BIG DATA GUIDELINE

AIRAH

Acknowledgements

Key contributors:
Roshan Gill, M.AIRAH
Evren Korular, M.AIRAH
Josh Wall
Danny Chak, M.AIRAH
Tristan Webber
Sean Reed, Affil.AIRAH
Chris Stamatis, M.AIRAH
Ken Thomson, F.AIRAH
Dr Stephen White, L.AIRAH
Rob Huntington, M.AIRAH
Brad Schultz, M.AIRAH
Con Tsalikis
Nicholas Lianos, Affil.AIRAH

Contact us:
Please reach out to AIRAH and the Big Data and Analytics STG for feedback and comments.
Email airah@airah.org.au

About AIRAH

AIRAH represents professionals and practitioners working in the heating, ventilation, air conditioning, and refrigeration (HVAC&R) and building services industries.

Our mission is to lead an Australian HVAC&R industry that is highly skilled, safe and sustainable. We have a proud history of more than 100 years representing, connecting, and educating professionals and practitioners who are of fundamental importance to our comfort, health, and safety.

About the Big Data and Analytics Special Technical Group

There have been a lot of questions and issues raised about big data and analytics – particularly in relation to how we proactively control and optimise building HVAC&R systems.

The Big Data and Analytics Special Technical Group brings together professionals to better understand issues, identify what needs to be done, and collaborate on combined efforts to move the industry forward.

ISBN: 978-0-949436-57-3
First edition published 2024

Contents

Abbreviations list

API – Application Programming Interface
BMCS – Building Management and Control Systems
DCH – Data Clearing House
EMS – Energy Management System
ERP – Enterprise Resource Planning
ETL – Extract Transform Load
IoT – Internet of Things
MSI – Master Systems Integrator
MSPs – Managed Service Providers
NEM – National Electricity Market
OT – Operational Technology
PLCs – Programmable Logic Controllers
ROI – Return on Investment
SFTP – Secure File Transfer Protocol

1. What is it?

Modern buildings generate a large amount of data. From air conditioning to lighting, air quality to security, a control system exists for each discipline of building services. Big data in the context of the built environment and this guide is about how we can combine and use these disparate data sets from differing sources.

As technology evolves, we are able to generate and collect data at an ever-increasing rate, but our methods of computation and analysis have remained the same. Big data involves using new tools and methods to structure, process and visualise these large and disparate data sets, with the goal of uncovering knowledge and information that we couldn't have arrived at otherwise.

The field of big data encompasses the entire life-cycle of data including measurement, collection, storage, security, transformation, analysis, automation, interpretation and presentation. It assists personnel, businesses and companies in identifying operational characteristics, developing strategies and implementing solutions that increase energy efficiency, minimises operational risk and increases occupant experience within the built environment.

For most consumers of big data, value is realised by extracting actionable insights from otherwise unstructured data. The field is well established in many industries, but applications within the built environment are relative newcomers to the use of this powerful tool. Applications in the built environment are evolving at a rapid rate as vendors and consumers continue to refine technology and preferences. Most value propositions use the data to manage risk, identify opportunities, optimise operational efficiency or all of the above.

Clients of big data often derive value from:

- Energy and water cost savings
- Energy and water consumption risk management
- Maintenance cost savings
- Rapid identification of underperforming plant and operational risks
- Tuning and optimisation
- Visibility and benchmarking of performance indicators
- Monitoring and management of indoor air quality and occupant comfort

This guide provides an overview of what big data is in the built environment, how to protect your data and how to begin your big data journey.

2. Understanding your data

In every industry, data science is playing a major role in helping businesses to optimise and digitally transform, a goal to which more and more organisations aspire. Making data accessible, meaningful, and actionable allows users to make better decisions and improve operations. The following three principles provide a good foundation:

1. Make data consumable – Data models and the insights they produce must be easily consumable by the average user. Having access to easily consumable, real-time insights, and visualisations of complex data sets can unlock new opportunities to help improve customer relationships and your bottom line.
2. Make data adaptable – Models should be self-learning and automated, so users can get the most from them. The models must learn and evolve, so the data you get is relevant to your users today and in the future. The models and data also need to be accessible through your enterprise platforms, so everyone can easily get to it.
3. Make data transparent – "Black box" solutions that hide their functions from users or that can't be re-used across the technology ecosystem are no longer acceptable. Users must be able to drill down to understand where the data behind the recommendation originated. When users understand the recommendation as well as the reasons for that recommendation, the experience is more meaningful.

