertainty involves measuring how much uncertainty exists in your data and its impact on analysis results. Techniques include:

- Standard errors and confidence intervals for point estimates and statistical inferences.
- Hypothesis tests to assess the significance of findings.
- Probability distributions and Bayesian methods for estimating uncertainty in parameters, models, or predictions.

[22] LinkedIn, "How do you manage data uncertainty", *Linkedin, https://www.linkedin.com/advice/3/how-do-you-manage-data-uncertainty-skills-data-science, retrieved 2024-06-04*

- Monte Carlo simulations to model and analyse the impact of uncertainty.

Quantifying uncertainty helps in understanding how data variability influences decision-making.

Reduce the Uncertainty

Reducing uncertainty involves enhancing data quality and reliability through:

- Data cleaning and imputation to address errors, missing values, and anomalies.
- Data transformation and outlier detection to improve data consistency.
- Data integration and standardization to ensure uniformity across data sources.
- Data validation and governance to establish and maintain high-quality data standards.

These methods help mitigate the effects of uncertainty, leading to more reliable and accurate data for analysis and decision-making. These methods are part of the ETL process.

5.2.4 Synthetic data generation[23]

To learn which methods would be best suitable, the methods should be evaluated against data. When real data is not available for use, synthetic data could be generated and used.

Synthetic data generation involves creating artificial data that replicates the characteristics, structures, and statistical properties of real-world data, ensuring adherence to data privacy regulations.

Syntetic data is artificial data that is generated based on real data. This can be used when there is not enough real data or when the data can't be used for some reason. Synthetic data can be used to test software or to train models for example.

There are four common methods to synthesise the data:

1. Generative AI, which creates synthetic data using Machine learning algorithms, such as Generative Pre-trained Transformers (GPT), Variational Auto-Encoders (VAEs) or Generative Adversarial Networks (GANs)

2. A rules engine generates data based on user-defined business policies.

3. Entity cloning extracts and anonymizes business entity data before replicating it.

4. Data masking anonymizes Personally Identifiable Information (PII) and sensitive data to produce compliant, new data.

[23] K2view, "*What is Synthetic Data Generation?*", last updated 2024-05-07, https://www.k2view.com/what-is-synthetic-data-generation/#:~:text=Synthetic%20Data%20Generation%3F-,Synthetic%20data%20generation%20is%20the%20process%20of%20creating%20artificial%20data,%2C%20models%2C%20or%20other%20techniques, retrieved 2026-06-04

There exist many different companies[24] with synthetic data generation software.

5.3 METHODS FOR DATA ANALYTICS, TRENDS AND FORECASTING[25]

There are several methods of predicting the condition of the facilities. Standard statistical methods identified for prediction and forecasting will be explained in this section. These methods are:

- Linear regression and nonlinear regression
- ARIMA Auto Regression integrated moving average

The following steps are important to follow before choosing a method to gain meaningful results:

1) Collect the data and ensure that the data is not aggregated or calculated mean and average data value.
2) Identify measurement method
3) Identify metadata around each data value
4) Plot the data and inspect outliers
5) Test different models and methods

5.3.1 Linear regression method

Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. For example, a modeler might want to relate the weights of individuals to their heights using a linear regression model.

Before attempting to fit a linear model to observed data, a modeler should first determine whether there is a relationship between the variables of interest. This does not necessarily imply that one variable causes the other, but that there is some significant association between the two variables. A scatterplot can be a helpful tool in determining the strength of the relationship between two variables. If there appears to be no association between the proposed explanatory and dependent variables (i.e., the scatterplot does not indicate any increasing or decreasing trends), then fitting a linear regression model to the data probably will not provide a useful model. A valuable numerical measure of association between two variables is the correlation strength which is a value between -1 and 1 indicating the strength of the association of the observed data for the two variables.

A linear regression line has an equation of the form $Y = a + bX$, where X is the explanatory variable and Y is the dependent variable. The slope of the line is b, and a is the intercept (the value of y when $x = 0$).

[24] K2View, Gretel, Mostly AI, Syntho, YData, Hazy

[25] Hyndman, R.J., & Athanasopoulos, G. *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia, 2018, OTexts.com/fpp2, Accessed on 2024-05-02

Least-Squares Regression

The most common method for fitting a regression line is the method of least-squares. This method calculates the best-fitting line for the observed data by minimizing the sum of the squares of the vertical deviations from each data point to the line (if a point lies on the fitted line exactly, then its vertical deviation is 0). Because the deviations are first squared, then summed, there are no cancellations between positive and negative values.

This is easily done in a statistical software or as an built in capability in a Business Intelligence tool.

Before attempting to perform linear regression, the data must pass through certain required assumptions.

- The variables should be measured at a continuous level. Examples of continuous variables are time, sales, weight and test scores.
- Use a scatterplot to quickly find out if there is a linear relationship between those two variables.
- The observations should be independent of each other (that is, there should be no dependency).
- The data should have no significant outliers.
- Check for homoscedasticity — a statistical concept in which the variances along the best-fit linear-regression line remain similar all through that line.
- The residuals (errors) of the best-fit regression line follow normal distribution.

Forecasting with Linear regression

After finding the best fitting regression line for the data set, and the data meets the requirements listed above assumptions, one can estimate future behavior and predict the explanatory value when the dependent variable meet the physical threshold value.

In the simplest case, the regression model allows for a linear relationship between the forecast variable y and a single predictor variable x:

$$yt=\beta 0+\beta 1x_t+\varepsilon_t.$$

An artificial example of data from such a model is shown in Figure 5-1. The coefficients β_0 and β_1 denote the intercept and the slope of the line respectively. The intercept β_0 represents the predicted value of y when x=0. The slope β_1 represents the average predicted change in y resulting from a one unit increase in x.

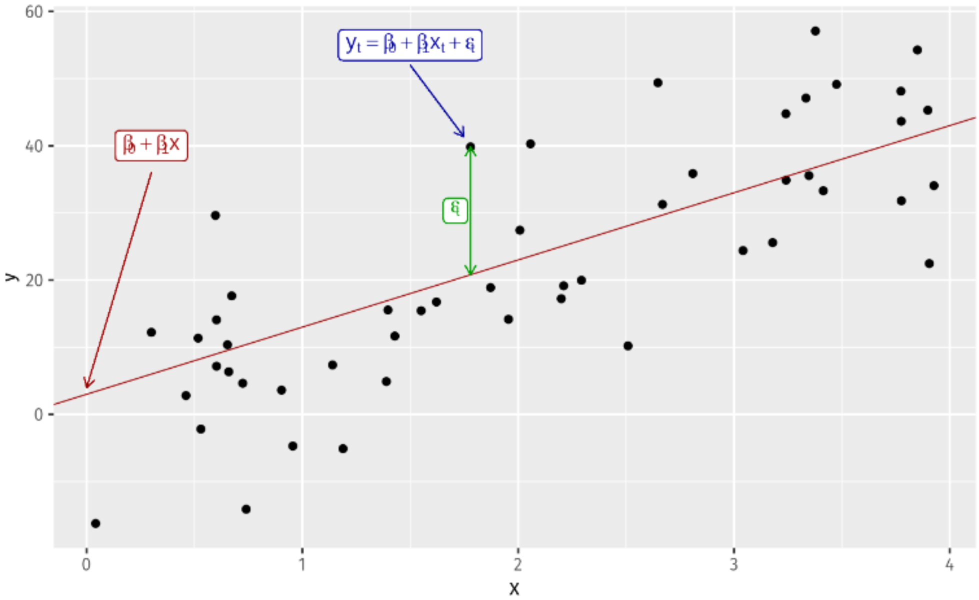


Figure 5-1: An example of data from a simple linear regression model.

Notice that the observations do not lie on the straight line but are scattered around it. We can think of each observation y_t as consisting of the systematic or explained part of the model, $\beta_0+\beta_1 x_t$, and the random "error", ε_t. The "error" term does not imply a mistake, but a deviation from the underlying straight line model. It captures anything that may affect y_t other than x_t.

5.3.2 Multiple linear regression

When there are two or more predictor variables, the model is called a multiple regression model. The general form of a multiple regression model is

$$Y_t=\beta_0+\beta_1 x_{1,t}+\beta_2\, x_{2,t}+\cdots+\beta_k x_{k,t}+\varepsilon_t,$$

where y is the variable to be forecast and $x_1,\ldots,x_k$ are the k predictor variables. Each of the predictor variables must be numerical. The coefficients $\beta_1,\ldots,\beta_k$ measure the effect of each predictor after taking into account the effects of all the other predictors in the model. Thus, the coefficients measure the marginal effects of the predictor variables.

Least squares estimation

In practice, of course, we have a collection of observations but we do not know the values of the coefficients $\beta_0,\beta_1,\ldots,\beta_k$. These need to be estimated from the data.

The least squares principle provides a way of choosing the coefficients effectively by minimising the sum of the squared errors. That is, we choose the values of $\beta_0,\beta_1,\ldots,\beta_k$ that minimize.

$$\sum\varepsilon^2=\sum\,(y_t-\beta_0-\beta_1 x_{1,t}-\beta_2 x_{2},t-\cdots-\beta_k x_{k,t})^2$$

This is called least squares estimation because it gives the least value for the sum of squared errors. Finding the best estimates of the coefficients is often called "fitting" the model to the data, or sometimes “learning” or “training” the model. The line shown in Figure 5-1 was obtained in this way.

5.3.3 Nonlinear regression

Although the linear relationship assumed is often adequate, there are many cases in which a nonlinear functional form is more suitable.

Log-log regression

The simplest way of modeling a nonlinear relationship is to transform the forecast variable y and/or the predictor variable x before estimating a regression model. While this provides a non-linear functional form, the model is still linear in the parameters. The most commonly used transformation is the (natural) logarithm. A log-log functional form is specified as

$$\log(y)=\beta_0+\beta_1\log(x)+\varepsilon$$

In this model, the slope β1 can be interpreted as an elasticity: β1 is the average percentage change in y resulting from a 1% increase in x. Other useful forms can also be specified. The log-linear form is specified by only transforming the forecast variable and the linear-log form is obtained by transforming the predictor.

Piecewise linear regression

There are cases for which simply transforming the data will not be adequate and a more general specification may be required. Then the model we use is

$$y=f(x)+\varepsilon$$

One of the simplest specifications is to make the function f piecewise linear. That is, we introduce points where the slope of f can change. These points are called knots.

Cubic regression spline

A smoother result can be obtained using piecewise cubics rather than piecewise lines. These are constrained to be continuous (they join up) and smooth (so that there are no sudden changes of direction, as we see with piecewise linear splines). In general, a cubic regression spline is written as:

$x_1=x$
$x_2=x^2$
$x_3=x^3$
$x_4=(x-c_1)^3$
$x_k=(x-c_k-3)^3$.

Cubic splines usually give a better fit to the data. However, forecasts of y become unreliable when x is outside the range of the historical data.

