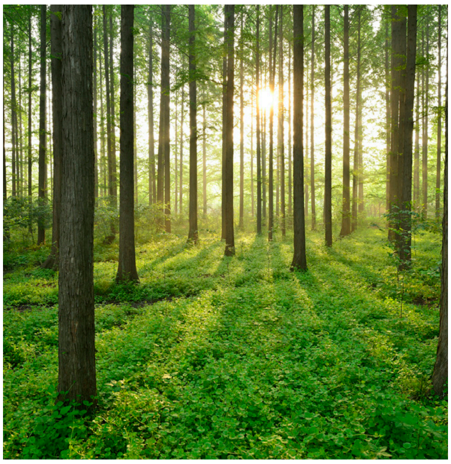
USES FOR BIG DATA IN THE NORDIC NUCLEAR POWER PLANTS

REPORT 2026:1166



ENERGIFORSK NUCLEAR SAFETY RELATED I&C, ENSRIC

Energiforsk

Uses for big data in the Nordic nuclear power plants

Experiences, challenges and lessons learned

TOMMIE LINDQUIST AND MARKUS ERIKSSON

ISBN 978-91-89917-09-5| © Energiforsk February 2026

Energiforsk AB | Phone: 08-677 25 30 | E-mail: kontakt@energiforsk.se | www.energiforsk.se

Foreword

This report forms the results of a project performed within the Energiforsk Nuclear Safety Related I&C (ENSRIC) Program. The ENSRIC Program aims to increase the knowledge of aspects affecting safety, maintenance and development of I&C systems and their components in the Nordic nuclear power plants. Part of this is to investigate possibilities to facilitate and simplify the work that is performed in the nuclear business.

The rapid digital transformation is creating new opportunities and challenges for the Nordic nuclear power plants. Big data can support safety, efficiency and asset management, which makes its secure and effective use increasingly important for the sector.

This study maps current and emerging uses of big data in Nordic nuclear power plants, identifies technical and organisational requirements, and offers practical recommendations for secure, scalable, and user-oriented solutions tailored to the nuclear context.

Results show that adoption is at an early stage and recommends focused pilot projects for example in predictive maintenance. Flexible tool ecosystems, robust data management, and strong human involvement are key for successful implementation and future progress.

The study was carried out by Tommie Lindquist and Markus Eriksson, RISE. The study was performed within the Energiforsk ENSRIC Program, which is financed by Vattenfall, Uniper, Fortum, TVO, Skellefteå Kraft and Karlstads Energi.

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

Summary

The ongoing digital transformation of the Nordic nuclear power sector is creating new opportunities to enhance safety, efficiency, and asset management through the use of big data. This report presents a comprehensive overview of the current state and future potential of big data applications in Nordic nuclear power plants (NPPs), drawing on literature, interviews with industry experts sharing practical experiences from both the nuclear sector and other safety-critical industries.

The project set out to map existing and emerging uses of big data in nuclear power plants, identify the technical and organisational requirements for successful implementation, and offer practical recommendations for secure, scalable, and user-oriented solutions. The analysis encompasses both direct applications, such as predictive maintenance, process optimisation, and decision support, as well as supporting measures including data integration, governance, and change management.

Big data adoption in Nordic NPPs is still at an early but evolving stage. While large volumes of process and maintenance data are collected, most analysis remains manual and siloed. Lessons from other safety-critical sectors show that success with big data initiatives depends on combining domain expertise with advanced analytics, starting with focused pilot projects, and maintaining a strong human-in-the-loop principle. A key finding of this report is that, instead of relying on a single platform or solution, organisations should develop flexible ecosystems of specialised tools that can be adapted as needs evolve. To overcome technical and organisational barriers, it is essential to ensure robust data quality management, clear governance structures, and user-focused approaches. Furthermore, integration of data sources and retention of raw data are identified as critical enablers for future progress.

Drawing on these insights, the report recommends that Nordic NPPs initiate well-defined pilot projects with demonstrable value, prioritising predictive maintenance as an initial focus. Solutions should be modular, interoperable, and developed in close collaboration with end users to ensure that they address real operational needs, are accepted in practice, and support effective adoption and change management.

By applying these recommendations and drawing on relevant experiences, Nordic NPPs can make more effective use of big data to improve safety and efficiency in their operations.

Keywords

Big Data, Nuclear Power Plants, Data Integration, Analytics Software, Decision Support, Predictive Maintenance, Process Optimisation

Sammanfattning

Den pågående digitala omställningen av den nordiska kärnkraftsektorn öppnar nya möjligheter att stärka säkerhet, effektivitet och tillgångsförvaltning genom användning av big data. Denna rapport ger en heltäckande översikt av nuläget och framtida potential för användning av big data i nordiska kärnkraftverk, baserat på litteraturstudier, intervjuer med branschexperter och praktiska erfarenheter från både kärnkraftssektorn och andra säkerhetskritiska branscher.

Projektets syfte var att kartlägga befintliga och kommande användningsområden för big data i kärnkraftverk, identifiera tekniska och organisatoriska krav för framgångsrik implementering samt ge praktiska rekommendationer för säkra, skalbara och användarvänliga lösningar. Analysen omfattar såväl direkta tillämpningar, såsom prediktivt underhåll, processoptimering och beslutsstöd, som stödjande åtgärder som dataintegration, styrning och förändringsledning.

Resultaten visar att användningen av big data i nordiska kärnkraftverk fortfarande befinner sig i ett tidigt utvecklingsskede. Trots att stora mängder process- och underhållsdata samlas in sker de flesta analyser fortfarande manuellt och i isolerade silos. Erfarenheter från andra säkerhetskritiska branscher visar att framgångsrika big data-initiativ bygger på att kombinera domänkunskap med avancerad analys, att starta med mindre fokuserade pilotprojekt och att upprätthålla en mänsklig delaktighet i beslutsprocessen. En viktig slutsats i denna rapport är att organisationer, istället för att förlita sig på en enda plattform eller lösning, bör utveckla flexibla ekosystem av specialiserade verktyg som kan anpassas efter verksamhetens behov. För att övervinna tekniska och organisatoriska hinder är det avgörande att säkerställa datakvalitet, tydliga styrstrukturer och användarcentrerade arbetssätt. Dessutom identifieras integration av datakällor och bevarande av rådata som kritiska möjliggörare för framtida framsteg.

Mot bakgrund av dessa insikter rekommenderar rapporten att nordiska kärnkraftverk inleder väl definierade pilotprojekt med tydligt mätbart värde, där prediktivt underhåll prioriteras som ett första fokusområde. Lösningarna bör vara modulära, interoperabla och utvecklas i nära samarbete med slutanvändarna för att säkerställa att de möter verkliga operativa behov, kommer att accepteras i praktiken och stödjer effektiv förändringsledning och implementering.

Genom att tillämpa dessa rekommendationer och dra nytta av både interna och externa erfarenheter kan nordiska kärnkraftverk använda big data mer effektivt för att förbättra säkerhet och effektivitet i sin verksamhet.

List of content

Energiforsk

List of abbreviations

AI	Artificial Intelligence
CV	Computer Vision
DL	Deep Learning
I&C	Instrumentation and Control
IoT	Internet of Things
LLM	Large Language Models
ML	Machine Learning
NPP	Nuclear Power Plant
RAG	Retrieval-Augmented Generation
SHAP	SHapley Additive exPlanations
SQL	Structured Query Language
VIF	Variance Inflation Factor

1 Background and objectives

1.1 BACKGROUND

The Nordic nuclear power industry is currently facing increasing demands for operational efficiency, safety, and lifetime extension. At the same time, the volume and complexity of available data are growing rapidly, driven by digitalisation, the proliferation of sensors, and the integration of advanced monitoring systems. This development brings both opportunities and challenges: while data-driven approaches can support predictive maintenance, fault detection, and improved decision-making, implementing such systems securely and effectively in safety-critical environments remains a complex task. To address these challenges, this project was initiated within the ENSRIC programme, focusing on safety-related instrumentation and control as well as long-term asset management. The work aims to provide insights and recommendations for how big data can be used to support safe, efficient, and cost-effective operations in Nordic nuclear power plants (NPPs).

1.2 OBJECTIVES OF THE PROJECT

The main objective of the project is to explore how big data technologies can be applied in Nordic NPPs to support safe, efficient, and cost-effective operations. Specifically, the project aims to:

- Define and contextualise the concept of big data as it applies to NPPs
- Identify current and emerging applications of big data in safety-critical industries, with a focus on predictive maintenance, condition monitoring, and decision support.
- Assess the technical and organisational requirements for successful implementation in NPPs.
- Compare best practices from other industries and international NPPs to identify transferable strategies.
- Investigate how critical low-volume data can be prioritised over high-volume, low-priority streams.
- Identify and evaluate relevant tools, platforms, and vendors suitable for effective big data management and analytics in the NPP context.
- Provide recommendations for secure, scalable, and user-oriented big data solutions tailored to the Nordic nuclear context.

1.3 REPORT STRUCTURE

Chapter 1 establishes the context, the underlying challenges, and articulates the objectives that motivate this study. Chapter 2 offers an account of the methodology employed, including both a literature review and a series of interviews conducted.

Chapter 3 is exploring the concept of big data as it pertains to the nuclear power sector. It provides definitions, identifies key characteristics, and presents an overview of current practices in Nordic NPPs, with particular attention to both technical and organisational dimensions. Chapter 4 examines practical applications and lessons learned, drawing upon experiences from the nuclear industry as well as other safety-critical sectors. Chapter 5 presents a study of available tools, platforms, and vendors relevant to big data management and analytics, integrating insights derived from both the literature and the interviews. Chapter 6 addresses critical considerations for implementation, including technical, organisational, and regulatory factors that influence the successful adoption of big data solutions in NPPs. The report concludes with Chapter 7, which provides a synthesis of the principal findings and a set of practical recommendations intended to guide Nordic NPPs in their efforts to implement big data solutions. Chapter 8 contains a complete list of references, and the Appendix offers summaries of interviews with industry representatives.

2 Methodology

The methodology employed in this project consists of a combination of semi-structured interviews with industry representatives and an extensive review of the relevant literature, thereby ensuring both practical perspectives and a comprehensive understanding of the current state of knowledge.