2.1 – What do you need?

There are three elements to a digital strategy that can deliver value from data. Following a "data to decisions to action" process these elements include:

1. Data capture
2. Data management
3. Decision and action.

In combination, these elements empower users with access to diverse data streams that can be automatically processed and analysed to deliver new insights to the workforce.

Modern-day building systems generate a vast amount of data. Collecting this data is the first step towards understanding how your data could be used to drive performance improvements, achieve building outcomes and effect change.

Data takes a number of different forms, and there are quite possibly just as many different methods of collecting it. All networked operational technology creates data; however, that doesn't mean that it is stored, accessible or transmitted via an open protocol.

Some possible data sources include:

- Building management and control systems (BMCS)
- Electrical metering and energy management systems (EMS)
- Lighting systems
- Vertical transportation systems
- Security and access control
- Industrial automation systems/Programmable logic controllers (PLCs)
- Fire systems
- IoT networks and devices

- Third-party data services e.g., weather data, solar irradiance data, energy meter interval data, National Electricity Market (NEM) data
- Data lakes, serving as cloud-based storage of historical operational data from one or more building systems.

In addition to the data source itself, there are other important considerations with respect to data accessibility. These include:

- Integration method – How the data and analytics contractor will access your raw data
  - Edge integration – A device is physically installed onsite to poll the network for data
  - Cloud integration – An application programming interface (API) is used to access data from an existing database
  - Push integration – extract transform load (ETL) methods are used to perform software integrations where a system has the capability to automatically offload data via email or secure file transfer protocol (SFTP).
- Protocol – Some network protocols are more accessible than others. There are many operational networks that do not use open protocols, and this can make some systems uneconomical to include within the scope of digital transformation projects.
- Database engineering – Not all data used within operational systems is exposed for use by analytics. For some integration solutions, incumbents will need to perform preliminary engineering work to facilitate third-party data and analytics services to access the data, and this often represents a significant, unexpected cost of a data transformation.
- Security and internet access – Discussed in other sections in greater detail; however, any cloud-based analytics provider needs to access data from outside of the site network.

2.1.1 – Big data or smart (targeted) data?

There is a lot of talk about "big data" in the industry but what does it really mean? Is it beneficial to collect every data point possible or only selected points? There can be hundreds of thousands of points in a building, and it is often a waste of resources to collect everything, not to mention the scope of data collection can have significant impacts on the deployment costs.

"Smart data" can still be "big data" but rather than collecting everything, only target points are collected and stored for use. Building attributes such as size, types of activities, location etc., matter because they have a direct impact on a building's expected performance and costs. However, the volume of ingested data always needs to meet the direct benefits from resulting savings (i.e., direct utilisation savings or indirect savings from proactive maintenance practices). Therefore, data collection volumes must be justified by evidence-based outcomes with proven return on investments (ROIs).

Requirements must be considered case by case, because at this stage of the analytics evolution there is no one-size-fits all scenario. Solutions and vendors that promise cost savings are to be expected.

Owners must examine the promised returns against their challenges. For instance, extensive monitoring with big data sensors that perform predictive analytics (e.g., vibration sensors) come at substantial cost that for a lot of cases can put net-present value in the red. In other words, the value of the asset or the risk of it not performing needs to marginalise the investment that will improve its monitoring.

With more and more data being generated, it's important not to get carried away with the collection of it. The collection of all inputs and outputs of a control system is extremely important and should always be performed.

Some examples:
- Energy consumption
- Water consumption
- Sensors
- Actuator positions
- Motor speeds
- Motor enable/status.

It's also important to collect "control" points such as
- Set points
- Schedules
- After-hours calls
- Occupancy rates.

Considering the transformational changes in buildings, time series data and trend logs have now become industry standard – arguably, big data will also find its true path with similarly proven ROIs.

The industry is now expanding to smarter platforms that support an integrated interpretation of this information, not only combined with operational data, but also high-level data and external, environmental data. These data sources can range from enterprise resource planning (ERP) systems of vendors providing technical asset maintenance, to IoT sensor information that complement existing devices and controllers.