Example Boston Marathon
The Boston Marathon example compares different methods on a dataset of winning times in the Boston Marathon. Forecasting the coming years winning time with different linear and nonlinear models:

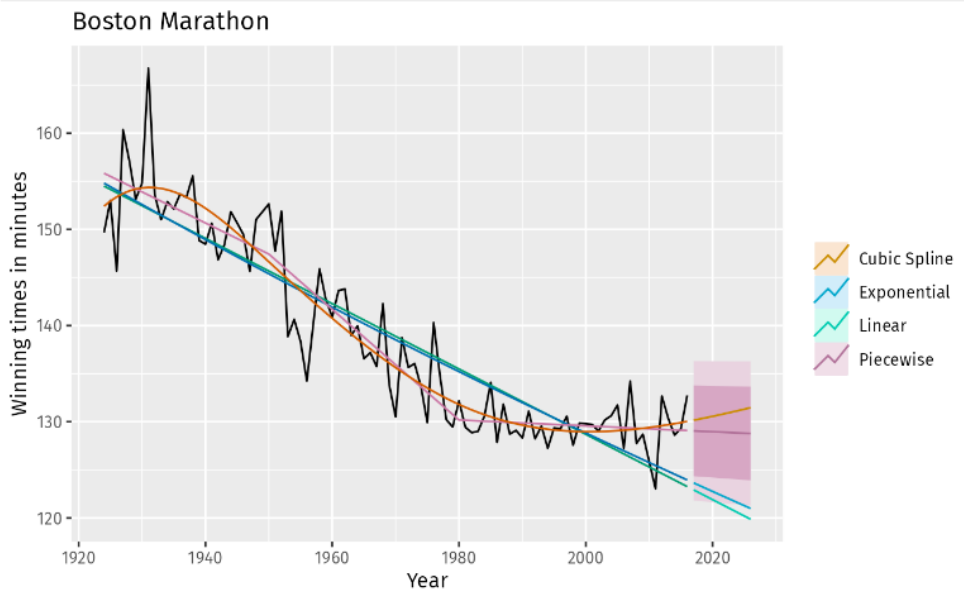


Figure 5-2: Fitted lines and forecasting with different methods for how the winning times of the Boston Marathon have changed over the years.

Figure 5-2 above shows the fitted lines and forecasts from linear, exponential, piecewise linear, and cubic spline trends. The best forecasts appear to come from the piecewise linear trend, while the cubic spline gives the best fit to the historical data but poor forecasts.

Natural cubic smoothing splines

There is an alternative formulation of cubic splines (called natural cubic smoothing splines) that imposes some constraints, so the spline function is linear at the end, which usually gives much better forecasts without compromising the fit.

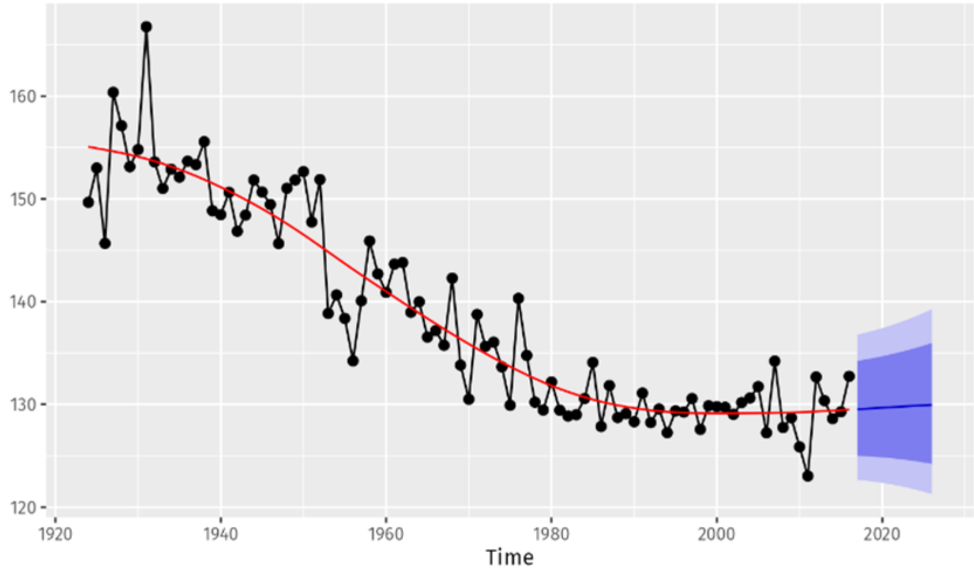


Figure 5-3: Natural cubic smoothing splines for forecasting on the Boston Marathon data.

The residuals plotted in Figure 5-3 above show that this model applied on the Boston Marathon winning data time series has captured the trend well. The wide prediction interval associated with the forecasts reflects the volatility observed in the historical winning times.

5.3.4 ARIMA

ARIMA, is an Autoregressive Integrated Moving Average, is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends. A statistical model is autoregressive if it predicts future values based on past values.

ARIMA models provide another approach to time series forecasting. Exponential smoothing and ARIMA models are the two most widely used approaches to time series forecasting and provide complementary approaches to the problem. While exponential smoothing models are based on a description of the trend and seasonality in the data, ARIMA models aim to describe the autocorrelations in the data.

The ARIMA model consists of a combination of:

- **Backshift notation** - The lag identifies the dependencies and differences between consecutive events
- **Autoregressive** - A linear combination of *past values of the variable*. The term *auto*regression indicates that it is a regression of the variable against itself.
- **Moving average** - Using past values of the forecast variable in a regression, a moving average model uses past forecast errors in a regression-like model. Thus the forecast test itself on the values.

Energiforsk

Backshift notation, differencing

The backward shift operator B is a useful notational device when working with time series lags:

$$By_t=y_{t-1}.$$

Some references use L for “lag” instead of B for “backshift”.) In other words, B, operating on y_t , has the effect of shifting the data back one period. Two applications of B shift the data back two periods. For monthly data, if we wish to consider “the same month last year,” the notation is B12. The backward shift operator is convenient for describing the process of differencing. Backshift notation is particularly useful when combining differences, as the operator can be treated using ordinary algebraic rules. In particular, terms involving backshift can be multiplied together.

Autoregressive models

In a multiple regression model, we forecast the variable of interest using a linear combination of predictors. In an autoregression model, we forecast the variable of interest using a linear combination of past values of the variable. The term autoregression indicates that it is a regression of the variable against itself.

Moving average models

Rather than using past values of the forecast variable in a regression, a moving average model uses past forecast errors in a regression-like model. We refer to this as an MA(q) model, a moving average model of order q, where q is the number of past forecasts. Notice that each value can be thought of as a weighted moving average of the past few forecast errors. However, moving average models should not be confused with the moving average smoothing. A moving average model is used for forecasting future values, while moving average smoothing is used for estimating the trend-cycle of past values.

Non-seasonal ARIMA models

If we combine differencing with autoregression and a moving average model, we obtain a non-seasonal ARIMA model. ARIMA is an acronym for AutoRegressive Integrated Moving Average (in this context, “integration” is the reverse of differencing).

5.3.5 Uncertainty in models and predictions[26]

When applying a statistical model, the aim is often to identify and describe the relationship between dependent and independent variables in the data and make predictions. These models will always contain errors, but by choosing a model that fits the data well and applying certain techniques, one can learn more about the uncertainty of the model and it’s predictions and how to reduce it.

[26] Gareth James et al., *“An Introduction to Statistical Learning, with Applications in R”,* (Springer), 2013, https://www.stat.berkeley.edu/~rabbee/s154/ISLR_First_Printing.pdf, (retrieved 2024-06-07)

Resampling: Resampling methods involve repeatedly drawing samples from a training set to refit the model. By repeatedly drawing different samples from the training data, the variability of a linear regression fit can be estimated. Two common methods are bootstrapping and cross-validation. Bootstrapping is commonly used to provide a measure of empirical accuracy for a parameter estimate. Cross-validation, such as k-fold cross-validation, can be implemented to estimate the test error of a statistical learning method, enabling the evaluation of its performance and the selection of the appropriate level of flexibility.

Feature engineering: Feature engineering involves creating new features or transforming existing ones to improve the model's predictive power. By identifying and including relevant interaction terms and polynomial features, the model can capture complex relationships within the data. Additionally, applying transformations such as logarithms can stabilize variance and address non-linearity, leading to more accurate predictions and reduced uncertainty in forecasts.[27]

Prediction intervals: Prediction intervals can be calculated to provide a range within which future observations are expected to fall. The prediction interval includes both the irreducible error and the reducible error.

Outlier detection: Identify and address outliers that may disproportionately affect the model. This can be done using residual plots or Cook's distance.

Data quality and quantity: By increasing the sample size and improving the data quality, the model will generally provide more reliable estimates.

5.3.6 Numerical methods and physical models

One approach to analyzing measurement data involves comparing the measured data with a physical model of the structure. This model can range from a simple physical representation to an advanced finite element analysis (FEM) of all or parts of a construction. In an FEM model, various physical responses—such as moisture transport, temperature distribution, shrinkage, and strength development—are considered.

However, the accuracy of the model depends on well-defined boundary conditions and material properties. If these parameters are accurately defined, an FEM model can provide valuable insights into the behavior of the construction. Measured properties, such as movements, crack widths, and temperatures, can be compared with the model. By calibrating the model against measurements, it can be refined.

Once a credible model is established, it becomes possible to predict how the design will respond to specific events. If measured values fall outside the expected range, further analysis may be warranted. Additionally, setting alarm values above or below certain measurement thresholds can help identify anomalies.

[27] Andrew Cole, "Feature Engineering, improving a Linear regression", towards data science (blog), 2020-07-01, https://towardsdatascience.com/feature-engineering-combination-polynomial-features-3caa4c77a755, (retrieved 2024-06-07)

5.3.7 Tools for data analytics and trends

An organization needs a tool capable of reporting and analytics to be able to present and analyse data. Well-known software vendors are offering analytics and business intelligence tools. The tools that are identified as market leaders in the Gartner assessment[28] should be considered. More advanced statistical models can be implemented in special statistical tools provided by applications specialized in statistical models eg Multivariable regression models etc. Other software vendors will provide applications for trend and forecasting using standard programming languages like Python and R.

5.4 TECHNOLOGIES FOR VISUALIZATION AND DATA MANAGEMENT

The interviews revealed a keen interest in enhancing orientation during inspections, with solutions such as 3D models and Digital visualizations being mentioned. The feasibility of similar solutions was investigated by interviewing personnel within a company[29] working with digital models, such as Building Information Modelling (BIM).

Creating 3D models of areas or buildings is not uncommon in the industry, and extensive experience exists. There are several different solutions which can be applicable. In one solution with 3D models, the user can easily switch between the rooms or areas. In another solution, the user experiences a more game-like situation where the user can walk from room to room in the model. Depending on the type of solution, the time and effort to create the model will differ.

Integration of automized 3D-modelling from photogrammetric data

The automated generation of 3D models from photogrammetric data amplifies the capacity of BIM by integrating detailed visual representations of deviations, such as cracks, straight into the digital model. This minimizes manual measured drawings, accelerates registration of deviations and improves the visual communication and decision support used for maintenance and reparations.

Integration of historical data from PDF reports into 3D models

Integrating information regarding existing cracks and deviations from old PDF models into newly scanned 3D models requires a process including data extraction, text and image recognition, as well as digital modelling. This approach allows for an update of the plant's digital representation to include all known defects over time.

5.5 MANAGEMENT TOOLS IN OTHER INDUSTRIES

Massive concrete structures also exist in other places, therefore, interviews have been conducted to understand how data collection and analysis

[28] Gartner, "Gartner Magic Quadrant & Critical Capabilities", *Gartner* https://www.gartner.com/en/research/magic-quadrant, retrieved 2024-06-10
[29] AFRY

works in other industries. The solutions used in other industries will be presented in the following paragraphs.

5.5.1 Trafikverket's BaTMan Tool

At Trafikverket a digital management system called BaTMan (Bridge and Tunnel Management) is in use. The inspections can be registered straight away in this system which also allows for an easy tracking of data over time. New inspections of e.g., a bridge are built upon previous inspections which enables an easy overview, as well as the possibility of trending.