2.1 LITERATURE REVIEW APPROACH

The literature review was carried out to establish the current state of practice, definitions, methods, and challenges related to the use of big data in safety-critical industries, with a particular focus on NPPs. The review covered the following types of sources:

- Technical reports and guidance documents from international organisations and industry associations relevant to nuclear safety, plant lifetime management, asset integrity and digitalisation.
- Peer-reviewed scientific publications on predictive maintenance, condition monitoring, process optimisation, data integration, and asset performance management.
- Documented case studies from safety-critical industries such as nuclear power, petrochemical processing, and aviation, including examples of applied analytics and digital monitoring in operational environments.
- Publicly available material from technology suppliers and vendors offering data-driven solutions for monitoring, diagnostics and decision support in industrial plants.

The findings from the literature review are presented alongside the interview results in Chapters 4, 5 and 6.

2.2 INTERVIEW APPROACH

To complement the literature review and capture current industrial practice, semi-structured interviews were conducted with representatives from Nordic nuclear power plants as well as from other safety-critical industries and an NPP operator.

Each interview covered, in particular:

- Current and planned use of data for predictive maintenance, lifetime management and process optimisation.
- Data sources, data storage solutions, and system integration practices (including on-premises, hybrid and cloud-based architectures).
- Organisational roles and responsibilities for data ownership, analysis and decision-making.

- Barriers and enabling factors for big data in nuclear power plants (technical, organisational, regulatory, cybersecurity-related).
- Approaches to prioritising critical information from safety-relevant systems, even when such systems generate comparatively low data volumes.

Interviews typically lasted 60 minutes and were conducted via Teams. Notes, transcripts and recordings were analysed thematically.

The interviews conducted for this study included representatives from the following organisations:

- Ringhals, a Nordic nuclear power plant operator based in Sweden.
- TVO, a Finnish nuclear power plant operator.
- Bruce Power, a Canadian company operating the world's largest nuclear power facility.
- INEOS Inovyn, a European producer of petrochemicals with operations across Europe.
- Borealis, an international producer of polyolefins, base chemicals, and fertilisers.

The insights from interviews with TVO and Ringhals are presented in Section 3.3, while Chapters 4, 5 and 6 exclusively draw on interviews conducted with Bruce Power, INEOS Inovyn and Borealis. A complete list of organisations and individuals who participated in the interviews is provided in the Appendix .

2.3 LIMITATIONS

The analysis is based on publicly available literature and interviews, without access to confidential operational data. Technical benchmarking of specific tools or platforms has not been performed; instead, the assessment relies on reported experiences and best practices. The recommendations are tailored to the Nordic context and may not be directly applicable elsewhere. Economic cost-benefit analyses and detailed cybersecurity or regulatory issues are not addressed.

3 Big data in the context of nuclear power

3.1 DEFINITION AND CHARACTERISTICS

Big data is often defined in the literature as extensive and rapidly generated datasets that exceed the capabilities of conventional processing methods [1], [2], [3], [4], [5]. Data sources in this context includes traditional industrial big data (e.g., real-time sensor observations), system-level models like fault trees, results from multi-physics simulations at both component and system levels, material characteristics, performance tests, inspection and maintenance reports, operating procedures, causal models, and more [1]. This data is collected and applied at different time and spatial scales as well as different levels of abstraction and it is available in variety of modalities including text, numerical data, images and tables [1], [3], [6].

When defining big data, the term tends to be further to broken down using ether the three Vs: volume, velocity, and variety, or the recently more common five Vs expanding on the previous with the inclusion of veracity and value [3], [4], [5].

- Volume: The sheer amount of data generated from numerous sensors and systems which can, in some cases, be in the range of petabytes or more [2]. It is fairly common that organisations are sitting on a large amount of data but lack the capability to analyse it [4].
- Velocity: The speed at which data is generated and needs to be processed, where limitations in processing capabilities relative to the volume and speed of the data, can be a challenge [3], [5].
- Variety: The different types of data generated from various sources like sensors, models, and maintenance logs [1], [7]. The generated data include structured, semi-structured, and unstructured data in the form of numbers, text, images, audio or videos [1], [3], [5], [6].
- Veracity: The reliability and trustworthiness of the data. It involves identifying and managing faulty, noisy, or incomplete information within large, fast-moving, and diverse datasets, as well as assessing the credibility of the sources that generate the data [4], [5].
- Value: The potential insights and benefits derived from analysing this data. The infrastructure to gather and store this data can be costly, hence an important characteristic of big data is the valuable insights that are possible to derive from it [4], [5].

The nuclear industry is relatively unique in the size, variety, scope, technical sophistication, organization, and quality of available data and models that capture system performance under normal operations and a wide-range of adverse event conditions [8]. As such, these five Vs of big data are particularly relevant in the context of NPPs, as big data in this context refers to the large volumes of heterogeneous data generated from various sources such as sensors, maintenance records, operational logs, and incident data [7].

The nuclear power sector has accumulated a vast and growing volume of data, models, and analytical tools, ranging from probabilistic risk assessments to advanced simulations and comprehensive databases, reflecting the industry's technical sophistication and the complexity of its operations. Figure 1 illustrates the overall flow of big data services in NPPs, highlighting the key stages from data acquisition to processing and storage [1].

The data is gathered from a large variety of sources including plant sensors, experiments, models, maintenance reports and incident data [1], [7], [9]. The data is also available in variety of modalities ranging from text-based, to numeric data, to images and tables. The velocity of the data is also evident in real-time monitoring systems [10]. However, despite investments from nuclear plants in gathering the data, much of this information remains underutilised, often confined within isolated systems or used in static ways.

Managing this data effectively also requires attention to veracity, since the data can include outliers, communication errors, incomplete information and noisy signals [11]. Ensuring data quality is essential for reliable analysis and informed decision-making. Finally, the value of these vast datasets for nuclear power plants stem from the potential insights and benefits derived from analysing this data to improve maintenance processes, operational efficiency, document management and decision support

In short, big data in the NPP context represent the large and heterogeneous collection of information streams consisting of sensor readings, maintenance reports, model outputs, and other plant documentation that exceed the capabilities of traditional processing tools, however when validated and processed can support various plant analysis and decision making.

Recent advances in data science and analytics offer new opportunities to integrate and interpret these diverse resources, enabling deeper system understanding and more informed decision-making. Harnessing these capabilities could play a crucial role in enhancing both the safety and operational efficiency of nuclear power plants.

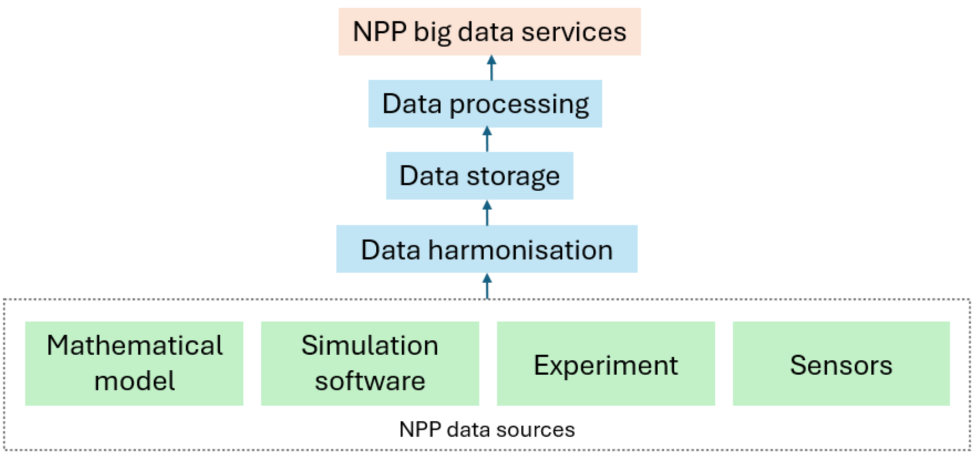


Figure 1. Flowchart of NPP big data services. Adapted from [1].

3.2 BIG DATA ANALYTICS

While big data engineering and big data analytics are closely related, they serve distinct roles within a data-driven organisation. Big data engineering focuses on building and maintaining the technical infrastructure required to collect, process, and store large volumes of data. This includes tasks such as data acquisition, cleaning, integration, and the development of data pipelines and storage solutions.

In contrast, big data analytics is concerned with extracting value from this data by applying statistical methods, machine learning (ML), deep learning (DL) and domain expertise to identify patterns, generate predictions, and support decision-making. In essence, engineering ensures that high-quality, accessible data is available, while analytics transforms that data into actionable insights that can improve operational efficiency and safety.

Figure 2 presents the main components of big data systems in nuclear power plants, distinguishing between the processes of data engineering and data analytics.

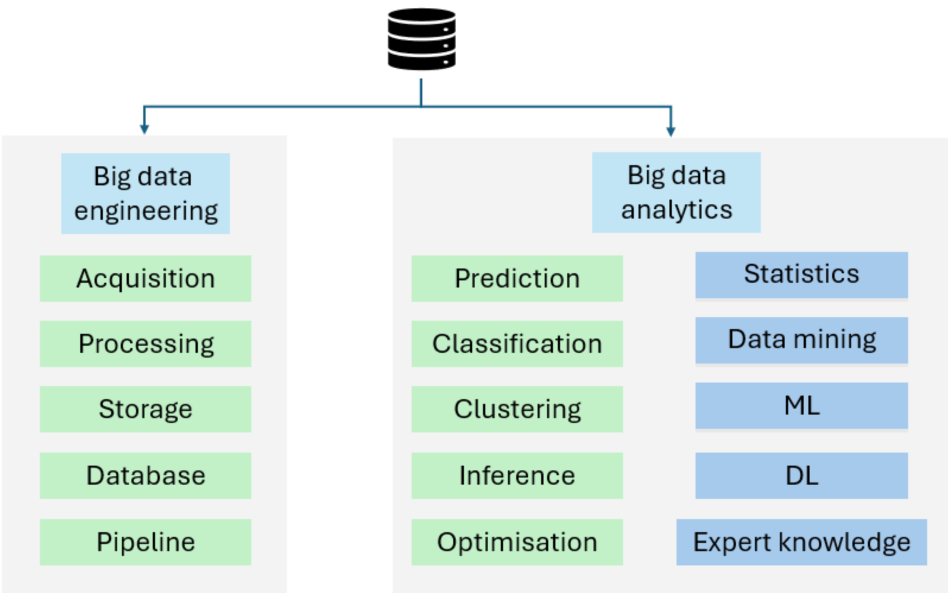


Figure 2. Paradigm of big data components. Adapted from [1].