2.2 – Data management

Data analysis in smart buildings and asset management hinges upon the critical process of data cleansing and structuring. Data cleansing is often regarded as a preliminary step, and entails normalising and transforming data to eliminate irrelevant information. However, numerical values from sensors lack significance without semantic attachment. This semantic structuring is achieved through formal ontologies, which describe building infrastructure comprehensively, encompassing both physical and digital entities. The fragmented nature of data ownership and creation within the built environment poses a formidable challenge to data acquisition. Subsequently, a considerable portion of time is typically devoted to cleaning acquired data and ensuring its compatibility for processing.

2.2.1 – Data cleansing

Data cleansing is often viewed as a pre-processing step for the data analysis. Data cleaning refers to the process of normalising and/or performing transformations on the data to remove or replace irrelevant information.

Numerical values returned by sensors are useless without a meaning. It is necessary to attach semantic information to the data extracted from sensors, smart devices, and the like. The formal ontology used at the core of a data source helps to describe building infrastructure, including physical and digital entities.

Due to the fragmented nature of data ownership and creation in the built environment, acquiring data continues to be a significant challenge. Once data is acquired, a significant proportion of time is typically spent on cleaning the data to be in a suitable format for processing.

2.2.2 – Data naming and structuring

Each stream of data will usually need to be made human-readable and machine-readable. Data is made human-readable through the use of naming conventions. This should be done in a structured way so users will be able to understand what the data is and how it relates to the broader system. A naming convention will usually be inherited from an organisation's specification or the systems that are being ingested (e.g., from a BMCS). These requirements can often be in conflict, and at scale are practically guaranteed to contain inconsistencies, so applying a naming convention for human-readability does not mean that data will be machine readable.

Machine readability is achieved through semantic tagging or a schema/ontology. The tagging and schema characterises the type of data and its relationship to other data streams independently of a naming convention. This helps to provide a digital model of relationships between data streams without a reliance on every point of a certain class having an identical name. This is also referred to as a metadata model.

Metadata models help to ensure that data is consistently structured and ready to be processed by queries and applications. Structured data makes it easier to use across applications of all types through the standardisation of semantic data models and web services; it can be processed at the Edge Layer or in the Cloud. Tagging enables many operational value points and contributes to a number of business outcomes.

Data tagging covers several applications, including sensor networks, IoT, smart homes, commercial building automation and monitoring, grid-interactive efficient building applications, occupants and behaviour, and asset management and audits.

The majority of schemas have been developed in the past 10 years, and they are in different stages of development. The below list shows the schemas found during the review process for both data-driven smart buildings, and asset management and classification applications.

Ontologies targeting data-driven smart building operation and maintenance (and year created):

- Google Digital Building Ontology (DBO) (2020)
- Virtual Buildings Information System (VBIS) (2020)
- ASHRAE Standard 223 (Proposed Standard) (2019)
- Building Topology Ontology (BOT) (2017)
- Real Estate Core (REC) (2017)
- Brick Schema (2016)
- SOSA (Sensor, Observation, Sample, and Actuator)/Semantic Sensor Network (SSN) ontology (2015)
- Project Haystack (2014)
- Smart Appliances REFerence Ontology (SAREF) (2013)
- Industry Foundation Classes (IFC) (1996).

Ontologies serve the purpose of organising knowledge and defining the relationships between concepts in a specific domain. They help in structuring information, facilitating data integration, and enabling more effective search and reasoning.

The benefits of ontologies include:

1. Clarification: Ontologies clarify the meaning of terms and concepts within a domain, reducing ambiguity and enabling more precise communication and knowledge sharing.
2. Data integration: They facilitate data integration by providing a common framework for different systems and sources to understand and exchange information.
3. Semantic interoperability: Ontologies enable semantic interoperability by allowing disparate systems to interpret and process information in a consistent and standardised manner.
4. Knowledge management: They support effective knowledge management by structuring information, facilitating search and retrieval, and enabling more advanced reasoning and analysis.

Several schemas have developed standardised descriptions of metadata typically stored in large commercial BASs. Collecting and identifying these schemas can be a challenging task. Another challenge in collecting information about metadata schemas is that they may change over time. Similar to software packages, these schemas may go through different versions, and the papers that described their original implementation may not be up to date.

Comparing the ontologies quantitatively is often difficult due to the different purposes underlying their design, intended application and organisations behind their development. A recent survey of core metadata schemas for data-driven smart buildings was published as part of the IEA ECB Annex 81 – Data Driven Smart Buildings [1].