A system such as BaTMan also allows for data to be searchable, which can save valuable time. The system as of today cannot log information straight away with the usage of a phone or tablet, it only works on computers. However, according to an interviewee talking about BaTMan, there seems to be plans on releasing an app, that would enable the system to be used on phones or tablets.

5.5.2 Öresundsbron – IBM Maximo

At Öresundsbron a modern CMMS (Computerized Maintenance Management system) named IBM Maximo is in use, which works as a digital model. With the help of drones, 360° pictures are taken and put into the system, where the quality of the pictures allows the user to zoom in on details if needed. The user can choose a specific part of the bridge, where the system then will show all existing data points within that chosen area. The system allows for the assets and objects to be classified and categorized in many dimensions. Information about objects and related systems and objects, location, vendor, suppliers, cost, resources etc is also maintained within the CMMS system. The integration to ERP and other systems is standard. All data around maintenance work is collected and managed. The inspections are scheduled and prepared in the CMMS system. During the inspection rounds, the observations, photos and measurements related to the inspected objects can be recorded. The information is stored in the CMMS.

Potential safety hazards that could arise with an inspection are reduced with the usage of drones and a digital model. This is due to the possibility of accessing harder-to-reach areas without having to go there in person. Moreover, a digital model can be of assistance when wanting to obtain a comprehensive overview of the entire structure. At Öresundsbron, IBM Maximo has resulted in a reduction of physical visits at the bridge, according to an interviewee. This has in turn led to reduced costs, higher security, and less frequent traffic disruptions. In addition, the availability of an overview has resulted in a better understanding of the structure as a whole, while also decreasing the organization's reliance on individuals.

See Figure 5-4 - Figure 5-8 below for user case pictures of the system, which also shows different functions within the system.

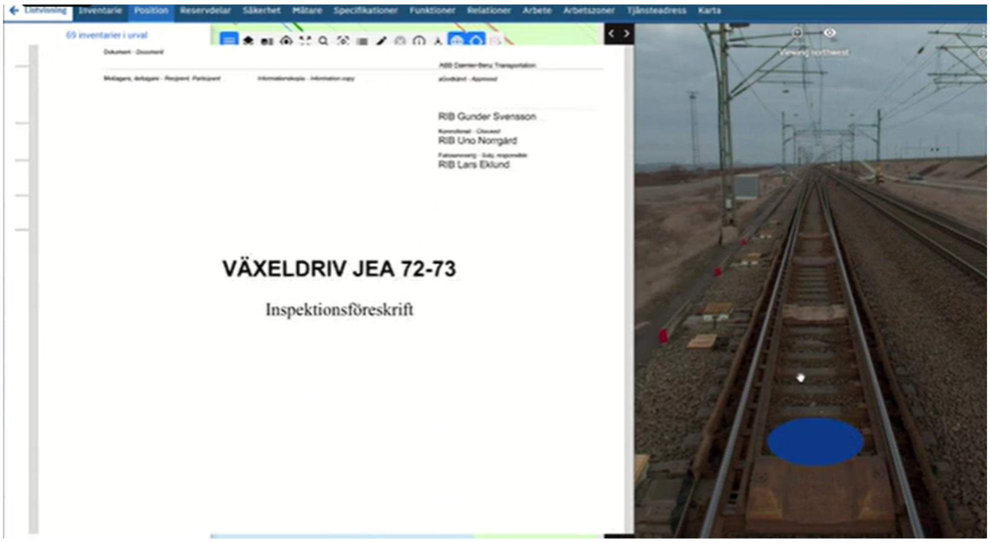


Figure 5-4: Example pictures from IBM Maximo , showing photos from the train tracks which can be seen when moving through the model

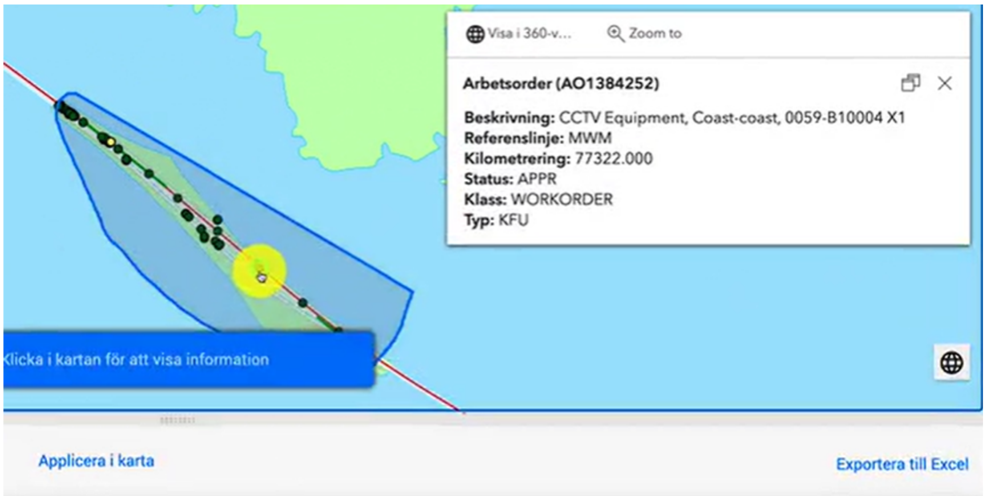


Figure 5-5: Example pictures from IBM Maximo, showing the 2D map with different inventories and their position.

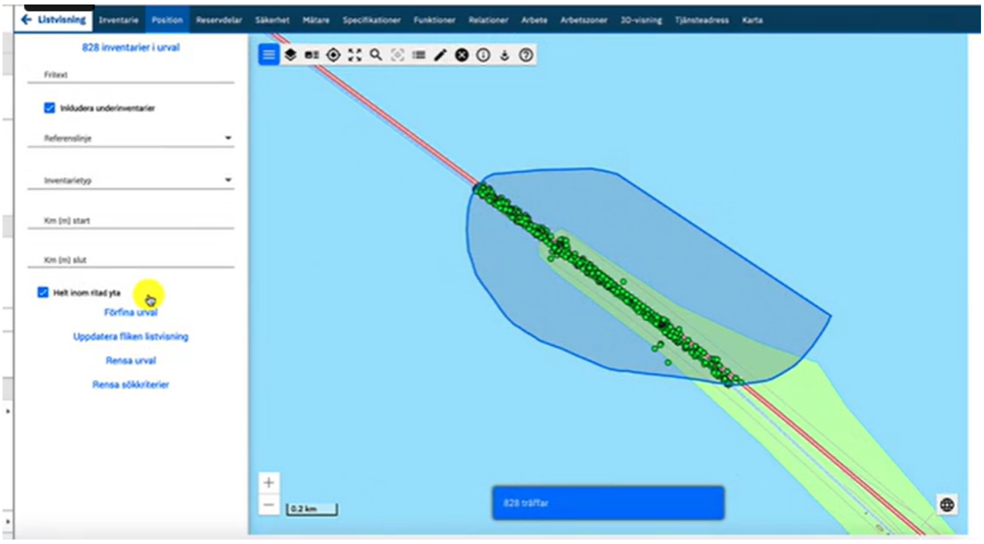


Figure 5-6: Example pictures from IBM Maximo, showing the 2D map with different inventories and their position.

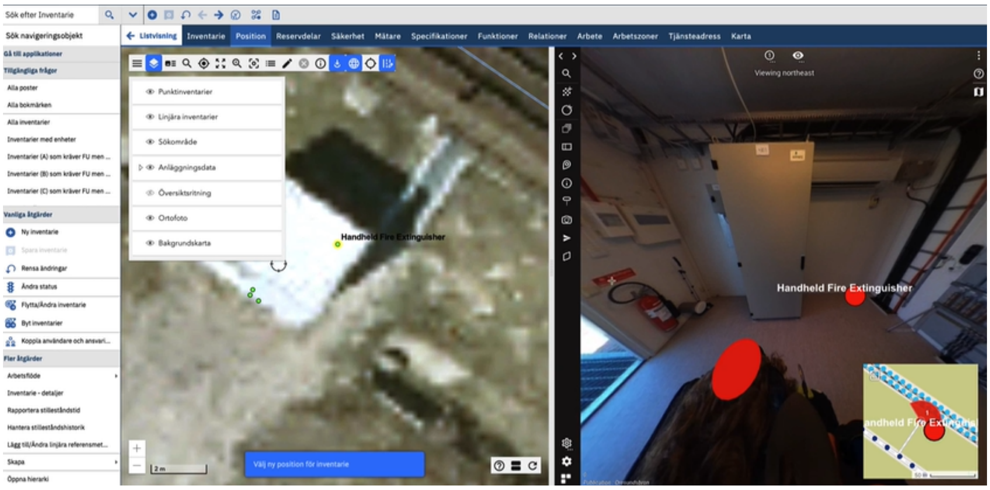


Figure 5-7: Example pictures from IBM Maximo, showing a 2D overview picture and 360 deg pictures, in a house next to the railway, pointing out inventory such as Handheld Fire Extinguisher.

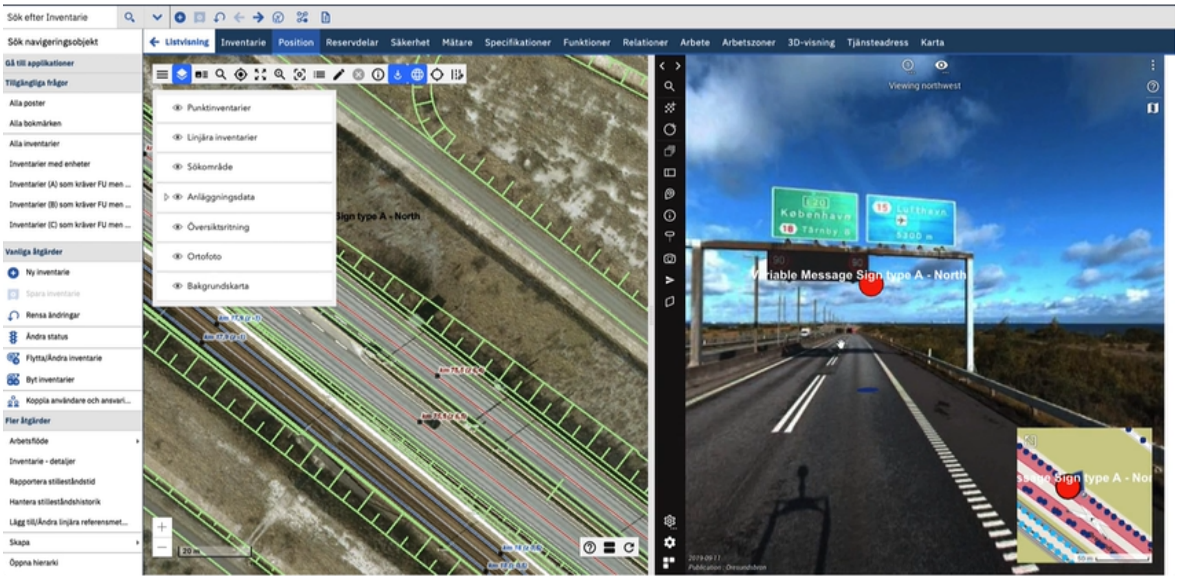


Figure 5-8: Example pictures from IBM Maximo, showing 2D maps and 360 deg pictures, in this case pointing out "Variable message signe typa A - North".[30]