3.3 CURRENT USAGE OF BIG DATA IN THE NORDIC NPPS

The following section describes the current use of big data in Nordic nuclear power plants, drawing on insights from interviews with representatives of both TVO and Ringhals. The use of big data in Nordic nuclear power plants is at an early but evolving stage, with both organisations actively exploring how to leverage data for operational improvements, maintenance, and decision support.

More information from the interviews can be found in the Appendix section.

3.3.1 Data collection and storage

Both organisations collect large volumes of process data from sensors and control systems, typically at relatively low sampling rates (e.g., minute intervals). This data is primarily used for manual analysis, such as investigating events, validating models, and comparing historical situations. Maintenance data and other operational information are stored in separate systems, resulting in data silos that are not yet fully integrated. Almost all data is stored on-premises, with long-term retention policies; in some cases, data is averaged to save disk space, although there is a trend towards retaining more raw data as storage costs decrease.

Current applications:

- Manual analysis: Data is mainly used for manual investigations and validation of models. Automated, real-time analytics are limited, and most analyses are performed in response to specific events or operational needs.
- Predictive maintenance: While there is interest in condition-based and predictive maintenance, most maintenance remains scheduled at fixed time intervals. Some interviewees note that pilot projects and exceptions exist, such as anomaly detection for heat exchangers and condensers, but these are not yet widespread.
- Process optimisation: Plants use data to optimise performance and energy output, but this is typically done through manual trend analysis rather than advanced analytics.
- Model validation: Historical data is used to manually validate simulation models and support operational decision-making.
- Integration and data silos: A major challenge is the lack of integration between different data sources, such as process data, maintenance records, and inspection logs. Data silos hinder the ability to perform holistic analyses and limit the potential for advanced applications like predictive maintenance and digital twins. There are ongoing discussions and some initial steps towards integrating these systems, but full integration has not yet been achieved.

3.3.2 Tools and analytics platforms

Analysis is typically performed using tools such as Excel, MATLAB, Python scripts, and process modelling software such as AspenTech. Some pilot projects have tested anomaly detection models and other advanced analytics, but these are not yet in regular use.

3.3.3 Organisational roles and data governance

Responsibility for data management is still being defined. While some roles are established, overall governance remains fragmented, and there is recognition that clearer structures are needed as data usage expands.

3.3.4 Security and regulatory considerations

Both organisations are cautious about cloud solutions due to security and regulatory requirements. Most data is kept on-premises, and there is a strong focus on information classification and access control.

4 Use of big data in NPPs and other safety-critical industries

Even if big data is not currently applied to a large extent in the Nordic NPPs, it is being applied in other safety critical industries and NPPs outside the Nordic region and have been so for a long time. Lessons can therefore be learned from these industries with regard to best practices for application of big data which can be carried over to the Nordic NPPs.

This chapter presents a comparative overview of big data applications in nuclear power plants and other safety-critical industries. It focuses on areas such as predictive maintenance, operational optimisation, document management, and decision support, combining findings from the literature with practical insights from international industry interviews.

The interview-based insights presented in this chapter are based exclusively on the interviews with Bruce Power, INEOS Inovyn and Borealis, and do not include input from TVO or Ringhals.

4.1.1 Values and cost of big data for NPPs

In NPPs, many different fields are applied and therefore a lot of diverse data is generated ranging from plant process signals to maintenance logs, procedures, drawings, and event reports [1]. There are therefore many and varied applications of big data analytics in NPPs which can provide improvements to economics, efficiency and safety. Big data analytics can further help by modelling complex systems by creating nonlinear relationships [1].

When it comes to the cost of big data, the extra staff required have been of interest to the reference group. However, it has proved difficult for the interviewees from INEOS Inovyn, Borealis and Bruce Power to specify a number for the extra staff required to work with big data and data analytics. Since their processes have been established for a long time, often decades, and this work is deeply integrated in their day-to-day business, any "before" situation is difficult to define. Nonetheless, one number that was mentioned during the interviews, was that a team of 10 was needed for the cross functional work of IT projects, change management and data science.

The following key areas where big data offers significant value have been identified:

- Predictive Maintenance
- Operational Optimisation
- Document Management
- Decision Support

These areas are discussed in greater detail in the following sections.

4.2 PREDICTIVE MAINTENANCE AND LIFETIME MANAGEMENT

The possibility of using data analytics and big data for predictive maintenance is perhaps the largest of all studied fields, it has been mentioned in all interviews as well as a vast body of work in the literature.

4.2.1 Insights from literature

Early anomaly detection using big data analysis for critical systems can reduce unplanned outages, optimise maintenance schedules and therefore lower costs [1], [12], [13]. ML can be applied for training on historical data to detect deviations from normal conditions and faulty conditions [13]. The trained model can then be applied for continuous monitoring, to utilise the historical analysis for detecting anomalies and potential faults. Furthermore, ML techniques, such as long short-term memory and support vector regression, can be applied to predict future behaviour and plant conditions to better inform maintenance planning and reduce down time [13].

Big data can also help in the detection of visual anomalies using Computer Vision (CV) [12]. Examples include monitoring for anomalies in the fuel rods during long time storage or help in performing visual inspections in areas inaccessible due to for example, high radiation exposure. In these cases, drones could be used together with CV to proses the footage to further analyse the area. Even Large Language Models (LLMs) can be useful in predictive maintenance, fault detection, and prevention [12]. Data from processes such as ML analysis could be incorporated into prompts to an LLM, enabling queries that help interpret and understand the data.

Practises observed in other industries include the aerospace and aviation industry with a decades long history of online health monitoring and integrated system health management [14]. The integrated system health management utilises historical records combined with sensor data to assess and predict the health of components and subsystems. It supports diagnostics and remaining useful life estimations. CV is also applied within the industry to detect aircraft defects, lightning damage, dents, cracks and holes [15].

NPPs can employ similar ML models for continuous monitoring and forecast equipment degradation, enabling identifying faults earlier and assisting with risk assessment for maintenance scheduling. Experiences from the CV utilisation in aerospace and aviation is also of great interest, that through analysing images, it can support in finding anomalies and defects.

The petrochemical industry is another industry which is continuously applying big data for predictive maintenance and fault detection, based on historical and real-time sensor data. However, the petrochemical industry also brings up using big data analysis for complex hybrid modelling, combining physics-based models with ML to enhance predictive power and reliability [16]. ML analysis through adaptive wavelets and artificial neural networks are used to detect defects like corrosion and cracks in pipelines while predictive analytics based on historical incident data and real-time monitoring are used for online monitoring and to forecast hazards conditions for critical equipment.

4.2.2 Insights from interviews

Interviewees describe several ways in which big data and ML have enhanced predictive maintenance and lifetime management. At Borealis, predictive maintenance is increasingly supported by models provided directly by equipment suppliers, especially for standard components that are manufactured in large numbers. This approach leverages supplier expertise, while internal teams focus on integrating these models with plant operations. Predictive maintenance is used to detect anomalies early, enabling interventions before failures occur and reducing costly downtime.

Bruce Power has implemented CV-systems for several maintenance-related tasks. These systems automatically identify equipment anomalies, such as misaligned cables on robots, and to monitor safety compliance by detecting whether personnel are wearing hard hats. CV also supports automated reading of gauges and dials during operator rounds, and vibration sensors are used for real-time monitoring of pumps and other critical assets. These solutions have significantly reduced analysis time and improved detection rates, allowing maintenance teams to address issues more efficiently and proactively.

INEOS Inovyn notes that predictive maintenance is most effective where instrumentation is sufficient and failures are frequent enough to justify investment. In their operations, process data is continuously collected and analysed to identify trends and potential issues, but the business case for predictive maintenance is strongest in environments where unplanned outages are particularly costly.

Interviewees also highlight that predictive maintenance models are most valuable when they provide actionable insights for technicians and engineers, such as forecasting when a component will reach a critical threshold or offering clear recommendations for intervention. Health indices without a clear time limit before which an action must be taken, are less valuable.

4.3 OPERATIONAL EFFICIENCY AND OPTIMISATION

Applying analytics to operational data allows for proactive adjustments, improved energy output, and streamlined processes in NPPs.

4.3.1 Insights from literature

By enabling more informed, data-driven decisions across the operational lifecycle, big data analytics can improve performance and energy output and support better operational decisions [1], [13], [7]. Whitin this context, ML trained artificial neural networks may be used to optimise fuel management and provide optimal targets of different operational variables based on the NPP operational state [1]. Insights derived from big data analysis may therefore provide operators with extra insights allowing them to adjust, optimising energy usage and improving overall operational efficiency [13]. Automation through big data can also reduce the man hours needed for routine tasks, allowing for more effective use of human resources [7].

Whitin other industries, real-time sensor data from rigs inthe petroleum industry is continuously analysed to minimize non-productive time and enhance safety through automated drilling state detection and predictive analytics [3], [16]. Further, ML models are used to analyse sensor data to optimise operation efficiency, for example by forecasting production decline, and can therefore assist with allocating resources more optimally.

4.3.2 Insights from interviews

Interviewees report that operational efficiency and optimisation are best achieved by combining process expertise with targeted analytics. At Borealis, ML-driven forecasts are used to optimise production schedules based on the current electric energy price, directly reducing variable costs and improving output. Data-driven models also help predict product quality, which shortens laboratory analysis times and enables faster adjustments in production.

Bruce Power highlights the value of CV and sensor data for real-time monitoring, enabling earlier detection of process deviations and supporting more proactive operational decisions.

INEOS Inovyn notes that continuous analysis of process data allows for incremental improvements in operating procedures and more reliable identification of trends and anomalies.

4.4 DOCUMENT MANAGEMENT

The NPPs in the Nordics have internal repositories that holds up to several millions of documents [12]. In order to process this large and varied amount of data, effective tools are necessary.

4.4.1 Insights from literature

NPP operators have up to millions of stored documents where the utilizations of big data analysis could provide with automated search, classification, summarization and digitalization [12]. LLMs utilising Retrieval-Augmented Generation (RAG) can retrieve internal documents and, based on the applications, the LLM may then use the retrieved information to assist in analysing failure reports for assessing root cause, assist in reviewing of work routines as well as inform maintenance plans [12]. LLMs can also be of assistance for the creation of these documents, in particular in more complex scenarios where the current automation for report generation is insufficient. In addition to LLMs, CV could also be applied within this context to assist in creating digitalised versions of older documents and maybe in particular older drawings [12].

Care should however be taken when applying these systems as access control and classification rules need to be enforced [12]. Only information from documents a person has the right to access can be displayed to this individual.