2.2.3 – Data platforms

Data platforms exist to support cloud-based data consolidation, data management and machine-readable access to data. These data platforms help to bridge the gap between on-premises control system operations technology (OT) and web-based information technology (IT) systems (so called IT/OT convergence). This enables improved user-interfaces and access to IT-based tools for data management and visualisation.

There is no one-size-fits-all pathway to implementing a data platform. Large organisations may elect to build a data platform from scratch as a ground-up IT project to meet the specific needs of an organisation, whereas smaller organisations will typically use a third-party service.

As a use case, the Data Clearing House (DCH), led by AIRAH and the CSIRO, is an open data platform for streamlining the data-management interface between building owners and software analytics service providers.

It enables building owners to capture and integrate multiple sources of data in a standardised format and to manage who gets access to the data. This enables building owners to gain control over their own data, and to improve data governance across the industry.

It enables analytics providers to receive data through APIs, avoiding much of the data-management difficulties and associated service set-up costs. It aims to overcome interoperability issues and avoid duplication of IT management infrastructure.

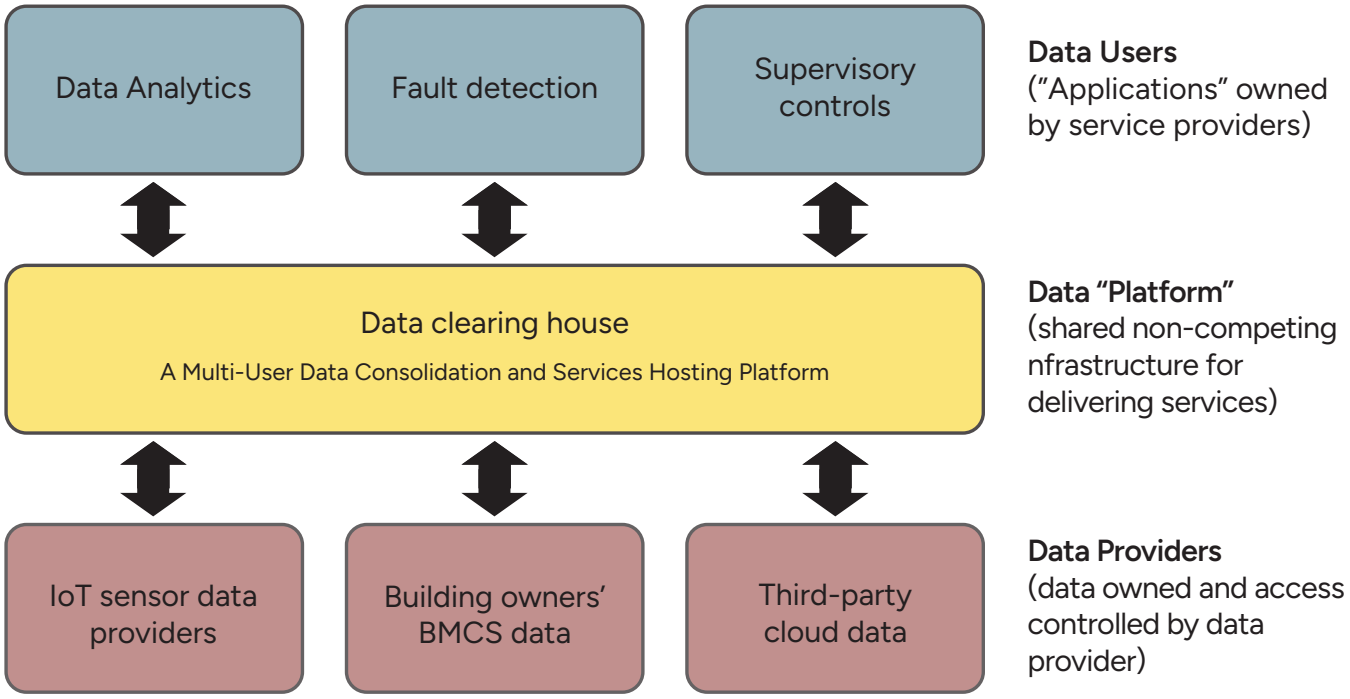

Figure 1: Data Platform element interactions

2.3 – Decision and action

Decision making and actions around the use of data, should be considered before starting on the journey of implementing a "big data" project. Sometimes collecting the wrong data and applying it to your situation can result in the wrong decision being made. The data and the analytics are just tools to allow the user to make decisions. The decisions and subsequent actions need to be targeted to the outcomes that best suit the operation of the target business.