5.5.3 Screening Eagle

Screening Eagle Inspect is a fully customizable inspection workflow software. The software enables the user to record observations with forms, comments, photos, sketches, 3D scans and NDT sensor data in a structured way. [31] The data and pictures can be summarized in a dashboard as can be seen in Figure 5-9

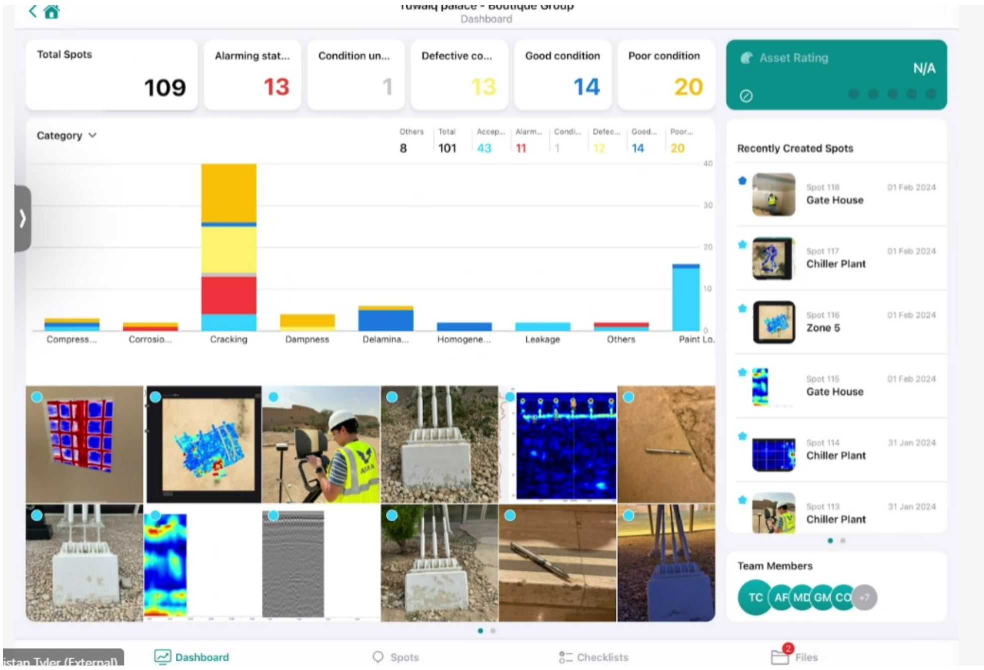


Figure 5-9: Dashboard example from SCI. It shows an overview of conditions.

[30] Screenshots from user case video of the system IBM Maximo which is in use at Öresundsbron.
[31] Screening Eagle Technologies, "*Inspect*", https://web.screeningeagle.com/home, Retrieved 2024-03-11

The software is generally used with an iPad or iPhone[32], where it can be used offline during the inspection. Locations of observations can be added on maps or 2D drawings, as can be seen in Figure 5-10. By taking photographs within the program, 3D models can be created in a simple way, and deviations such as cracks can be analysed by embedded AI software. The 3D models can be used as a base layer to add information on. The location of deviations is therefore easy to find as the information and images can be visualized in the room/plant on the maps or the 3D images. A 3D model created by using an Ipad camera, which was done during the interview of an employee at Screening Eagle can be seen in Figure 5-11. Furthermore, Sensors can also be connected through the software, automatically adding information.

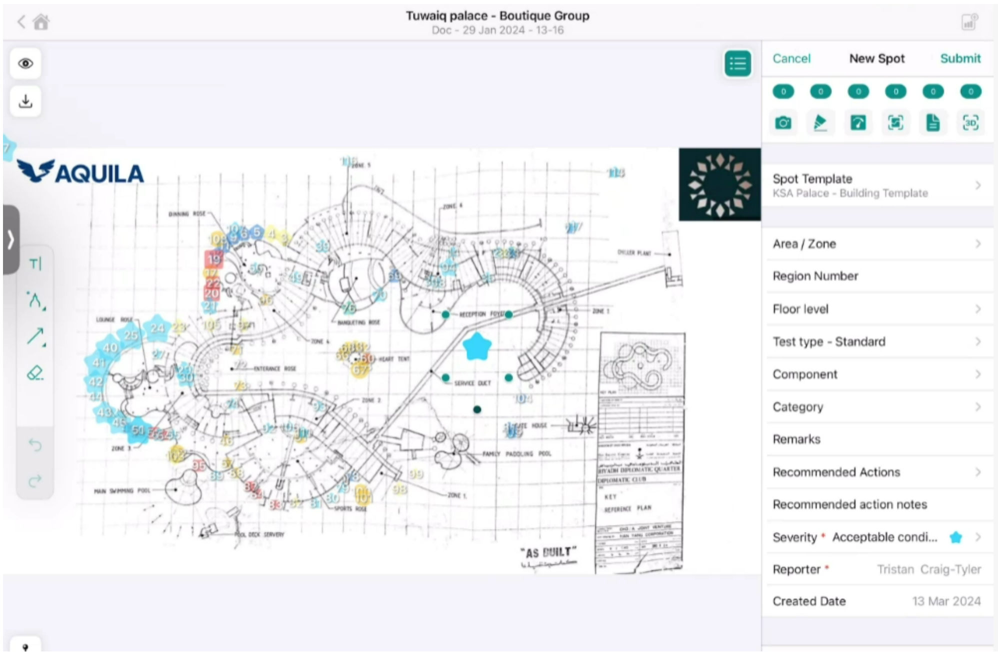


Figure 5-10: Map of example utility in Screening Eagle Inspect with observations added to the man in the shape of numbers.

[32] The app is currently available in App Store but not in Google Play Store

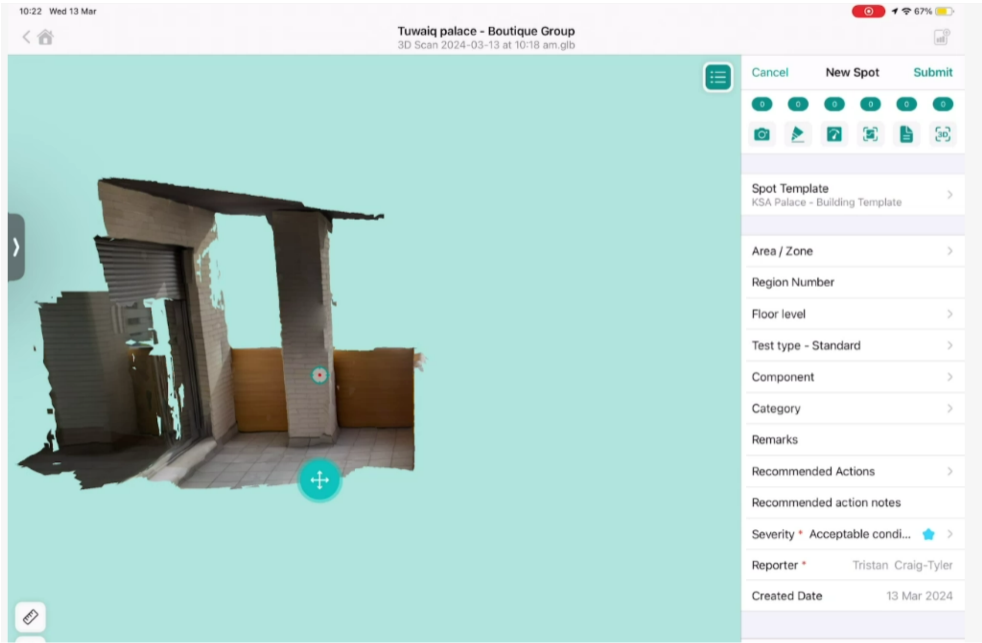


Figure 5-11: Example from Creating a 3D model of a Balcony using the iPhone/iPad camera and the tool 3D tool from SCI. Quick example created during the interview.

Furthermore, reports from the inspections can be auto-generated from the software by using templates customized by the user, and an approval workflow can be administered, see Figure 5-12. The software allows for simple filtering for certain information on previous inspections, for example, certain images or locations.

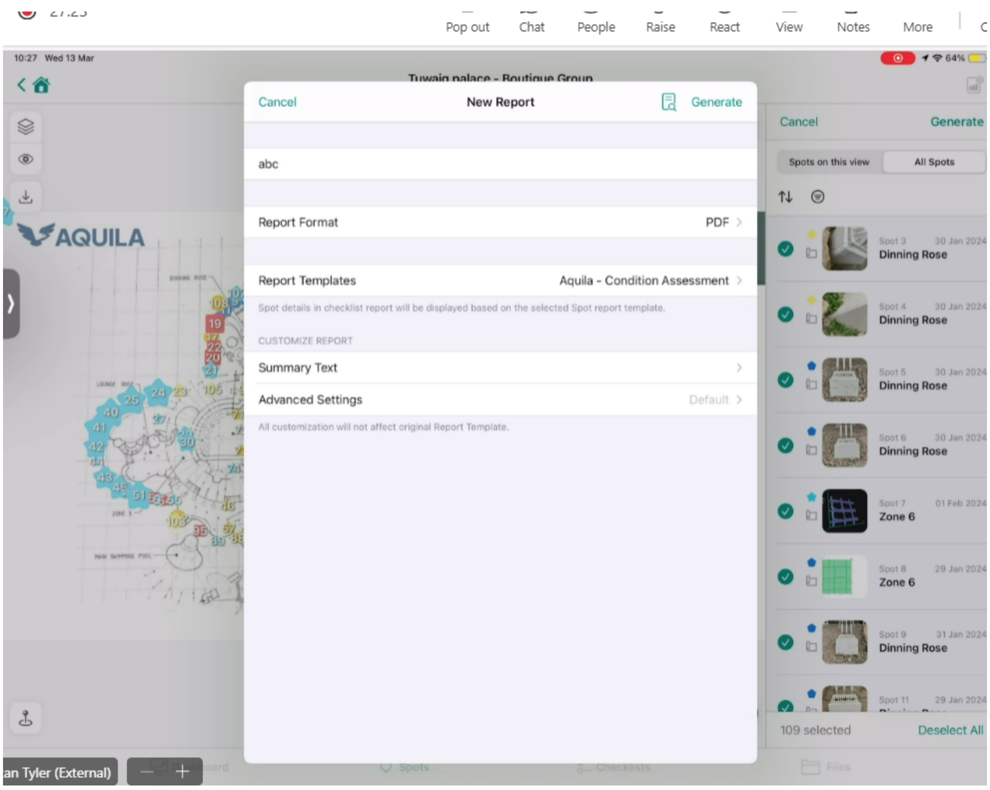


Figure 5-12: Creating a report in Screening Eagle Inspect – example.

By using an API, the program can be connected to an asset management system which is the most common way for large customers to use the software. However, the Screening Eagle Inspect software can, to some extent, also be used for asset management, which some smaller companies take advantage of. The software can also be connected to Power BI through an API, which enables data analysis to be applied and trends to be created. Furthermore, the following ideas/plans exist for the software:

- Integration with Bentley SYNCHRO
- Integration with Autodesk Construction Cloud
- Integration with IBM Maximo
- Integration with Procore
- Presence on the Workato marketplace

Screening Eagle Inspect is web-based and hosed with Amazon Web Services, using the highest possible security for cloud solutions on the market, with the possibility of hosting the data on servers in Sweden.

Screening Eagle has customers within the nuclear industry in Europe that use the solutions connected to Amazon Web Services. This has been made possible through collaboration on how to implement the solution with maximum security and comply with all regulations. Furthermore, one of Screening Eagle's customers is requesting an on-prem solution. According to the interviewee from Screening Eagle, this is possible, however, it is not recommended as it is a very expensive solution.