4.4.2 Insights from interviews

Bruce Power is actively digitising plant documentation, integrating data from various sources, and developing digital twins to create a unified, accessible data environment. CV is used to automate the reading of gauges and dials during operator rounds, and to support the digitalisation of paper-based records.

Borealis and INEOS Inovyn also emphasize the importance of centralising documentation and making it easily searchable, enabling faster access to operational information and supporting more efficient analysis. The most valuable approaches combine digital archives with tools that allow users to quickly retrieve, review, and analyse documents relevant to maintenance, compliance, and operational decision-making.

4.5 DECISION SUPPORT

Integrating big data into decision support can strengthen the operator’s role, offering actionable insights without replacing human judgement.

4.5.1 Insights from literature

Big data analysis provides a further in depth analysis, which provide additional resources for making informed decisions [1], [12]. LLMs can assist in diagnostics and troubleshooting by providing explanation and suggestions in a natural language and RAG can be utilised to assist in in retrieving relevant information with regard to ensuring regulation adherence [12]. Beyond LLMs, does other big data analysis also provide resources for decision support. It was for example noted that artificial neural networks can be applied for optimal operation decision support, but also aid in providing context for fault supervision and diagnosis as well as support with corrective actions during failure [1].

However, this application requires careful review to ensure that safety and security requirements are strictly adhered to. Ther are significant dangers in overreliance on automation resulting in a relaxation on the human supervision [12]. The objective should therefore be a collaborative relationship between the operators and the big data analysis to provide with further insight for efficient operation without overreliance on it [1].

4.5.2 Insights from interviews

The interviews provide several perspectives on decision support in safety-critical industries. All three organisations emphasise that the main purpose of decision support systems is to strengthen the role of the human operator, not to automate decisions.

Borealis describes how their analytics solutions are designed to deliver actionable information to process engineers and operators, but the final decision is always made by staff with subject matter expertise. They highlight the importance of integrating decision support into established workflows, ensuring that recommendations are clear and relevant for those who use them in daily operations.

Ineos Innovyn also stresses that decision support tools are used to guide, not control, operational processes. Their approach is to provide engineers with forecasts and recommendations, such as predicting when a component will reach a critical threshold but always leaving the responsibility for action with the human expert. They note that models are most valuable when they offer transparent and interpretable outputs, allowing users to understand the reasoning behind each recommendation.

Bruce Power provides further examples of how decision support is implemented in practice. Their systems use advanced analytics and ML to detect anomalies and support maintenance planning, but the organisation is clear that automation is intended to augment human expertise rather than replace it. The interviewee from Bruce Power describe the "human-in-the-loop" principle as central to their strategy, particularly for safety-critical decisions and outage planning. They emphasise that decision support tools must be transparent and actionable, and that operators must retain control over all critical processes.

4.6 STRATEGIES FOR PRIORITIZING CRITICAL LOW-FLOW DATA OVER STREAMS WITH HIGH-VOLUME LOW-PRIORITY DATA

In the context of this study, the question of how to prioritise low-volume, high-priority data streams over high-volume, low-priority ones has not emerged as a significant challenge among the organisations interviewed. This is largely because AI and big data analytics are currently applied as decision support tools, rather than for autonomous control or real-time decision-making, see Section 4.5. As a result, the following discussion draws primarily on insights from the literature, where such strategies have been explored in greater technical detail.

In automated decision-making and machine-learning systems, high-frequency, high-volume data streams can overshadow low-frequency but highly informative signals. Handling this issue and ensuring that this critical minority data is properly prioritised can be separated into two related problems:
(1) Classification of data into imbalanced minority (less frequent data that potentially contain more information) and majority (high frequency data) classes
(2) Learning from imbalanced classes when the model's goal depends on the minority information.

Class imbalance is a well-studied problem, and [17], [18] are influential works which describe methods for classification to alleviate a bias towards the majority class present in "base" classifiers. When new data enters a process and is to be assigned to a minority or majority class based on its information content, this classification is usually performed by a ML-model trained on historical data. However, if the training dataset is similarly skewed, the classifier will naturally minimize total error by predicting the majority class more frequently, since a greater chance of misclassifying minority cases contributes little to the overall total error. As a result, the classifier systematically mislabels minority observations, even when these are the most important to identify. To alleviate this issue, solution as generally presented on two levels, the data-level and the algorithm-level.

Data-level methods rebalance the training set so that the classifier is exposed, during training, to a more even distribution of majority and minority cases. This rebalancing includes undersampling, oversampling, and hybrid approaches.

- Undersampling removes majority-class examples to reduce dominance, but risks discarding important information.
- Oversampling increases the representation of minority data, often by interpolation or synthetic generation, but can lead to overfitting where there the model becomes dependent on the training data and fails when unseen data is introduced.
- Hybrid methods attempt to combine both techniques to balance information loss with risk of overfitting to minority samples.

Algorithm-level methods keep the data unchanged but modify the learning mechanism. A common technique is to alter the cost function, the metric used to determine the performance, so that misclassifying minority examples incurs a higher cost. This forces the model to learn a decision boundary that is more sensitive to the minority class and counteracts the natural bias toward the majority. These methods are often more application-specific and harder to implement but are often more resource effective compared to rebalancing. Importantly, data-level and algorithm-level strategies are not mutually exclusive and can be combined for better performance.

Even after data has been assigned to minority and majority classes, a second imbalance problem arises in avoiding a majority class bias when training models that use this data [19], [20]. This may for example be in failure forecasting for predictive maintenance, where failures are rare events, and the minority data may contain disproportionately more useful information. However, standard ML methods often fail to predict or quantify such events because the model learns normal behaviour well but easily ignore rare events [19].

Because this challenge is fundamentally similar to the class-imbalance problem in classification, the solutions are also analogous. At the data-level, the training set used for the predictive model can be rebalanced so that failure-related patterns appear more frequently, increasing their influence during training [20]. At the algorithmic-level, the cost function can be modified to assign higher weight to errors associated with the minority class, forcing the model to emphasize identifying these rare but critical conditions [20].

Energiforsk

5 Application and vendor study

This chapter provides a review of the tools, platforms, and vendors utilised for big data management and analytics within safety-critical industries. Drawing upon both the literature and empirical findings from industry interviews, the chapter compares available solutions, describes practical experiences, and discusses key considerations relevant to the selection and implementation of big data technologies.

The interview-based insights presented in this chapter are based exclusively on the interviews with Bruce Power, INEOS Inovyn and Borealis, and do not include input from TVO or Ringhals.

5.1 INSIGHTS FROM LITERATURE

5.1.1 Existing solutions for big data management delivery

Storing, managing and interpretating the vast amount and varied data produced by NPPs is both challenging and time-consuming [21]. Effective methods for big data management are therefore necessary without compromising on data security, where the physical location of data and computation hardware becomes an important consideration. The following are a description of different methods for managing big data with different benefits and drawbacks.

On-premise solutions
On-premise deployment is the preferred approach for strict security and aligns well with regulatory and cybersecurity requirements, as it allows full control over data and infrastructure [8]. Systems hosted on-site can operate independently of an internet connection and maintain safe operability even when disconnected from external networks. This approach, however, comes with high hardware requirements and demands resource allocation for continuous maintenance and management, such as acquiring and installing additional hardware when, for example performing a storage upgrade.

Cloud-based solutions
Cloud computing provides dynamic scalability and cost efficiency, reducing the need for on-site hardware [7]. By leveraging web-connected servers, users can access resources from anywhere with an internet connection. Cloud-based systems are available in two primary forms: public and private clouds.

- Public cloud
  Public clouds offer virtually unlimited storage and computing resources, along with ready-to-use managed services such as scalable storage and machine learning pipelines [7]. There is no hardware management, eliminating the need for management and maintenance by the operator [7]. However, the reliance on third-party providers and the need for the data to leave the premise to be process by the third-party introduces networking and privacy risks, making public clouds unsuitable for sensitive data [7], [22].

- Private Cloud
  A private cloud is built by and operated for a specific operator or organisation, offering enhanced security, control, and governance [9], [22]. This approach ensures secure data storage, controlled access, and tighter infrastructure governance. However, private clouds are less scalable compared to public alternatives and typically involve higher costs due to the need for infrastructure investments and ongoing maintenance [22]. It can be hosted on-premises, at a dedicated off-site data centre, or managed by a third-party provider, with the different deployment method offering varying amount of security in exchange for increased cost and reduced flexibility.

Edge computing
Edge computing brings data processing closer to the source, such as sensors, rather than transmitting all data to centralised servers [23], [24]. This approach reduces latency as it reduces transmission delays and improves bandwidth efficiency as less data needs to be transmitted to the centralised servers. This helps enable real-time processing. Additionally, data is processed locally which enhances security and reducing data transmission minimise the risk of leakage during transmission. However, the distributed nature of edge networks makes them more vulnerable to physical tampering. Further, Edge devices often have limited computing and storage capabilities. Coordinating a large number of devices can also be complex and the initial setup costs can be significant.

Hybrid approaches
Hybrid solutions combine some of the advantages of both on-premise and cloud solutions. In such setups, sensitive data is processed locally, while less critical workloads are offloaded to cloud environments where security requirements are more flexible. This approach allows for resource sharing between on-premise and external infrastructures, providing a balance between performance, scalability, and data protection [22]. However, hybrid systems introduce challenges such as complex security management, which require careful integration to not compromise on data security [22].

5.1.2 Tools, applications and vendors providing big data management solutions

NPPs incorporates a varied knowledge from many different fields and can therefore derive value from many different applications of big data. In addition, there are often many different tools available for different applications within big data. This section will present tools available for different applications and management of big data as well as different vendors operating in the presented categories.

Computer vision (CV)
CV is an application within big data where computers process visual data such as images and videos. Several use cases for computer vision in NPPs has been presented including automated anomaly detection of critical components, assistance in monitoring inaccessible areas and digitalisation of older documents [12]. A range of tools exist to support these applications. For instance, Trueflaw and EPRI have collaborated on an, non-destructive ultrasonic inspection system utilising edge-processing witch highlights potential defects to assist inspectors,

making their jobs faster and more reliable. Field trials were conducted in April 2022 at a U.S. NPP [25], and the system has also been deployed at Ringhals [26].

Other tools include Anomalib, an open-source library for anomaly detection and localisation with algorithms that can be used off-the-shelf or tailored towards specific needs [27] and Clip4Str, an optical character recognition model based on OpenAI's CLIP model with multi-modal capability, meaning it can process multiple types of input data [12]. Similarly, YOLO (You Only Look Once) provides real-time object detection capable of identification of multiple objects [28]. Another example is LandingLens, a computer vision platform developed by Landing AI for quality inspection and defect detection. It features of the shelf use without prior experience in ML, but has a closed nature with limited access to specific internal parameters [29].