The results of any data analytics tools must lead to some action. This can be automated through the dispatch of supervisory control/automation commands or placement of work orders. Or it could be through information provided to a human responsible for managing assets.

Any action that requires human decision making must be provided through compelling visual interfaces that are easy to access.

Every organisation on a digital transformation journey should evaluate available solutions against the original project objectives. Data and analytics software packages will typically incorporate more features than are required to achieve project objectives, and this makes objective assessment of competing products a non-trivial process. For example, an organisation may need to compare proposals from two vendors against business requirements, simplified in the graph over page.

Table 1: Data Project proposal comparison

Feature	Vendor "A"	Vendor "B"
# points deployed	1,200	800
# algorithms available	50	30
Self-serve access to histories	Yes	Yes
Self-configured energy alerts	No	Yes
NABERS tracking function	Yes	No
Ability to customise algorithms	No	Yes
HVAC specialisation	Yes	Yes
EMS specialisation	Yes	Yes
VT specialisation	Yes	No
Experience in office buildings	Yes	Yes
Experience in industrial buildings	No	Yes
Amortised annual cost	Lower	Higher

Determining the best value for your particular project goals can be complex, and the best value will not necessarily be the lowest cost. The operational technology (OT) data and analytics market is competitive, so assessing business requirements against the market is recommended. In the event specific business needs are not captured by off-the shelf offerings, it is also possible to create a specification and seek out vendors who are willing to assist with bespoke solutions.
In any case, before proposals are accepted, contracts that bind the vendor to deliver the specific features against which a proposal has been evaluated should be drafted. Specialist consultants are able to assist with these scopes in the event they need to be outsourced.

[1] IEA. "Survey of metadata schemas for data driven smart buildings (Annex 81) Energy in Buildings and Communities Technology Collaboration Programme", June 2022. annex81.iea-ebc.org. https://annex81.iea-ebc.org/Data/publications/IEA%20Annex%2081%20Survey%20of%20Metadata%20Schemas.pdf (accessed Apr. 14, 2023)
[2] Victorian Digital Asset Strategy. vic.gov.au. https://www.vic.gov.au/victorian-digital-asset-strategy
[3] Queensland Government Data and Information Guideline - BIM Projects. forgov.qld.gov.au. https://www.forgov.qld.gov.au/information-and-communication-technology/qgea-policies-standards-and-guidelines/bim-projects-data-and-information-guideline

3. How to protect your data

The majority of buildings have poorly maintained and unsecure operational technology (OT) networks. Simply put, there is minimal budget allocation for building infrastructure networks, hence their capabilities are often limited to simple network operations. Because most are backed by home-grade network modems and routers, only the higher-end property portfolios tend to apply best IT practices such as implementing network segmentation, blocking inbound access with firewalls, etc.

This situation creates complexities for HVAC systems because data governance and security are often constrained by the above practices. Security – and specifically cybersecurity – plays a bigger role in this exercise because buildings now require remote vendor access to reduce costs.

With each individual network having a unique combination of technologies, there is no "one-size-fits-all" approach to this task. Every party involved in the installation of a device on a network has their own standards and methodologies for naming and configuration – in addition to manufacturers using differing standards to make the data available for collection.

Additionally, networks tend to suffer from increased payload traffic, since HVAC network systems require specific design considerations to avoid downtimes and failures. One prime example is BACnet broadcast storming, which can occur when specific messages are broadcasted globally, or for a specific network segment. Most vendors consider payload implications for a simplistic BMCS. As such, building analytics tend to be unfit for such architecture.

So, the question remains: What is the best way to secure data and ensure best practice?

The above examples demonstrate that the problem is not specific to one function (i.e., HVAC), or pertain to only one facet. In fact, all OT networks are suffering from similar challenges, with each vendor troubleshooting their own functional challenges and sometimes, creating headaches for another.

Due to these issues building owners should also be looking to collaborate with OT-managed service providers (MSPs) that can apply best practices and guidelines across the building and for all vendors to abide upon. Because remote access extends the building network to the user's machines (e.g., laptops), vigilance is required to ensure no end is left open.