The interviewee[33] working with concrete had previous experience working with Screening Eagle in various segments such as road, railway, tunnels, hydropower and housing to aid in structuring the data from visual inspections and has been satisfied with the software. The software aided in many ways, such as:

- Organizing the information of each spot efficiently, including time stamps.
- Data processing (i.e.reports) as a software function, thereby enhancing the reporting and documentation part.
- Opportunity to define any inspection and reporting feature from scratch.
- Exporting in fixed versions (pdf) or in workable versions (csv, docx)

[33] From AFRY

5.6 OTHER TECHNOLOGIES

As concrete structures are important in many other industries too, there are also other types of solutions that the project group has briefly looked into. These can be seen as possibilities that the nuclear industry could also benefit from.

5.6.1 Usage of drones to create a 3D model of dams in Sweden

Dams in Sweden have utilized drones to scan the entire structures using photogrammetry. As a result, cracks can be visualized and analysed in a 3D model. There is manual work needed to mark all the cracks, which is time-consuming, however, in the end, the whole structure and its deviations are easier to grasp and analyse. Moreover, as the structure is massive it saves time and energy to not need to climb around and manually look for cracks at the spot. Instead, bigger deviations can be pinpointed and further investigated if there is an identified need for this.

5.6.2 AI-recognition for cracks and deviations in concrete

At the beginning of 2023, Kiwa[34] released an article presenting a new AI technology where drone pictures can be used to follow up on defects in the concrete as well as categorizing these. The technology is supposed to improve the inspection quality, save time and create better control and overview of the condition of the concrete. The software works through an automated image analysis process with the help of machine learning. Moreover, within the 3D image that is created one can choose a specific deviation and, in the model, see the size of it. The usage of such software facilitates the analysis of the concrete, as small deviations easier can be identified, due to the high accuracy.

[34] Kiwa, *Världens första AI-igenkänning för stora betongstrukturer i 3D*, *Kiwa*, last updated 2023-01-30, https://www.kiwa.com/se/sv/media/nyheter/varldens-forsta-ai-igenkanning/, retrieved 2024-02-02

6 Opportunity analysis and evaluation of technologies and solutions

6.1 NATURAL LANGUAGE PROCESSING

For the nuclear industry, which has an abundance of reports which are hard to extract information from and trend, NLP could be an excellent solution. The model suggested, called a Large language model, has similar capabilities to ChatGPT. It is used as a chatbot where the user enters queries and gets responses from the model, based on the documents it has access to. The model would be able to:

- Improve searchability - Look at all the documents and systems it has access to and give you information about the documents based on your query. Such as asking for documents from a certain date, inspection room, anomaly, etc. The model will also tell where the data is to be found.
- Summarize - The model can summarize important reports or sections.
- Trend – Possibility to find different trends.
- Assist - When entering a room for inspection, the assistant could help the inspector by informing them about previous deviations found in the room.

How well an NLP solution can satisfy the needs is largely dependent on the input data from e.g., previous inspections. The grading system that the facilities have (1-5, 0-4 etcetera) for each deviation found can aid in trending. Where unique IDs exist for a room or structure over time, it is rather easy to create trends. If no unique IDs exist, it should still be possible to trend by analysing the text. Combining the model with a map/image of the rooms of the inspection could enable the solution to point out or map to the crack or deviation.

AtomAI

This tool is not yet available in Swedish, therefore, the tool might not yet be able to be implemented in the Nordic nuclear facilities. The company has not yet expanded its business outside of the US, and the models are thus in their current state only applicable in nuclear plants with English as the written language. In discussions, the company indicated that they are currently investigating the market potential outside of the US. As it is a start-up created in 2021, it is reasonable to assume that its current main focus is the American market. Due to the potential of this type of application, it might be of interest to further investigate whether or not this could be a suitable solution for nuclear facilities and learn more about the company’s products.

6.2 THE PROCESS OF TRANSFORMING DATA INTO INFORMATION

Trending, time series analysis and other statistical models, required to predict the future, rely on quality historical data. The nuclear facilities need to collect, store and organize the data in a data repository and establish ETL processes to retrieve, clean and store the historical data. In the data repository, the measurements and observations are extracted into data that is cleaned, mapped, categorized, indexed and organized in a form that together with current data enables analysis.

6.2.1 Extracting data from current sources and new data sources

It is essential to establish processes to collect measurements and inspection data in a structured way. The nuclear facilities need to identify information areas and define key performance indicators. The KPIs need to be analysed and monitored. Measures and models that the regular inspections and technical department, internal and external stakeholders follow up today need to be identified and assessed. In the assessment, the nuclear facilities should also identify data relevance and importance, the quality of data and metadata, and identify how far back in time historical data need to be retrieved.

The process would be able to:

- Extract historical data and metadata
- Retrieve current and future data.
- Assure data quality
- Handle changes in the system, which will create new or changed sources to be integrated
- Provide a data model for Trending and Analysis,

Thus, this gives the opportunity to:

- Provide the business with one single source of information
- Improve availability – the information available in a user-friendly BI tool, see Figure 6-1
- Share information – analysis easily shared and reproduced
- Improve searchability – metadata together with visualisation and BI tools enables search of inspection results in time and space etc
- Improve traceability – robust ETL process with transparent transformation of data to information provides traceability.
- Trend – possibility to find deviations, variations, trends and patterns, see Figure 6-3
- Improve forecasting and simulation based on statistical models, see Figure 6-3
- Establish alarm limits values and model constraints
- Visualise data see examples below

Examples of visualizations:

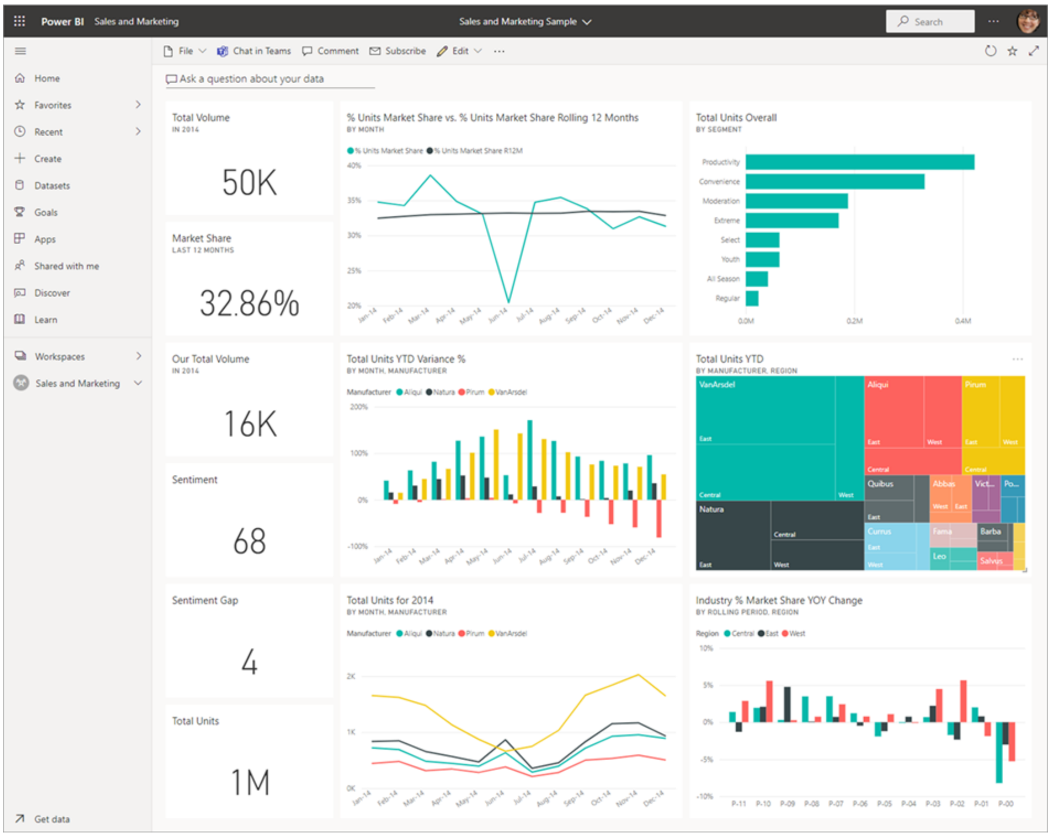


Figure 6-1: Power BI dashboard, showing an analysis of company data, such as total units, market share, total volume, industry % Market share YOY change and more.

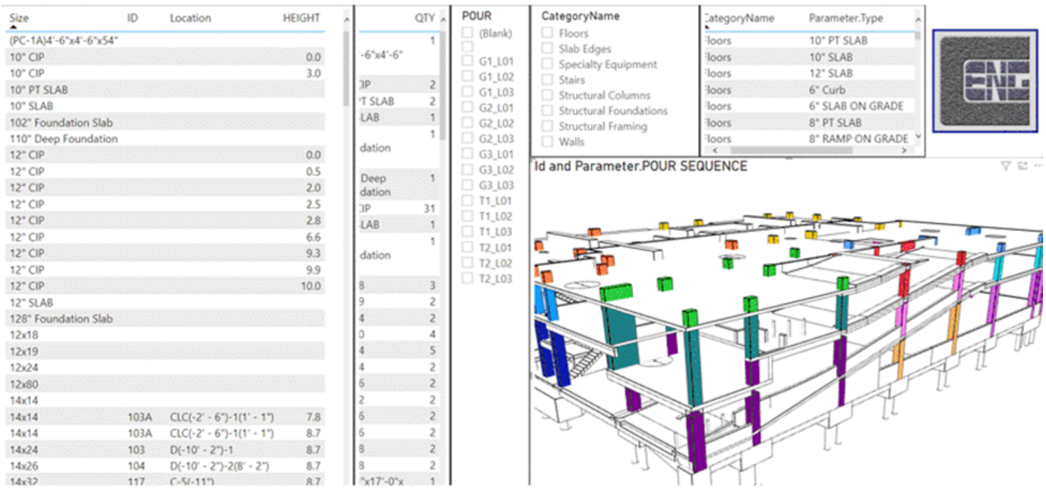


Figure 6-2: Power BI 3D visualization.

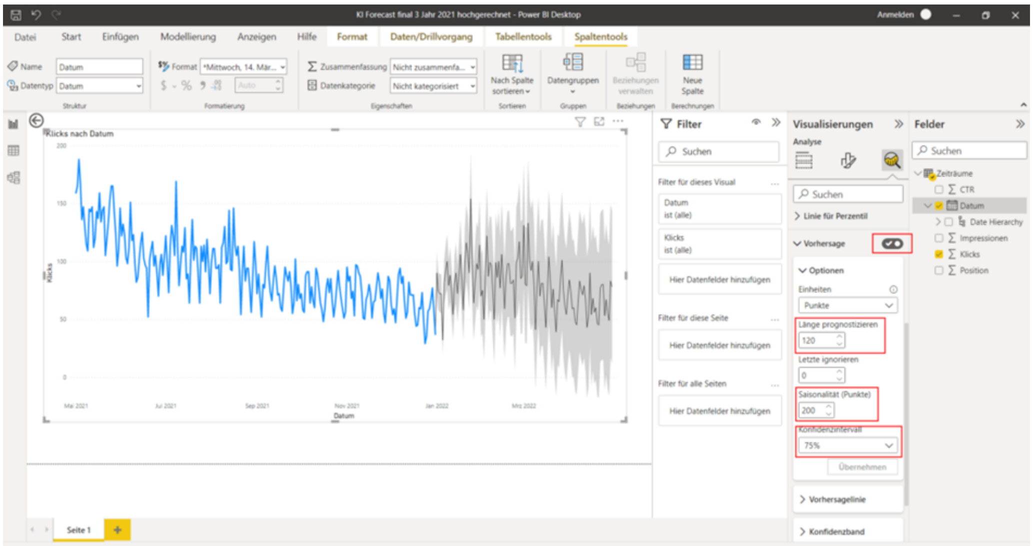


Figure 6-3: Power BI trending and forecasting capability.