Natural language processing

Natural language processing is an application within big data where machines interpret and generate human natural language. A typical implementation can involve a LLM together with RAG to search, summarise, and reason over internal documentation. Presented use cases include analysing maintenance and failure reports to asses root causes and assist in maintenance planning [12]. The Viking and Poro LLM are open-source multilingual large language models developed by a collaboration between Silo AI, the University of Turku and High Performance Language Technologies [30], [31]. They are developed as an effort to create models for languages other than English and Poro has been trained on both Finnish and English, while Viking also include the other Scandinavian languages.

Visualisation

Visualisation involves presenting the human-in-the-loop with the information of the big data analytics in a useful format to help them make better informed decisions [7]. Visualisation platforms provide dashboards and interactive analytics for operations, maintenance, and inspection of data. One solution is Grafana, an open-source platform for monitoring and visualisation that can integrate data from time-series sources and Structured Query Language (SQL) databases [32]. It enables users to create dashboards, visualize metrics, and configure alerting thresholds. Beyond the open-source version, Grafana is also available in cloud-based and enterprise editions. Another example is IBM BigSheets, a web-based analytics enables non-technical users to explore and analyse big data through a familiar spreadsheet-like interface [3].

Cloud platforms

As presented, cloud solution can help in providing a scalable big data management method and can be separated in public cloud and private clouds where the private cloud offers a higher degree of security in exchange of a higher cost and reduces flexibility. For the public cloud, Microsoft Azure, Amazon Web Service and Google Cloud are three leading cloud platforms.

Within private cloud environments, virtualisation technologies such as VMware and VirtualBox enable the pooling of central computing resources including storage, processing power, and networking into multiple virtual servers [33]. These virtual environments then dynamically allocate resources based on demand.

Another component of the private cloud big data infrastructure is the database, a structured system for storing, managing, and retrieving data [34]. Databases are generally divided into two main categories, relational SQL databases and non-relational NoSQL databases. SQL databases are optimised for consistent and structured data while NoSQL databases are designed to handle large volumes of unstructured or semi-structured data, making them particularly suitable for big data applications [34]. Examples of SQL databases are PostgreSQL and MySQL, while examples of NoSQL databases include MongoDB and Apache Cassandra.

In addition, for analysing the large datasets, distributed computing frameworks like Apache Hadoop and Apache Spark can be used [35]. Apache Hadoop, developed by the Apache Software Foundation, is an open-source framework for storing and processing structured and unstructured data across distributed environments, offering a high degree of scalability [35]. Apache Spark, which operates on top of Hadoop, is also open source and provides a unified analytics engine for large-scale data processing, supporting a wide range of applications and integration with diverse data storage systems [35]. The petrochemical industry, for example, which is utilizing big data analysis throughout various parts of the proses leverage big data platforms such as Apache Hadoop and Apache Spark alongside machine ML to process seismic datasets for subsurface imaging and fracture modelling [3], [16].

Enterprise asset management

Enterprise asset management is an application of big data to manage assets throughout its lifecycle, including planning, procurement, installation, maintenance and disposal [36]. Goals with the asset management include to maximise reliability and to minimise downtime and maintenance costs. Modern system often incorporates predictive maintenance through ML and visual inspections through computer vision. There are several vendors of enterprise asset management including, IBM’s Maximo [37], Hitachi’s Ellipse EAM [38] and Asset Suite EAM [39], Hexagon EAM [40] and Oracle eAM [41]. Further, Hitachi’s Asset Suite EAM is designed for mission-critical plant environments, while IBM’s Maximo Application Suite also offer Maximo for Nuclear Power Plants to target industry-specific challenges [37], [39].

5.2 INSIGHTS FROM INTERVIEWS

There is a clear shift away from relying on a single platform to meet all analytics requirements. Instead, organisations are increasingly constructing modular ecosystems composed of specialised tools, which often combine both commercial products and proprietary solutions. Honeywell and AspenTech are cited as preferred choices for process control and simulation applications. For plant-wide data integration and monitoring, OSIsoft PI (now AVEVA PI) and GE Digital are utilised. Meanwhile, engineers continue to favour Excel, MATLAB, Python, and Minitab for ad hoc analyses and prototyping tasks.

While cloud platforms are used for analytics and decision support, on-premise solutions remain the preferred option for critical operations, owing to concerns about security and reliability. At Borealis, for example, an IoT platform has been implemented so that factory-specific data is stored locally in parallel with the analytics platform.

External cloud platforms have mainly been used for analytics, decision support, and pilot projects, where they offer flexibility and enable global access to data and tools. The main motivations for using external cloud solutions are to quickly test new analytics capabilities and to facilitate collaboration across multiple sites. However, their use is limited to non-critical applications due to security, data sovereignty, and existing infrastructure considerations.

Furthermore, there is a trend towards the development and deployment of custom microservices and in-house solutions, especially in legacy plants, as a means to avoid vendor lock-in and to customise analytics capabilities to better suit specific operational needs.

In Table 1 software for data analytics that are being used by the companies in the interview study are being listed.

Table 1. Software tools and platforms used for data analytics and process optimisation in safety-critical industries. Based on interview results.

Software / Platform	Description	Applications
Honeywell UniSim Design	Advanced simulation, hybrid modelling	Used for combining simulation and real process data for model calibration
AspenTech	Process optimisation, advanced control, simulation	Used for real-time optimisation
OSIsoft PI (now AVEVA PI)	Time-series data historian, plant data integration	Frequently mentioned for plant-wide data collection and integration
GE Digital	Data integration, monitoring, analytics	Used for plant monitoring and diagnostics
Excel, Minitab, MATLAB, Python	Offline statistical analysis, modelling, data science	Used by engineers for ad hoc analysis, modelling, and prototyping
Custom / In-house Solutions	Tailored analytics, microservices, integration	Modular, custom-built microservices for specific problems
Altair Analytics Workbench	Data processing, quality analytics	Used for flexible, user-driven analytics and prototyping
Orise Digital AI	Anomaly detection, predictive analytics, process optimisation, asset health monitoring	Used in process industries; integrates with major industrial data platforms
Yanomaly asset health monitoring	Industrial anomaly detection, predictive modelling	Used for deploying and maintaining AI solutions

6 Implementation considerations and lessons learned

This chapter presents lessons learned on the use of big data in safety critical industries and discusses key considerations for implementing big data solutions in NPPs. It addresses technical, organisational, and regulatory factors that influence successful adoption. The findings are based on a review of published literature as well as interviews with industry representatives.

The interview-based insights presented in this chapter are based exclusively on the interviews with Bruce Power, INEOS Inovyn and Borealis, and do not include input from TVO or Ringhals.

6.1 INSIGHTS FROM LITERATURE

The following lessons learned are derived from the literature review, focusing on secure and effective implementation of big data systems in safety-critical industries.

Big data adoption is still in early stages in NPPs with limited real-world deployment. Its deployment is, however, being increasingly considered as it can provide value for different aspects of the NPPs including fault prevention, predictive maintenance, optimisation and decision support [9], [15]. As NPPs are a safety critical industry, secure and robust implementation of these systems are essential. A study of the aerospace and aviation industry, as well as the petrochemical industry, brings up several issues regarding how big data analytics should be managed. By further studying NPP specific information on how big data should be implemented as well as generic recommendations for secure big data applications, many similar points are brought up.

A fundamental aspect is data quality and availability. Data in industrial environments can often be noisy, incomplete, or inconsistent, which makes cleaning strategies such as outlier detection, interpolation, and standardisation important for reliable model inputs where poor-quality or biased data can lead to mispredictions and errors [3], [42]. Of importance is also feature selection, as choosing the right input parameters improves model accuracy and reduces computational load. Techniques like SHAP (SHapley Additive exPlanations) and VIF (Variance Inflation Factor) are useful for identifying relevant features and mitigating multicollinearity in high-dimensional sensor data involving correlated sensor readings, thereby improving model stability [13].

Interpretability and explainability are other considerations for safety-critical systems. Models that operate as black boxes are unsuitable for nuclear applications, where transparency and explainability is essential [9], [14], [43]. SHAP values can help explain model decisions and validate trustworthiness [13]. In addition, complexity should be minimized, and big data solutions should not replace simpler systems unnecessarily. Further, breaking large solutions down into smaller modules improves explainability, maintainability, and fault isolation [14],

[42], [43]. Lifecycle management is another critical factor where frequent retraining and reconfiguration introduces risks of bias or overfitting [42], [43]. Static validation may also be insufficient as system behaviour may evolve over time, requiring continuous monitoring for anomalies and model drift.

Developing a human-centred design is also of importance to ensure that complex outputs are translated into actionable insights through user-friendly interfaces and dashboards. Clear distinctions between measured and predicted information are necessary, as predictive outputs carry uncertainty [3], [7]. The risk of overreliance on automation resulting in operator complacency should also be considered throughout. Big data should augment, not replace, human decision-making, and operators require training to work effectively with these systems, identify failures, and follow intervention protocols [12], [21], [43]. Successful implementation also depends on collaboration across disciplines, bridging domain expertise with data scientists and IT specialists [3].

When having distributed assets, processing data closer to the source can reduce latency and manage large, fast-moving streams more effectively [3], [16]. However, big data should not be used for autonomous control, emergency decision-making, or safety-critical operations [15]. Its role should remain supportive, enhancing preventive measures such as predictive maintenance while keeping the human in the loop [15], [43].

NPPs have long held a methodical prosses when it comes to change ensuring strict adherence to security, however the rapid developments in big data can at times make security a secondary concern [42], [43]. Big data expands the attack surface as new vulnerabilities can be exploited to deliberately deteriorate model performance or enable the extraction of sensitive model information. Big data integration should therefore not be rushed, and a secure infrastructure and cybersecurity are vital.

Finally, regulatory gaps further complicate adoption. No unified standards exist for big data in nuclear systems. Current approaches rely on low-level software standards [43]. In the Nordics, regulations from a national and EU level are unspecific regarding technologies used in NPPs. Operators bear the responsibility for establishing IT security teams to evaluate and approve new technologies [12].