Hence, property owners and managers should apply best-practice security either via in-house teams or via third parties in implementing the most cost-effective and safest cyber security approach possible. Pro-activeness in this regard is critically important to prevent "decision drift" and ease of convenience to rule in regards to cyber securing their big data.

3.1 – Onsite network security

So far, we have considered the concept of data holistically from a systems perspective. However, it is worth considering onsite security and specifically its underlying assets so we can further understand and address risks under various scenarios.

The network is the conduit that allows information to flow between the systems servers, controllers, devices or peripherals, the enterprise system, and the outside world. Intruders able to tap into the network can disrupt the flow of information. The example architecture diagram over page illustrates these practices.

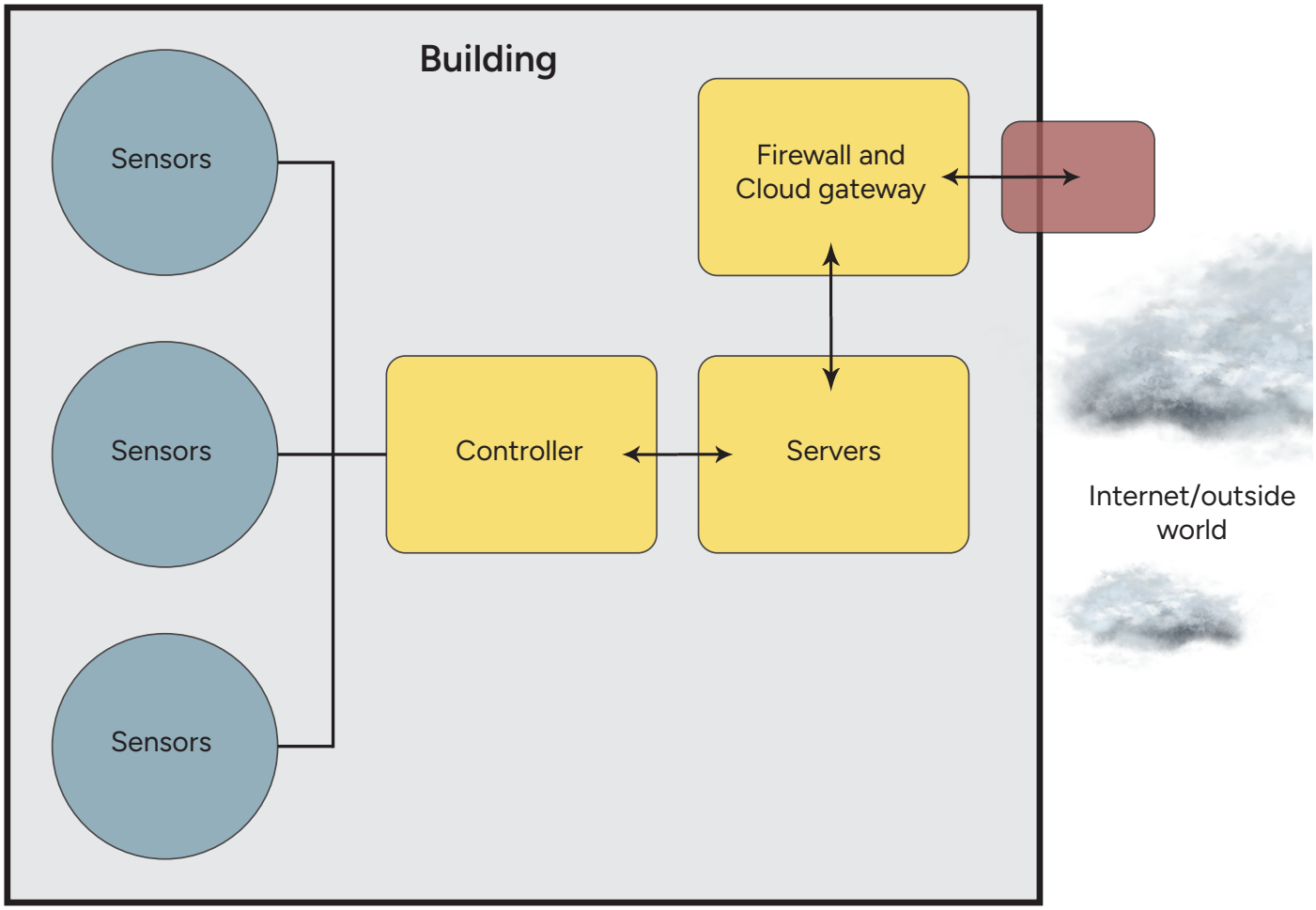

Figure 2: Simplified building OT/IT network architecture

As the amount of data available to third parties increases, so does the risk of misuse or abuse. Often paired with the ability to read data, comes the functionality to write. This means devices can be controlled and programmed remotely over a network. Without proper security measures, these datapoints and controls could be inadvertently exposed to unknown third parties.