Regression and forecasting capabilities are built-in functions in many standard BI platforms. More advanced models can be implemented with specialized tools or add-ons. However the quality and traceability of ingoing data needs to be established.

6.3 THE METODS FOR DATA ANALYTICS, TRENDS AND FORECASTING

There exist several statistical methods for trending and forecasting data, which could be applied to concrete inspection data and measurement data. A selection of methods are described in Chapter 5.3.

The following steps are important to follow before choosing a method to gain meaningful results:

1) Collect the data and ensure that the data is not aggregated or calculated mean and average data value.
2) Identify the measurement method
3) Identify metadata around each data value
4) Plot the data and inspect outliers
5) Test different models and methods

When a suitable method is found, it could be used for current and future analysis.

Without deeper insight about what datapoints and meta data exist for each measurement type, it is hard to say which methods could be used for which measurement type and what the results would be. One earlier study by Energiforsk[35] has identified data sources from measurements but every data point needs to be extracted to continue the work of recommending methods.

[35] Peter Lundqvist and Nick Zaraei, Compilation of Concrete Test Data from The Nuclear Industry, Energiforsk 2021:761, https://energiforsk.se/media/29656/compilation-of-concrete-test-data-from-the-nuclear-industry-energiforskrapport-2021-761.pdf, retrieved 2024-04-26.

To effectively analyse data from concrete structures and identify variations, trends and make predictions, it is essential to have time series data from the same structure and measurement points. Ideally, this data should include a substantial number of historical measurements. In cases where time series data is unavailable, which is often the case as measurements are made from different locations, one could compare nearly identical concrete structures (same exposure environment, original quality of concrete, etc.) to detect deviations or anomalies and make trends and predictions. However, this approach is precarious, and it is essential that the parameters or specifications of the concrete are very similar, that the same measurement method is used and that the external conditions and variables can be isolated/locked, such as exposure environment and original quality of the concrete. If this is the case, creating time series data based on these measurements from different locations and analysing that data could be possible.

6.3.1 Metod of analysis for different measurements

Carbonization: Given the data, a regression model as specified in section 5.3 could be applied to try and estimate when a critical level will be reached. As the carbonization speed gradually decreases as a diffusion process as

$$x = k_{carb} * \sqrt{t}\,,$$

the carbonization speed can be described as a straight line with a slope coefficient 0,5 by taking the logarithmic of the equation. Using this relationship, log-log regression can be applied to the data.

Chloride concentration: The depth of a certain chloride concentration behaves according to a diffusion process and probably follows

$$x = k_{chloride} * \sqrt{t}\,.$$

Given this, a good approach is to use a log-log regression. An appropriate concentration for chloride to monitor can for example be how a hypothetical threshold value (when corrosion becomes possible) moves into the concrete. There is no true threshold value, but 0.4% Cl by weight of cement is often used.

Compressive strength: Similar to chloride concentration and carbonization, a reasonable approach is to assume that the compressive strength behaves proportionally to $\sqrt{t}$, or a similar declining relationship. Therefore, a good approach is to try a log-log regression.

Tendon measurements: As prestress losses in tendons follow a linear-log function, for the logarithm of time, a good approach is to apply log-log regression.

Visual inspection: For visual inspection data, graded on a scale from 1-5, it is also preferred to identify deviations with similar conditions and external factors by comparing metadata. By combining them into a dataset and training a model, preferably testing several methods proposed in 5.3, predictions of degradation could be made. Different regression models could be tried to see how they fit the data.

Crack lengths from inspections: When analyzing crack lengths and how they change over time, it's important that the data quality is good, as different inspectors might not have the same level of thoroughness. Preferably photogrammetry can be used and then a regression model could be applied. As several variables affect the crack, a multivariate regression model would probably be suitable. A numerical model could also be created but it is important to know the reasons behind the crack and the boundary conditions of the structure.

Crack width: To predict crack widths, a numerical model could be created. Similar to modeling crack lengths, knowing the reasons behind the crack and the boundary conditions of the structure are important. Furthermore, no statistical model connected to physical relationships can be recommended in this study. Different regression models can be applied to the data to try and find a method that fits the data.

Moisture measurements: Similar to the analysis of crack widths, knowledge of boundary conditions is needed to be able to create a numerical model to correspond to the physical properties. Different regression models can be applied to data to try and find a model that fits the data well.

Analysis of other measurements: For measurements of strains and movements in the reactor container and temperatures, it is not straightforward how to forecast and trend. More studies would be needed to be able to recommend appropriate methods as the physical relationships between variables are not as easily described as in the cases mentioned above. Given the data, it would be easier to find a suitable statistical model. Different regression models can be tested in most cases to try to find trends and relations, however, it's hard to connect these measurements to physical relationships.

6.4 TECHNOLOGIES FOR VISUALIZATION AND DATA MANAGEMENT

In chapter 5.4 BIM was presented, which is a type of model that could be used to better visualize and analyse concrete data at the nuclear facilities. The following paragraphs will briefly present how this tool could be used.

The advantages of utilizing BIM for managing inspection data at nuclear facilities

Building Information Modelling (BIM) can be a game-changer within infrastructure projects, especially in the nuclear sector, as it makes planning, design, conduction, and maintenance more efficient. BIM allows the integration of inspection data such as historical data, pictures, descriptions and grading of ocular lesions into a digital model of the nuclear site. This enhances data handling, traceability, and maintenance planning, as well as ensures efficient communication and handling of risks between all stakeholders.

Digital site visits within areas of high-risk

A digital model allows for virtual site visits within the nuclear sites, which can be especially valuable within areas that have high degrees of radiation. This type of technology reduces the need for physical inspection as well as contributes to more efficient and secure maintenance planning.

Strict cybersecurity requirements

Due to the strict security requirements at nuclear facilities, cybersecurity is important. All data and models must be handled and stored within secure environments to be safe from cyber threats and secure the plant's defence and operational efficiency.

Conclusion

A combination of BIM, digital models and programs that automatically generate 3D models from photogrammetric data can ensure an extensive improvement in the handling of inspection data at nuclear facilities. These technologies could improve the precision, accuracy and efficiency of the data modelling, strengthening the security, and facilitating maintenance planning and decisions, which in turn leads to more secure and more durable facilities.

6.5 MANAGEMENT TOOLS IN OTHER INDUSTRIES

Chapter 5.5 presented existing management tools in other industries. In the following paragraphs these solutions will be analysed and presented regarding their potential usefulness within the nuclear industry.

6.5.1 Trafikverket's Batman Tool

BaTMan was created to be used for bridges, tunnels, and other structures. However, there are still potentials with the tool that could be useful at nuclear facilities. There are differences between a bridge and a nuclear facility regarding the definition of each element. Bridges have easily distinguishable elements that can be sorted into groups, as a bridge is not as big and complex as a powerplant. For a system such as BaTMan to be used at a power plant, all different types of elements therefore would need to be defined or numbered in an easy, comprehensible way.

The Energiforsk report from 2021 also investigated BaTMan, though not with trend analysis and predictions in focus. It concluded that *"Even though the system is becoming slightly outdated, it is noteworthy that on a conceptual level BaTMan is partly similar to what might be sought after."*[36] The project group comes to a similar conclusion that it is a slightly outdated, but well-working program, within its current area of usage. Therefore, there is value in investigating the potential of creating a similar program, specifically constructed to be used in the nuclear industry.

The main issue that needs to be considered if creating a new system, that BaTMan currently lacks, is the security level vital for usage within the nuclear industry. This aligns with conclusions from the 2021 report: "*It is concluded that the system is not built with the purpose of being private and secure, as many actors need access to the information. […] Moreover, although the system is well suited for documentation and tracking of a large number of structures and points of damage, it might not be optimal for*

[36] Peter Cvitanovic, Marit Täpp & F. Midböe. 2021. *Making inspections of concrete structures in nuclear powerplants more efficient.* Energiforsk.

nuclear power plants because of the system's inherent openness." Since then, the system has been updated to improve the level of security, according to an interviewee, who continues to explain that BaTMan is not encrypted because many people need access and find the information useful. However, sharing information is not allowed without proper authorization. To gain access, one must have completed specific courses. Users can only view their own sections. Access to other sections must be assigned, and approval is required for all parts. There are different levels of permissions, allowing some users to add content while others are only allowed to read information.

Similar to inspections at the nuclear facilities, deviations are graded on a scale of 1-5, and photos are taken. Within BaTMan the information can then be used to do trend analysis on the severity of a crack, the number of cracks at a specific place or other such data. There are both pictures and trend-graphs connected to each element put in the system, these different things can also be used to filter which allows for easier analysis. These are applications that align with what this project is investigating, which strengthens the conclusion of the tool's usefulness. When BaTMan was implemented in 2005 historical data was imported into the program. This study has not investigated the specifics of how this was done, but according to an interviewee, it was a partly automatic process. As there is a lot of historical data stored at the nuclear facilities, which is currently not analysed, it could be of interest to further investigate the details of how historical data was imported into BaTMan.

6.5.2 Öresundsbron – IBM Maximo

IBM Maximo has the potential to support the inspection process at the nuclear facilities. Together with a 3D model the assets and inspection objects are searchable and visualized, which is a problem having been highlighted by the interviewees. Moreover, follow-up work orders can be initiated and prepared with the information from the inspection. Today there are at least two different systems in use at each nuclear facility; one maintenance program and one document program. A CMMS such as IBM Maximo would enable all information to be stored in only one place, where the work orders also can be planned and scheduled.

Inspections and preventive maintenance should always be the subject of continuous improvements. The usage of drones and AI technology to scan the entire concrete structure at nuclear sites could facilitate a faster and more efficient inspection process. This would allow the inspector to allocate more time towards analysing the data collected during the inspection.

Trends and analysis

The Maximo database together with 360 Visualization tools provides measures and metadata around inspection objects. Modern analytic tools can use standardized APIs to retrieve data from the Maximo database. An analytic tool or BI platform gives users capabilities for collecting, analysing, visualizing, and sharing data. The BI platform's user interface is fairly simple, allowing users to visualize any data and seamlessly integrate the visuals into the applications they use every day.

Modern BI platforms have statistical functionalities for trending, predictive analytics, detecting variations, deviations and much more.

Even if IBM Maximo is a good program for visualization, it should be further investigated regarding the amount of security within the system. If it is concluded to not comply with security regulations in nuclear, the tools within the system can still act as an inspiration regarding how to improve the data handling at the nuclear facilities. Moreover, a BI platform would facilitate the analysis potential of collected data from inspections.

6.5.3 Screening Eagle Inspect

The Software was demoed during a meeting with the Screening Eagle team, where it was demonstrated that the system is highly customizable and well-suited to the specific requirements of nuclear facilities. By connecting the deviations found during visual inspections to a location on a map or in a 3D model, it is easy to get a well-organized history of the deviations, as well as measurements from sensors and tests such as Schmidt hammer, rebounding hammer etc. Templates can be created within the program which enables an automatic creation of reports, something that most likely will save time for the inspector. The software can filter data from previous inspections, the comparative analysis is thus facilitated which in turn will save time for the inspector.