6.2 INSIGHTS FROM INTERVIEWS

Successful implementation of big data and AI solutions in safety-critical industries is shaped as much by organisational and human factors as by technical considerations. Across both literature and interviews, there is strong consensus that the most effective approach is to begin with small, well-defined pilot projects, engaging end users early to build trust and gather practical feedback. Change management is often a significant challenge, as staff accustomed to established routines may resist new digital tools; involving motivated “change agents” and prioritising user experience can help overcome these barriers.

When introducing new analytics or AI-driven systems, it is essential to focus on practical applications that deliver actionable insights for technicians and engineers. Selecting use cases where instrumentation is already sufficient and the business case is strong is preferable to attempting broad, generic solutions from the outset. Early successes help build momentum and expertise, enabling gradual expansion to more complex systems and assets. Supplier-provided models are often effective for standard equipment, while in-house solutions may be more suitable for unique, critical, or complex components.

Technologies such as CV and advanced analytics have demonstrated clear value in faster fault detection, improved planning, and reduced downtime. Proven supplier models or ML-tools for anomaly detection and condition monitoring enable rapid deployment and measurable results.

Other recurring lessons include the importance of integrating legacy systems, ensuring high data quality, and avoiding unnecessary data reduction. Security, regulatory compliance, and data sovereignty also remain central considerations, with most organisations preferring on-premises or sovereign cloud solutions for critical systems, while leveraging modular, hybrid approaches for analytics and decision support. Notably, there is a strong consensus that AI and analytics should support, not replace, human decision-making, especially in critical operations, and that keeping the "human in the loop" is essential for both safety and acceptance.

6.3 LIMITATIONS REGARDING RECOMMENDATIONS ON SOFTWARE TOOLS AND VENDORS

While the original project proposal anticipated that the study would result in concrete recommendations regarding the most suitable software tools, applications, and vendors for big data analytics and decision support in Nordic NPPs, the findings from the interviews and the broader analysis indicate that such recommendations are neither practical nor meaningful within the current context.

A recurring theme in the interviews with industry representatives was the importance of adapting software solutions to the specific problem at hand. The suitability of a particular tool or vendor depends on a range of factors, including the nature of the operational challenge, the type and quality of available data, existing IT infrastructure, organisational requirements, and security considerations. As highlighted in Section 6.2, there is a clear consensus that no single platform or vendor can address all needs; instead, organisations benefit from constructing modular ecosystems of specialised tools, often combining commercial products with in-house solutions, tailored to their unique circumstances.

Furthermore, the selection of vendors is highly context-dependent. The interviews pointed out that the choice of vendor is influenced not only by technical requirements but also by factors such as legacy system compatibility, data sovereignty, regulatory compliance, and the organisation's long-term strategic goals. Because of this, it was not feasible to find and interview vendors or to recommend specific suppliers, as the optimal choice will vary significantly between organisations and use cases.

7 Summary and recommendations

This report has mapped the current state and future potential of big data applications in Nordic NPPs. Through literature review and interviews with industry representatives, it provides an overview of progress to date, key challenges, and practical experiences from within the sector. Lessons from other safety-critical industries confirm that a stepwise, user-focused approach is most effective.

Drawing on insights from both the literature review and interviews, the following practical steps are recommended to help Nordic NPPs make effective use of their big data resources:

1. Start with focused pilot projects
    - Select one or a few well-defined pilots with clear business value and measurable outcomes.
    - Prioritise areas where data and instrumentation are already available, and where results can be demonstrated quickly.
2. Define clear workflows for actionable insights
    - For every analytics or decision support initiative, establish a clear workflow that specifies who receives the information, what decisions need to be made, and who is authorised to act.
    - Avoid building dashboards or analytics tools without a plan for how insights will be used in practice. Even if the full solution is not clear from the outset, there must be a defined process for turning analysis into action.
3. Build modular and flexible tool ecosystems
    - Avoid seeking a single, all-encompassing analytics platform. Instead, develop an ecosystem of specialised tools that can be integrated and adapted as needs evolve.
    - Leverage both commercial and in-house solutions and focus on interoperability between systems.
4. Engage end users early and continuously
    - Involve operators, engineers, and maintenance staff from the outset to ensure solutions address real operational needs.
    - Use feedback from end users to refine tools and processes, and to drive acceptance and adoption.
5. Develop cross-functional teams

- Create cross-functional teams that combine domain expertise, data science and IT to bridge knowledge gaps and accelerate implementation.

6. Plan for scale from the outset
   - While starting small, ensure that solutions, data structures, and processes are designed with scalability in mind.
   - Consider how successful pilots can be expanded or replicated across other units or plants.
7. Maintain a pragmatic approach to change management
   - Identify and empower "change agents" who can champion digital initiatives and support colleagues through transitions.
   - Address resistance proactively by demonstrating value and providing ongoing support and training.

7.1 PRACTICAL RECOMMENDATIONS

To effectively begin leveraging big data, a Nordic NPP should start small, focus on a clearly defined use case, and build modular solutions that can be expanded as experience grows.

A targeted predictive maintenance pilot on a component or subsystem with sufficient historical and process data constitutes a suitable initial approach. Select an asset where failures have operational impact and instrumentation is already in place.

Before starting, define who will receive the analysis results, what actions are required, and who is responsible for execution. The output must be concrete and actionable for maintenance staff, specifying both what needs to be done and when.

Choose analytics software based on the specific problem and data available, there is no universal solution. Form a cross-functional team where domain experts and data scientists work closely together continuously throughout the project. This collaboration ensures that data is interpreted correctly and that solutions are tailored to real operational needs. Use supplier-provided models for standard equipment or custom tools for unique assets as appropriate. If necessary, integrate data from multiple sources (e.g., process signals, maintenance logs, inspection records) to ensure a complete basis for analysis; addressing data silos may be required for success.

If using AI or ML, start with unsupervised anomaly detection on existing data to quickly identify deviations. As labelled events accumulate, refine the approach with supervised methods. Ensure outputs are interpretable and integrated into existing maintenance workflows.

When selecting delivery models for any big data solutions, it is recommended to prioritise on-premise options for critical systems, as these provide greater control over data. It is, however, essential to maintain strict data classification and access controls to ensure that only authorised personnel can access sensitive information.

8 References

[1] D. A. Ejigu, Y. Tuo, and X. Liu, ‘Application of artificial intelligence technologies and big data computing for nuclear power plants control: a review’, *Front. Nucl. Eng.*, vol. 3, Feb. 2024, doi: 10.3389/fnuen.2024.1355630.

[2] D. T. Dagan and E. J. Wilkins, ‘What is “big data” and how should we use it? The role of large datasets, secondary data, and associated analysis techniques in outdoor recreation research’, *J. Outdoor Recreat. Tour.*, vol. 44, p. 100668, Dec. 2023, doi: 10.1016/j.jort.2023.100668.

[3] M. Mohammadpoor and F. Torabi, ‘Big Data analytics in oil and gas industry: An emerging trend’, *Petroleum*, vol. 6, no. 4, pp. 321–328, Dec. 2020, doi: 10.1016/j.petlm.2018.11.001.

[4] Ishwarappa and J. Anuradha, ‘A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology’, *Procedia Comput. Sci.*, vol. 48, pp. 319–324, Jan. 2015, doi: 10.1016/j.procs.2015.04.188.

[5] K. N. Singh, R. K. Behera, and J. K. Mantri, ‘Big Data Ecosystem: Review on Architectural Evolution’, in *Emerging Technologies in Data Mining and Information Security*, A. Abraham, P. Dutta, J. K. Mandal, A. Bhattacharya, and S. Dutta, Eds, Singapore: Springer, 2019, pp. 335–345. doi: 10.1007/978-981-13-1498-8_30.

[6] V. Sheokand and V. Singh, ‘Modeling Data Heterogeneity Using Big DataSpace Architecture’, in *Advanced Computing and Communication Technologies*, R. K. Choudhary, J. K. Mandal, N. Auluck, and H. A. Nagarajaram, Eds, Singapore: Springer, 2016, pp. 259–268. doi: 10.1007/978-981-10-1023-1_26.

[7] C. Walker, V. Agarwal, P. Ramuhalli, and N. Lybeck, ‘Development of an End State Vision to Implement Digital Monitoring in Nuclear Plants’, *Annu. Conf. PHM Soc.*, vol. 14, no. 1, Oct. 2022, doi: 10.36001/phmconf.2022.v14i1.3176.

[8] M. T. Bensi and K. M. Groth, ‘On the value of data fusion and model integration for generating real-time risk insights for nuclear power reactors’, *Prog. Nucl. Energy*, vol. 129, p. 103497, Nov. 2020, doi: 10.1016/j.pnucene.2020.103497.

[9] C. Jendoubi and A. Asad, ‘A Survey of Artificial Intelligence Applications in Nuclear Power Plants’, *IoT*, vol. 5, no. 4, pp. 666–691, Dec. 2024, doi: 10.3390/iot5040030.

[10] U. Himanshu, L. Lagos, S. Joshi, M. Esoofally, and K. Cooper, ‘Predictive analytics with big data-spark framework’, *Nucl. Plant J.*, vol. 36, pp. 29–31, Aug. 2025.

[11] P. Ramuhalli, C. Walker, V. Agarwal, and N. Lybeck, ‘Development of Prognostic Models Using Plant Asset Data’, ORNL/TM-2020/1697, 1661211, Sept. 2020. doi: 10.2172/1661211.

[12] L. Sütfeld and A. Thore, ‘On-Premise AI Solutions for Nordic Nuclear Applications’, Energiforsk AB, Stockholm, Sweden, Research Report 2025:1093, Mar. 2025. [Online]. Available: https://energiforsk.se/media/34237/2025-1093-on-premise-ai-solutions-for-nordic-nuclear-applications.pdf

[13] C. Walker, P. Ramuhalli, V. Agarwal, N. Lybeck, and M. Taylor, 'Development of Short-Term Forecasting Models Using Plant Asset Data and Feature Selection', *Int. J. Progn. Health Manag.*, vol. 13, June 2022, doi: 10.36001/ijphm.2022.v13i1.3120.
[14] K. Ranasinghe *et al.*, 'Advances in Integrated System Health Management for mission-essential and safety-critical aerospace applications', *Prog. Aerosp. Sci.*, vol. 128, p. 100758, Jan. 2022, doi: 10.1016/j.paerosci.2021.100758.
[15] S. Pandey, 'Machine learning and big data applications for nuclear power plant monitoring', June 2024.
[16] Y. Xu, M. Xing, M. C. Chen, M. Xuan Zou, and W. J. Huang, 'Petrochemical industry digital transformation from the perspective of big data : A survey', in *2022 International Conference on Frontiers of Communications, Information System and Data Science (CISDS)*, Guangzhou, China: IEEE, Nov. 2022, pp. 14–20. doi: 10.1109/CISDS57597.2022.00010.
[17] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, *Learning from Imbalanced Data Sets*. Cham: Springer International Publishing, 2018. doi: 10.1007/978-3-319-98074-4.
[18] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, 'Learning from class-imbalanced data: Review of methods and applications', *Expert Syst. Appl.*, vol. 73, pp. 220–239, May 2017, doi: 10.1016/j.eswa.2016.12.035.
[19] S. Rudy and T. Sapsis, 'Output-weighted and relative entropy loss functions for deep learning precursors of extreme events', Dec. 01, 2021, *arXiv*: arXiv:2112.00825. doi: 10.48550/arXiv.2112.00825.
[20] C. Shyalika, R. Wickramarachchi, and A. Sheth, 'A Comprehensive Survey on Rare Event Prediction', Oct. 05, 2024, *arXiv*: arXiv:2309.11356. doi: 10.48550/arXiv.2309.11356.
[21] D. A. Ejigu, Y. Tuo, and X. Liu, 'Application of artificial intelligence technologies and big data computing for nuclear power plants control: a review', *Front. Nucl. Eng.*, vol. 3, Feb. 2024, doi: 10.3389/fnuen.2024.1355630.
[22] L. Qian, Z. Luo, Y. Du, and L. Guo, 'Cloud Computing: An Overview', in *Cloud Computing*, M. G. Jaatun, G. Zhao, and C. Rong, Eds, Berlin, Heidelberg: Springer, 2009, pp. 626–631. doi: 10.1007/978-3-642-10665-1_63.
[23] S. Liang, S. Jin, and Y. Chen, 'A Review of Edge Computing Technology and Its Applications in Power Systems', *Energies*, vol. 17, no. 13, p. 3230, Jan. 2024, doi: 10.3390/en17133230.
[24] K. Cao, Y. Liu, G. Meng, and Q. Sun, 'An Overview on Edge Computing Research', *IEEE Access*, vol. 8, pp. 85714–85728, 2020, doi: 10.1109/ACCESS.2020.2991734.
[25] Electric Power Research Institute (EPRI), 'AI Tool Developed by EPRI Significantly Cuts Analysis Time in U.S. Nuclear Plant Field Trial'. Accessed: Oct. 27, 2025. [Online]. Available: https://www.epri.com/research/products/000000003002025510
[26] T. Ltd, 'Nuclear Industry First AI Powered Ultrasonic Inspection in Ringhals, Sweden'. June 2025. [Online]. Available: https://www.ndt.net/search/docs.php3?id=31354
[27] S. Akcay, D. Ameln, A. Vaidya, B. Lakshmanan, N. Ahuja, and U. Genc, 'Anomalib: A Deep Learning Library for Anomaly Detection', in *2022 IEEE*

International Conference on Image Processing (ICIP), Oct. 2022, pp. 1706–1710. doi: 10.1109/ICIP46576.2022.9897283.

[28] A. Nazir and Mohd. A. Wani, 'You Only Look Once - Object Detection Models: A Review', in *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)*, Mar. 2023, pp. 1088–1095.

[29] F. M. Segura, F. P. Segura, M. P. L. Zudaire, and F. V. Segura, 'Advances in Artificial Intelligence for automated knee osteoarthritis classification using the IKDC system', *Eur. J. Orthop. Surg. Traumatol.*, vol. 35, no. 1, p. 32, Dec. 2024, doi: 10.1007/s00590-024-04124-0.

[30] R. Luukkonen, J. Burdge, E. Zosa, V. Komulainen, P. Sarlin, and S. Pyysalo, 'Viking: A Family of Nordic LLMs'. Accessed: Oct. 27, 2025. [Online]. Available: https://huggingface.co/LumiOpen/Viking-33B

[31] R. Luukkonen *et al.*, 'Poro 34B and the Blessing of Multilinguality', *ArXiv Prepr. ArXiv240401856*, Apr. 2024, [Online]. Available: https://arxiv.org/abs/2404.01856

[32] M. Chakraborty and A. P. Kundan, 'Grafana', in *Monitoring Cloud-Native Applications: Lead Agile Operations Confidently Using Open Source Software*, M. Chakraborty and A. P. Kundan, Eds, Berkeley, CA: Apress, 2021, pp. 187–240. doi: 10.1007/978-1-4842-6888-9_6.

[33] P. Chen, X. Chen, J. Xie, W. Xiong, and T. Yu, 'Design and implementation of cloud platform for nuclear accident simulation', *Front. Energy Res.*, vol. 10, Jan. 2023, doi: 10.3389/fenrg.2022.1075224.

[34] A. Meier and M. Kaufmann, *SQL & NoSQL Databases: Models, Languages, Consistency Options and Architectures for Big Data Management*. Wiesbaden: Springer Fachmedien, 2019. doi: 10.1007/978-3-658-24549-8.

[35] P. Sewal and H. Singh, 'A Critical Analysis of Apache Hadoop and Spark for Big Data Processing', in *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC)*, Oct. 2021, pp. 308–313. doi: 10.1109/ISPCC53510.2021.9609518.

[36] N. Gallagher, 'What is Enterprise Asset Management (EAM)?' Accessed: Oct. 27, 2025. [Online]. Available: https://www.ibm.com/think/topics/enterprise-asset-management

[37] IBM Corporation, 'Energy and Utilities Asset Management Software with IBM Maximo Application Suite'. Accessed: Oct. 27, 2025. [Online]. Available: https://www.ibm.com/products/maximo/energy-utilities

[38] Hitachi Energy Ltd, 'Ellipse EAM'. Accessed: Oct. 27, 2025. [Online]. Available: https://www.hitachienergy.com/products-and-solutions/asset-and-work-management/lumada-eam/ellipse-eam

[39] Hitachi Energy Ltd, 'Asset Suite EAM'. Hitachi Energy Ltd, 2021. Accessed: Oct. 27, 2025. [Online]. Available: https://publisher.hitachienergy.com/preview?DocumentID=9AKK106930A8154&LanguageCode=en&DocumentPartId=A4-web&Action=Launch

[40] Hexagon AB, 'Enterprise Asset Management Solutions'. Accessed: Oct. 27, 2025. [Online]. Available: https://hexagon.com/solutions/enterprise-asset-management

[41] Oracle Corporation, 'Oracle® Enterprise Asset Management User's Guide'. Accessed: Oct. 27, 2025. [Online]. Available: https://docs.oracle.com/cd/E18727_01/doc.121/e13670/T259967T259970.htm

[42] UK National Cyber Security Centre (NCSC) and US Cybersecurity and Infrastructure Security Agency (CISA), 'Guidelines for Secure AI System Development', Oct. 2023. [Online]. Available: https://www.ncsc.gov.uk/files/Guidelines-for-secure-AI-system-development.pdf

[43] K. Lee, A. White, D. Finnigan, and M. Dennis., 'Considerations for Developing Artificial Intelligence Systems in Nuclear Applications', Sept. 2024. [Online]. Available: https://www.ncsc.gov.uk/files/Guidelines-for-secure-AI-system-development.pdf

Appendix

The Appendix provides a list of all interviews conducted for this study, together with the names of the participating organisations and brief summaries of each interview.

1. RINGHALS - NORDIC NPP OPERATOR

Date: September 16, 2025

Participants:

- Interviewer: Tommie Lindquist
- Jonas Olandersson
- Magnus Nilsson

Ringhals is an NPP operatator in Sweden

The interview with representatives from Ringhals focused on the company's use of big data within their nuclear power plants. Ringhals handles most data in a siloed manner and most analysis is carried out manually.

1. Data collection: Ringhals collects data from various sources, including process parameters that are continuously logged at a relatively low frequency (typically on a minute scale). This data is mainly used for manual analysis to investigate events and compare them with historical situations. It is also used to validate certain types of models created for analysis purposes.
2. Data processing: One of the key challenges is the conservative approach to processing the collected data.
3. Data reliability: Ensuring data reliability and accuracy is another significant challenge. Ringhals has implemented some data validation processes, including regular calibration of sensors and cross-referencing data from multiple sources to identify and correct discrepancies.
4. AI and anomaly detection: Ringhals has experimented with AI solutions and anomaly detection models in collaboration with Caverion, which have shown promising results. However, these models have been paused temporarily due to challenges related to explainability and the complexity of training the models

2. TVO - NORDIC NPP OPERATOR

Date: September 15, 2025

Participants:

- Interviewer: Tommie Lindquist

- Marko Savolainen
- Timo Vaahtera
- Jani Manninen
- Mauri Viitasalo

Teollisuuden Voima Oyj (TVO) is a Finnish energy company that owns and operates nuclear power plants in Finland.

The interview with representatives from TVO focused on the use of big data in the nuclear industry. The participants described how TVO collects and analyses data from various sources, including sensors, inspection data, and maintenance data. They have recently automated the data collection process from plants to offices to enable continuous analysis, which has significantly improved their ability to monitor and optimise plant operations.

TVO is exploring the use of big data for predictive maintenance, fault detection, process optimisation, and lifecycle management. They are also testing different analytical tools to determine the best use of the collected data. However, they face several challenges in implementing big data solutions:

1. Integration of Legacy Systems: One major challenge is integrating data from older systems where the data formats are not time-based or standardised. This makes it difficult to combine and analyse data from different sources effectively.
2. Data Reliability and Accuracy: Ensuring the reliability and accuracy of data is another significant challenge. TVO has to make sure that the data collected from various sensors and systems is trustworthy and has not been tampered with. This includes maintaining accurate time-stamping and synchronisation across different systems.
3. Organisational Roles and Responsibilities: Defining clear organisational roles and responsibilities for data management is still a work in progress. While some areas have well-defined roles, the overall picture remains somewhat disorganised, which can hinder effective data governance.
4. Common Tools for Data Analysis: TVO is currently in the testing phase for various analytical tools and has not yet standardised on a common tool for data analysis. Different parts of the organisation have different needs and preferences, making it challenging to find a one-size-fits-all solution.
5. Cybersecurity and Regulatory Compliance: As TVO moves towards testing more automated and cloud-based solutions, ensuring cybersecurity and compliance with regulatory requirements becomes increasingly important. They need to manage what data is stored in the cloud and ensure that it is protected from unauthorised access.
6. Data Storage and Management: Managing the increasing volume of data and deciding how long to store it is another challenge. While TVO collects a vast amount of data, they need to develop strategies for data archiving

and ensuring that only relevant data is retained for long-term analysis. In some cases, TVO saves data in the form of averages instead of actual samples to conserve disk space.