The Screening Eagle Inspect software effectively simplifies logging and visualization of visual inspections, as well as enhances searchability. The software is recommended to be connected to an asset management system, such as SAP or similar. This is done via API. This report has not evaluated the effectiveness of that connection. The software can also be connected to PowerBI through the API to facilitate analysis.

6.6 OTHER TECHNOLOGIES

There is a value in taking inspiration from other technologies, therefore, both AI-scanning of cracks and AI-recognition from e.g., Kiwa is worth further investigation.

7 Accounting of main risks and/or potential difficulties

The different propositions in this report are mainly based on the potential long-term benefit of better handling and sorting of the data, to get better analysis of all collected data. The following chapter will briefly go over the main risks and/or potential difficulties connected to each of the earlier presented solutions.

7.1 NATURAL LANGUAGE PROCESSING

As with all NLP models, there is always a risk of depending too much on it. It is important to understand that NLP is a tool and not a substitute for people. An NLP model does not provide an absolute truth, rather it is a fairly accurate model that is to be used as a facilitating tool.

All AI models will always depend on what kind of training data it has been exposed to, therefore, there is always a risk involved with an NLP model being biased or "poisoned" which could lead to a manipulated model behaviour.

NLP models can be implemented in several different ways with different levels of security risks, depending on how the model is used. A ChatGPT-based solution can be set up through API and Azure (see sections 5.1.1, 5.1.2, 5.1.3), consequently, this solution may not be possible to use in the nuclear industry as there are strict security requirements involved. However, there exist other types of NLP models, with lower risk involved, presented below.

Large language model

A large language model, as described in chapter 6.1, which is usually an open-source model retrieved from the internet can be implemented. It has been pretrained on a large corpus of text from the internet to learn to "understand" and process natural language.

As the model can be brought into a local environment at the power plant, following the necessary procedures, the security risk of implementing it can be reduced as it is on-prem. In the local environment at the powerplant, the chosen data, such as reports is retrieved and processed once to enable it to be read by the model. Upon giving the model the query, the model goes through the data to give you the answer. Hence, no data is changed or has the possibility of being corrupted and there is no connection to outside of the powerplant, except the model itself.

AtomAI

As this is a software solution specifically created for the nuclear industry no specific risks have been identified. However, as it is an American company and American nuclear plants implementing the software, this might be considered a risk if it were to be implemented in the Nordics. If the company's[37] solutions would be considered applicable and possible to implement in the nuclear facilities, a more thorough risk assessment should be conducted.

7.2 DATA REPOSITORY AND ETL PROCESSES

To facilitate a useful and in-depth data analysis the data need to be retrieved and stored in a data repository. The data needs to be organized and structured, following steps of the ETL process, or similar.

The risks that are induced with local repository and storage of data are loss of data, due to violation or breakdown, risks to be mitigated by authorization control management systems and backup systems or redundant databases.

Otherwise, there is a risk that the quality of the analysis will be low, thus, the risks with ETL rather lie in not using the process. The risk of relying on decisions made on dispersed, nontraceable person-dependent and redundant data instead of information from an ETL process relying on one single source of qualified data.

7.3 METHODS FOR DATA ANALYTICS, TRENDS AND FORECASTING

When applying a statistical model to your data, it's important to be aware of the following challenges and risks:

- **Assumptions** - Many statistical models rely on assumptions (e.g., normality, linearity) that might not hold in real-world scenarios. Violating these assumptions can lead to biased results.
- **Complex Relationships** - Some statistical models struggle to capture complex relationships between variables. For instance, linear regression assumes a linear relationship, which may not hold in all cases.
- **Limited Flexibility** - Traditional statistical models have limitations in handling non-linear patterns or high-dimensional data.
- **Overfitting** - Overfitting occurs when a model fits noise in the data rather than the underlying pattern or relationship. It can happen if the model is too complex or if there are too many predictors relative to the sample size.
- **Underfitting** - Conversely, underfitting occurs when a model is too simplistic to capture the true underlying patterns. It results in poor predictive performance.

7.4 TECHNOLOGIES FOR VISUALIZATION AND DATA MANAGEMENT

Digitalization always poses a risk if there is not enough cyber security, and the system therefore has the risk of getting hacked. With digital 3D models, important information is gathered in one place, which can be considered a risk if the system is hacked. However, this risk is not specifically associated with the 3D model, but rather with the system where the model is stored. Thus, cybersecurity plays a pivotal role in whether a BIM model or such poses a risk. All the different plants

[37] NuclearN

already have digital systems where data is being stored, the same storage place could also be used for a digital model.

7.5 MANAGEMENT TOOLS IN OTHER INDUSTRIES

There is always a big cost involved when changing to a new system, therefore, there is always a risk involved no matter which tool is chosen. Integration and to be able to transfer all the historical data can be difficult, with the risk of some parts of the information not getting transferred (e.g., metadata or such). A cost-benefit analysis would need to be conducted for all potential tools, as there are always both advantages and disadvantages when implementing something new.

The following sections will briefly present risks and/or potential difficulties with each specific management tool.

7.5.1 BaTMan

In Chapters 6 and 7, BaTMan is extensively discussed along with insights from the two other reports by Energiforsk. Especially Chapter 7 provides an in-depth exploration of the potential of the software. The conclusion drawn from BaTMan is that it is a good type of program, but it would need an update before using it in the nuclear industry. Either, by creating an entirely new program, or by updating parts of the program to better fit the needs and requirements within the nuclear industry. Strengths within the program are the possibilities of importing historical data, which is one of the main issues raised in this current study.

The primary issue and risk with the software lies in its design which is not specifically created for usage within the nuclear industry. Therefore, there has not yet been any test to investigate whether the program is secure enough or not, or how many parts of the software would need to be updated for it to work well at an nuclear facilities. The project group has not been able to find any reports that explicitly investigate the details of an actual implementation in the nuclear industry. Thus, this would need further investigation.

7.5.2 IBM Maximo

This software can be installed in a way where all information will be stored on-site, or it can be cloud-based where IBM provides a secure cloud. Therefore, the security issues can be equated to those of using a 3D model; If the nuclear facility has secure platforms there is no extra risk that comes with updating to this software. An investigation regarding whether or not there exist any less obvious risks has not fit into the scope of this report. Thus, a deeper risk assessment should be conducted if this software is to be used.

7.5.3 Screening Eagle Inspect

The software can only be used with cloud services, which might be considered a high risk by the nuclear industry. However, as the Software is connected to the cloud via Amazon Web services, which has the highest level of security when it comes to cloud services, the risk is reduced. The servers can most likely can be set up in Sweden.

Furthermore, there are already several customers within the nuclear industry using the software together with the cloud services. This has been possible due to good collaboration and careful consideration of the different aspects and requirements within the nuclear industry. An assessment regarding the security aspects of implementing this software at Swedish and Finnish powerplants should be carried out by the Utility owners.

7.6 OTHER TECHNOLOGIES

The main risks are not investigated and might be of interest in another study.

8 Recommended methods and next steps

Performing trend analysis and predictions on concrete data is more intricate than it may initially appear. An in-depth understanding of the context and the data's origin is essential to draw meaningful conclusions from the results of the data analysis. Standardized methods and techniques can be used for trending data, making predictions or identifying variations from individual measurement time series from a structure. The critical factors lie in data quality and type and availability, rather than the specific tool employed. Consequently, utilizing a robust BI platform would suffice for effective analysis.

8.1 HISTORICAL DATA

Before selecting a method and determining the next step, thoughtful consideration of the rationale behind utilizing historical data is imperative. Given the substantial investment required to transform data from old reports into a format suitable for trend analysis, it is important to consider its overall value and feasibility. A comprehensive assessment of the utility of trend analysis and historical data evaluation across various locations within the plant is warranted. The key consideration lies in determining whether every area within the facility holds equal significance, or if the primary focus should be on critical zones such as water flow canals and reactor containment areas. While it may be straightforward to import historical data, the feasibility of conducting meaningful analysis depends on whether robust quality management practices were employed during the data's initial collection.

When extracting historical measurement data, it is advisable to employ ETL processes to transfer both the data and associated metadata to a data repository in a centralized location preferably and in a standardized format suitable for comprehensive analysis. As this endeavour can be time-consuming and involve manual effort, nuclear facilities must strategically prioritize the data they intend to analyse, focusing on critical information and structures. The historical data should be made traceable and accessible to facilitate analysis and pattern detection in the data. Furthermore, the ETL process also needs to be integrated into the current system and to collect data from ongoing measurements.

The implementation of an NLP solution, acting as a chatbot, can enhance knowledge acquisition from historical data, facilitate analysis, and streamline communication with visual inspection reports and other relevant documents. Notably, a company[38] has successfully deployed such solutions in several U.S. plants. Due to the stringent security requirements, the proposed solution must be deployed on-premises, as previously outlined. The recommendation is to initiate a small proof of concept to assess the performance and accuracy of such a model. It is important to note that on-premises implementations of NLP solutions remain less common than their cloud-based counterparts, and their implementation can be more challenging and time-consuming.

[38] NuclearN.ai

To move forward in finding good methods for trending and forecasting of measurement data, a recommended next step is to make a limited set of data and metadata available from one or several different measurement methods as this is required to evaluate different methods of analysis. Several measurement points from structures with identical conditions with similar parameters would be needed. Alternatively, synthetic data could be generated to evaluate different methods on.

To forecast and trend data from chloride concentration, carbonisation and compressive strength measurements, it is recommended to try implementing a log-log regression model as these processes have a declining trend over time. Once a model accurately captures the relationship within the data, it can be used to predict future behavior and estimate when a specific threshold value will be reached.

For other measurements, it is more difficult to recommend a good method for making predictions and finding trends, as the physical relationship between variables is not as easily described. Preferably different regression models can be applied to the data to try and find a model that fits the data well. In some cases, a numerical model could be created given boundary conditions.

To reduce the uncertainty of the prediction, there are several different ways of improving the model fit, such as resampling, feature engineering and improving and cleaning the data.

8.2 FUTURE DATA

If there are not any explicit thoughts as to how useful it would be to import all data, it could be of a bigger value to focus forward. To make sure that enough data is collected and that it is stored in a system where trending and analysis easily can be made. The more data that is collected, the better. Therefore, it might be of use for all nuclear facilities to get inspired by Olkiluoto and invest in more sensors and more data collection. It is of value to do time-series analysis on more places or parts than only moisture penetration values. If the data collected is numerical, and if even more data is collected than that of today, trending can be facilitated in more places in the facilities.

To facilitate easier and faster analysis and better searchability of visual inspections, a more visual system where images and metadata can be stored in specific places in the model is recommended. Screening Eagle Inspect is specifically adapted for logging inspections in a structured way using iPad/iPhone[39] and has several functions which have been requested by the interviewees. The nuclear facilities need to assess whether it is possible to implement the system due to security aspects, but it is already used by other nuclear facilities in Europe. IBM Maximo is an asset management and maintenance system that is well-established with many functionalities such as scheduling work orders, visualization, storing metadata etcetera. Both Maximo and Screening Eagle Inspect can be connected to Power BI

[39] The app is available in App Store but not in Google Play Store

to facilitate analysis. It is recommended to further assess this system or develop a legacy CMMS system to further digitalize the inspection process.