3. BRUCE POWER - CANADIAN NPP OPERATOR

Date: October 23, 2025

Participants:

- Interviewer: Tommie Lindquist
- Nick Torenvliet

Bruce Power is a Canadian company that operates the world's largest nuclear power facility in Ontario, Canada. The company is recognized for its leadership in digitalization and advanced data analytics in nuclear power operations.

The interview with Nick Torenvliet at Bruce Power provided an in-depth look at the company's approach to big data, digitalisation, and advanced analytics in nuclear power operations. Nick shared practical experiences and lessons learned, highlighting both technical and organisational aspects that distinguish Bruce Power's practices and ongoing transformation.

1. Digitalisation and data integration: Bruce Power is undergoing a major digital transformation, focusing on digitising all plant documentation and integrating data from various sources. This includes the development of digital twins, time series analysis, and the use of computer vision and natural language processing. The company is actively working to break down data silos and create a unified, accessible data environment.

2. Advanced analytics and AI: The organisation has implemented advanced analytics for anomaly detection, predictive maintenance, and process optimisation. Nick described the development of automated analysis systems using machine learning, such as diffusion processes and variational autoencoders, which have significantly reduced analysis time and improved detection rates. Human factors and trust in automation are carefully considered, with ongoing dialogue with regulators to ensure acceptance and safety.

3. Computer vision and sensor deployment: Bruce Power uses computer vision for equipment monitoring, safety compliance (e.g., detecting hard hats), and automating operator rounds. The company is also expanding the use of vibration and other sensors, supported by an internal 3D locating system, to enable real-time monitoring and early detection of equipment issues.

4. Data storage and retention: The company strongly opposes data reduction or averaging, advocating for the retention of all raw data to preserve analytical value. While compression (e.g., Gzip) is used for storage efficiency, Bruce Power avoids any process that would result in

information loss, recognising the importance of data “noise” for uncovering hidden relationships.

5. Ecosystem of tools and microservices: Rather than relying on a single analytics platform, Bruce Power favours an ecosystem of specialised tools and microservices tailored to specific problems. This approach is seen as essential for legacy plants, while new builds could benefit from a more unified, cloud-based architecture. The company is cautious about vendor lock-in and prioritises data ownership and sovereignty.
6. Human-in-the-loop and change management: Automation is used to augment, not replace, human expertise. The “human-in-the-loop” principle is central, especially for safety-critical decisions and outage planning. Bruce Power emphasises incremental, agile development and the importance of maintaining staff engagement and domain knowledge throughout digital transformation.
7. Regulatory and security considerations: The company is attentive to regulatory requirements, data sovereignty, and national security. While cloud solutions are considered, Bruce Power prefers local or sovereign cloud providers to ensure control over sensitive data. The organisation is aware of the risks associated with external vendors and stresses the need for clear data ownership and access policies.
8. Lessons learned and recommendations: Key advice includes ensuring full access to all operational data, resisting vendor claims to data ownership, and building internal capabilities for data analytics. Bruce Power recommends starting with unsupervised methods for quick wins, gradually building labelled datasets for more advanced analytics, and always keeping the human in the loop for critical decisions. Agile, iterative development and a focus on actionable results are seen as essential for success.

4. INEOS INOVYN - EUROPEAN PETROCHEMICAL PLANTS OPERATOR

Date: September 15, 2025

Participants:

- Interviewer: Tommie Lindquist
- Joseph Rumer

INEOS Inovyn is a major European producer of petrochemicals, operating plants across Europe. The company manufactures essential chemicals used in a wide range of industries.

The interview with Joseph Rumer at INEOS Inovyn focused on the company’s approach to big data and digitalisation within their operations. Joseph shared practical experiences and lessons learned, highlighting both technical and organisational aspects.

1. Big data usage and definition: INEOS Inovyn has been collecting and analysing large volumes of process data from sensors for decades, but does not yet use “big data” in the modern sense of integrating diverse, unstructured data types or real-time analytics. Their current focus is on structured process data, with continuous monitoring and storage of thousands of measurements across all plants.

2. Key use cases: The most valuable applications are process optimisation (including root cause analysis and yield improvement) and predictive maintenance. However, widespread predictive maintenance is limited by the level of instrumentation on equipment and the infrequency of critical failures. Business cases for predictive maintenance are sometimes theoretical due to rare failure events and small sample sizes.

3. Data infrastructure and tools: Data is primarily stored on-premises in time series/historian databases. A cross-functional team of about 10 people manages data access, integration, and analytics strategy across all sites. Tools used include Excel, Minitab, MATLAB, Python, and industry-specific platforms such as Honeywell and AspenTech, with most analytics performed on-premises for historical reasons.

4. Cloud and data storage: While some pilot projects have used cloud-based analytics, there is no widespread adoption yet. Data is stored for 10–15 years, with compression based on changes in values rather than averaging. There is no strict policy against cloud storage, but on-premises remains the default.

5. AI and analytics: ML is used for process optimisation and predictive maintenance, but always as a guidance tool, never for direct control. Generative AI tools are used for basic searches, but there is no internal equivalent. Any AI-driven recommendations are reviewed by engineers before any action is taken.

6. Data quality and ownership: Ensuring data reliability is a challenge, especially when sensors fail. Data owners are responsible for verifying data before it is used in reporting. External data is reviewed using traditional supplier evaluation methods.

7. Domain knowledge and vendor selection: Successful analytics require strong domain knowledge to translate data insights into actionable information. There is no single recommended vendor; the choice depends on existing infrastructure, cybersecurity, and flexibility. Cybersecurity is always the top priority.

8. Vision and lessons learned: Looking ahead, the goal is to empower plant engineers with tools that flag anomalies across all data streams, enabling proactive investigation without requiring coding skills. The strongest lesson is the importance of building solid foundations and sufficient instrumentation before pursuing advanced analytics or machine learning. Existing data alone rarely yields new insights without the right groundwork.

5. BOREALIS – EUROPEAN PETROCHEMICAL PLANTS OPERATOR

Date: October 15, 2025

Participants:

- Interviewer: Tommie Lindquist
- Marcus Hedlund, Head of Industrial Digitalization

Borealis is an international producer of polyolefins, base chemicals, and fertilizers, with and operations across Europe and globally.

The interview with Marcus Hedlund at Borealis provides a view of how the company approaches big data, digitalisation, and advanced analytics in process industry operations. Marcus shared practical experiences and lessons learned, highlighting both technical and organisational aspects.

1. Data infrastructure and integration: Borealis has built a robust data infrastructure, integrating process data from plant sites up to a global level. Data is collected not only from process automation but also from ERP systems, laboratory systems, and other sources. The company emphasises the importance of collecting data based on actual needs rather than indiscriminately and continuously adds data to their data lake as new requirements arise.

2. Use case-driven approach: Rather than starting with technology, Borealis begins with identifying operational problems or improvement opportunities. Solutions are then developed using a mix of data-driven modelling, simulation, and process expertise. The company focuses on building scalable products and frameworks, not just ad hoc solutions, to ensure reusability across multiple plants.

3. Business value and optimisation: The main business drivers are increasing production, reducing variable costs (such as energy and raw materials), and improving product quality. Borealis uses AI models to forecast market conditions (e.g., electricity prices) and optimise production schedules. Quality prediction is a major focus, using data analytics to anticipate product properties and reduce time-consuming laboratory measurements.

4. Decision support and automation: Solutions are designed to either automate improvements (e.g., closed-loop control for energy optimisation) or provide actionable decision support to operators. The importance of defining clear workflows and decision responsibilities is emphasised, as dashboards alone are not sufficient unless they are integrated into operational processes.

5. Data quality and organisational roles: A significant portion of analytics work involves ensuring data quality. Borealis employs a central team of data engineers and data scientists who act as generalists, supporting local experts in various domains. The company uses a “hub and spoke” model, combining centralised data expertise with domain-specific knowledge from engineers and operators.

6. Ecosystem of tools, not a single platform: Borealis has moved away from seeking a single platform to solve all analytics needs. Instead, they build a modular ecosystem of specialised tools and products, integrating both in-house and commercial solutions. This approach allows flexibility and avoids vendor lock-in.
7. Cloud and on-premises solutions: Critical systems and closed-loop controls remain on-premises for security and reliability, while decision support and analytics are increasingly cloud-based to facilitate global collaboration and access.
8. Data retention philosophy: Borealis does not perform data reduction on process data, recognising that information loss can undermine future analytics. Data is typically retained for at least two major plant turnaround cycles (up to ten years), though most modelling uses data from the last two to three years for relevance.
9. Change management and user experience: Digital transformation is seen as an organisational change process. Borealis stresses the importance of involving end users early, starting small, and building on successes. User experience (UX/UI) is prioritised, especially for control room staff, to ensure solutions are intuitive and widely adopted.
10. Lessons learned and recommendations: start with clear business needs, avoid generalist-only teams for complex modelling, never reduce raw process data, build modular ecosystems rather than seeking a single platform, and focus on change management and user engagement for successful digitalisation.

USES FOR BIG DATA IN THE NORDIC NUCLEAR POWER PLANTS

This report provides a comprehensive overview of how big data could transform Nordic nuclear power plants, mapping current and emerging applications, technical and organisational requirements, and lessons learned from both the nuclear sector and other safety-critical industries. While adoption is still at an early stage, there are significant opportunities to enhance safety, efficiency, and asset management through predictive maintenance, process optimisation, and decision support.

A key finding is that, rather than relying on a single platform, organisations should develop flexible ecosystems of specialised tools that can be adapted as needs evolve. The report recommends starting with well-defined pilot projects, particular in predictive maintenance, prioritising modular and interoperable solutions, and engaging end users throughout the process. Emphasis is placed on robust data quality management, clear governance, and maintaining a strong human-in-the-loop principle. By leveraging both commercial and in-house tools and learning from international best practices, Nordic nuclear power plants can build secure, scalable, and user-oriented big data solutions that drive improvements in operational performance and long-term sustainability.

A new step in energy research

The research company Energiforsk initiates, coordinates, and conducts energy research and analyses, as well as communicates knowledge in favor of a robust and sustainable energy system. We are a politically neutral limited company that reinvests our profit in more research. Our owners are industry organisations Swedenergy and the Swedish Gas Association, the Swedish TSO Svenska kraftnät, and the gas and energy company Nordion Energi.

Energiforsk