Implementing BaTMan straight off is not possible and for several reasons not recommended. Both IBM Maximo and Screening Eagle Inspect are more adapted and modern systems, they are therefore preferred. However, if the nuclear facilities want to build their own system for structuring inspection data, BaTMan could serve as a good source of inspiration.

Implementing a combination of BIM and 3D models can have several advantages for nuclear facilities, not only for handling inspection data but also for other aspects of the facilities and structures. It is an effective way of visualizing the chosen parts of the plant in a digital format to facilitate virtual site visits. Furthermore, historical data can be included using data extraction. It’s moreover recommended to assess if this solution is of interest and if so, begin on a small scale.

9 Conclusions

There are many statistical methods that can be used for analyzing different kinds of data, such as various regression methods. To determine which methods are most suitable, the methods need to be applied and evaluated on real data, or on synthetic generated data if real data is not available. However, it is complex to analyse data from concrete structures to get an understanding of the full structure.

The outlook towards nuclear facilities in Switzerland and the USA indicates that they have a similar way of working with visual inspections. The best practice in the USA is according to a report from EPRI to collect the data into the structures monitoring program (SMP) database. This facilitates traceability, availability and analysis. To what extent the database is used is not investigated in this report but could be interesting to further investigate.

To facilitate the analysis of measurement data, both historical and current data should be handled by an ETL process to make the data more searchable, available, and traceable. The data and metadata should be stored in a repository in a centralized location preferably. This process would facilitate analysis in an analysis tool such as a business intelligence platform, encompassing activities such as identifying, variations, deviations, trending and making predictions. Statistical methods for trending and forecasting are presented in this report and the different methods need to be evaluated on the data to see which aligns best with reality. Furthermore, it is important to analyse which data is important and possible to trend and prioritize that, as ETL can be a time-consuming process. The quality of the data is an important component for accurate analysis.

To facilitate the analysis of visual inspections in the future, a system that better gathers all information and enables visualization is recommended. A system such as Screening Eagle or IBM Maximo would not only help the inspectors conducting the inspections but also facilitate the analysis of the inspections. Implementing BIM combined with 3D models could also be a viable solution. Furthermore, an NLP solution can be used to make historical data from visual inspections more available.

Appendix A: Method of conducting the interviews

This project serves as a pre-study regarding possible solutions at nuclear facilities to analyse qualitative and quantitative data.

To guarantee a high quality of data collected from interviews, the work carried out after the interview is equally significant as the interview itself. If possible, interviews should be recorded, to ensure a correct transcription of the data. Unfortunately, interviews were not recorded in this project as the interviewees did not grant their permission. To still maintain the correctness of the information used in the report, all information was shared with the interviewees after each interview, providing them with the opportunity to clarify matters, add afterthoughts or correct misrepresentations. This aligns with existing literature about qualitative research methods, which highlights the importance of summarising and transcribing the data as well as sharing the recorded information with the interviewees afterwards.[40]

PRIMARY DATA SOURCES/INTERVIEWEES AND POPULATION

The primary interviewees for this study consist of the steering committee and AFRY employees with experience in the power plant industry. Following the conduction of the interviews, new interviewees were identified due to colleagues referring to them because of their responsibilities at the plants41. Therefore, the interview sample grew as more interviews were being conducted. Moreover, the data from the interviews raised new questions and ideas about potential solutions. Therefore, interviewees from other industries were identified as well as other areas within AFRY.

INTERVIEWS

A semi-structured interview format was chosen to conduct the interviews, as it allows for the interviews to proceed in an unanticipated direction, which could yield valuable information.

Each interview was approximately 1,5 hours long, conducted over teams with two to three people from the project group attending. One member was responsible for taking notes, while another led the interview and asked follow-up questions. If a third member was present, they also took notes and/or focused on the interviewee’s responses. To ensure uniform interpretation of the answers, a summary session was held immediately after each interview to discuss the main points identified during the interview.

[40] Kevin Kelly & Brian Bowe (2011). Qualitative Research Methods in Engineering. 22.1207.1-22.1207.10. 10.18260/1-2--18752.

[41] This could be referred to as chain-referral sampling, which is a fast method that allows for finding fitting interviewees fast and effectively.

COMPILING INFORMATION

The interview data were compiled into useful information by labelling the data and sorting it into smaller units/groups. Compiling the data into smaller units facilitated the comparisons of answers and the identification of potential patterns in the data.

Appendix B: Interview Questions

This appendix contains the questions for the semi-structured interviews. The indented questions represent follow-up questions that were posed if needed. The interviews was mostly conducted in Swedish.

Introductory questions

- Who are you and what is your role in the organization?
    - Duties, number of years worked in this field, education

Project background

- Tell us about your vision of what you intend this project will achieve
    - Why has this project been developed/problem view from your point of view?
    - What is the hope that the tool/project result will be used for? (reason freely)
    - Why do you want a method for analysis?
    - Are there issues that the tool should be able to support?
- How is data currently collected? And why is it collected?
    - Are different data named differently? (labelling)
    - Is there data that is not currently collected but you would really like to collect?

Current situation

- What kind of data is currently available? How much can we share?
    - What does text data look like today? Describe/explain. (if not understood: free text/subj. describe? Is it labelled? Where. language jargon?)
- What is the current inspection situation?
    - Are the inspection requirements government-controlled? Requirements we need to take into account?
    - Who carries out the inspections (competence requirements?)? And how do they do it?
    - What measurements are carried out and how are they logged?
    - How often are measurements carried out and to what extent?
    - Why are the inspections carried out?
- What analyses and methods are used today for the collected data?
    - What kind of info is it, who conducts the analyses, etc.?
    - Ex: There is a change in crack/damage - How is that information processed?
- What's the issue? Time, quality, opportunity for trend analysis? Comparisons with other nuclear facilities? Streamlining?
- Current reporting needs?
- What kind of computer system do you use now? Is it connected to a spatial model/digital twin?

Desired situation

- What is your desired position in 2 years?
    - What are the requirements for analysis support?
- What is your desired position in 5 years?
    - Do you have a vision? Thoughts on the support of AI?
- What are your reporting needs (internal, external, regulatory requirements, etc.)?
- What are your analysis needs?

10 References

Caserta, Joe and Kimball, Ralph, "The Data Warehouse ETL Toolkit, Practical techniques for extracting, cleaning, conforming and delivering data", (Indianapolis: Wiley Publishing Inc, 2004), https://nibmehub.com/opac-service/pdf/read/The%20Data%20Warehouse%20ETL%20Toolkit%20_%20Practical%20Techniques%20for%20Extracting-%20Cleaning-.pdf

Cole, Andrew, "Feature Engineering, improving a Linear regression", towards data science (blog), 2020-07-01, https://towardsdatascience.com/feature-engineering-combination-polynomial-features-3caa4c77a755

Cvitanovic, Peter, Täpp, Marit and Midböe, Finn 2021. *Making inspections of concrete structures in nuclear powerplants more efficient.* Energiforsk.

EPRI, "*Structures Monitoring Program Inspection records - Best practices for Preparation, utilization and retention".* Published August 2021, https://www.epri.com/research/products/000000003002020988

Hassanzadeh, Manouchehr. Energiforsk, *Förvaltning av Kärnkraftens Betongkonstruktioner,* Energiforsk Rapport 2020:684.

Heath, Alex, "*All the news from OpenAI's first developer conference*", The Verge, last updated 2023-11-19, retrieved 2024-01-10, https://www.theverge.com/2023/11/6/23948619/openai-chatgpt-devday-developer-conference-news

Hyndman, R.J., & Athanasopoulos, G. *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia, 2018, OTexts.com/fpp2,

James, Gareth et al., "*An Introduction to Statistical Learning, with Applications in R*", (Springer), 2013, https://www.stat.berkeley.edu/~rabbee/s154/ISLR_First_Printing.pdf, (retrieved 2024-06-07)

Kelly, Kevin & Bowe, Brian. (2011). Qualitative Research Methods in Engineering. 22.1207.1-22.1207.10. 10.18260/1-2--18752.

K2view, "What is Synthetic Data Generation?", last updated 2024-05-07, https://www.k2view.com/what-is-synthetic-data-generation/#:~:text=Synthetic%20Data%20Generation%3F-,Synthetic%20data%20generation%20is%20the%20process%20of%20creating%20artificial%20data,%2C%20models%2C%20or%20other%20techniques

Kiwa. 2023. *Världens första AI-igenkänning för stora betongstrukturer i 3D*, Kiwa, last updated 2023-01-30, retrieved 2024-02-02, https://www.kiwa.com/se/sv/media/nyheter/varldens-forsta-ai-igenkanning/

LinkedIn, "How do you manage data uncertainty", *Linkedin, https://www.linkedin.com/advice/3/how-do-you-manage-data-uncertainty-skills-data-science.*

Lundqvist, Peter and Zaraei, Nick, Compilation of Concrete Test Data from The Nuclear Industry, Energiforsk 2021:761, https://energiforsk.se/media/29656/compilation-of-concrete-test-data-from-the-nuclear-industry-energiforskrapport-2021-761.pdf, retrieved 2024-04-26.

NuclearN, "AtomAssist", *NuclearN*, retrieved 2023-03-19, https://nuclearn.ai/our-product/nuclearn-atomassist/

Screening Eagle Technologies, "*Inspect*", https://web.screeningeagle.com/home, Retrieved 2024-03-11

Nuclear facilities in Sweden and Finland, *THAG,*

Interviewees:

Ahmad Ostovar, AFRY - NLP
Andreas Buhlin, AFRY – Digital twin
Andreas Svedberg, AFRY – 3D scanning
Conrad Malmstedt Ledin, AFRY - BIM
Erik Hansson, OKG – Oskarshamn
Florian Müller, AFRY – Screening eagle, Switzerland
Hannu Pyy, AFRY – Concrete expert nuclear Finland
Johan Söderman, AFRY – concrete expert, BaTMan
Johanna Spåls, Vattenfall - Ringhals
Jonas Bergfors, SKB
Jussi Ritola, AFRY – Concrete expert nuclear Finland
Lukas Karlsson, AFRY - AI
Marcus Edin, Vattenfall – Forsmark
Martin Kwasniewski, AFRY – Digital twin
Michael Blom, AFRY – Cybersecurity
Mika Oikari, AFRY – Concrete expert nuclear Finland
Monica Andreasson, AFRY – IBM Maximo
Niclas Täck, AFRY – Cybersecurity
Peter Lundqvist, Vattenfall
Per Thunstedt, Trafikverket
Pontus Tossavainen, Vattenfall – Ringhals
Shahad Antoan, AFRY - BIM
Thilda Jonsson, Vattenfall - Forsmark
Timo Kukkola, TVO - Olkiuoto
Timo Leppänen, Fortum
Tristan Tyler, Screening Eagle
Ulrik Brandin, OKG – Oskarshamn
Ulrika Stålhammar, AFRY - NLP
Urban Axelsson, OKG
Violeta de Lama, AFRY – 3D scanning

METHODS FOR COLLECTION AND EVALUATION OF CONCRETE DATA WITHIN THE NUCLEAR INDUSTRY

The Swedish and Finnish nuclear facilities need methods to analyse data from measurements and visual inspection. Currently, data exist in several locations but is often stored in reports. An important first step is to establish an ETL process to extract, transform and load the data into a data repository to facilitate the analysis. Various statistical methods can then be used to analyse the data to make trends and predictions, such as regression analysis.

To analyse future data from visual inspections and improve data collection, the recommendation is to use a software available on the market, which often includes 3D models and facilitates documentation through a tablet. Implementing BIM and 3D modelling could also be a viable solution.