Many devices on OT networks use insecure protocols such as Modbus, BACnet, C-Bus, LonWorks, OPC, SNMP, and KNX. It's recommended to use their secure counterparts, such as BACnet Secure Connect (BACnet/SC), Secure SNMP (SNMPsec), and KNX Secure. If an insecure protocol is necessary, countermeasures such as physical security, firewalls, VPN tunnels, and intrusion-detection mechanisms should be implemented to protect the device and communication. Due to the nature of the equipment being installed within OT environments, most security controls must be implemented at the network layer.

Embedded devices such as field controllers or smart IP sensors often won't facilitate end-point protection or other measures employed commonly in the IT space. In other words, they are not smart enough to restrict access on their own and therefore restrict functionality to selective users.

Controlling access in and out of OT environments should be the first step to securing an OT network. This will lead to an understanding of exposure and allow the network owner to adopt a zero-trust-access policy, whereby all access must be explicitly granted. This also ensures OT networks are not being misused by users with access to it. Receiving emails, browsing social media, etc., should be explicitly disallowed from an OT network as these are the most common attack vectors. After all, there is no need for non-business-as-usual activities to take place within an OT network. Secondly, network segmentation should take place to separate physical assets logically.

Virtual local area networks (VLANs) are often utilised to logically segment a network, not necessarily based on physical location or network switch connection. They are especially useful when a single network serves multiple purposes, such as supporting both a BEMS and EMS monitoring within an OT network.

VLANs help address data contamination issues by restricting the scope of broadcast messaging in specific protocols and managing bandwidth. This not only reduces the potential for an attack but allows flexibility in design when separation of IT and OT networks aren't feasible. However, it's important to note that VLANs should not be solely relied upon for security purposes. Regardless of the network architecture used, the goals remain consistent. All connections between OT and IT networks should align with the risk tolerance determined as acceptable by the customer. Any such design should involve conducting a zone and conduit analysis, as mandated by the IEC 62443 family of standards.

These two steps should be considered as the bare minimum when implementing security controls. If feasible, an OT-focused intrusion detection/prevention system (IDS/IPS) should be employed. These appliances monitor the network and autonomously ensure the OT environment remains in check.

By fingerprinting the environment (i.e., associating activities by user or device), these systems are able to understand what traffic should be traversing the OT network and block any anomalous behaviour patterns until reviewed by a network security engineer.

3.2 – Offsite network security

Firstly, offsite network security refers to the notion of remote access. As previously stated, securing the data source is incredibly important. What happens once it is removed from the environment? If we simply shift the data from one secured location to another that has not been secured have we achieved anything? Considerations for transporting data out of the secure environment should ensure traffic is encrypted both in-flight and at-rest. Further, access-control policies should be applied to ensure only trusted sources can both read/write that data.

When transporting data from one site to another, a secure virtual private network (VPN) tunnel is recommended. This tunnel should be configured to use modern standards of encryption and not be overly permissive in the access it allows between the source and destination networks. This type of encrypted communication ensures no prying eyes can view the data as it is transported, in the unlikely event they are able to intercept the communications. Regardless of the feasibility, the data should never be transmitted "in the clear" – without some form of strong encryption.

Once the data has been successfully extracted from the OT environment, it should then be encrypted for storage. This is often referred to as "at rest". This mitigates the risk of data leakage exposing the content. Again, encryption should take place, combined with access-controls restricting access to the data only to trusted entities. These entities should be audited on a regular basis to ensure no overly permissive or outdated access is maintained.

3.3 – What is needed in implementing a cyber-secure posture?

From the above, it is evident that data connectivity is attached to governance and security. Therefore, increasing a building's "smart" capabilities carries risks that require additional investment to be mitigated and controlled. The question therefore remains: "What do you need?".

The truth is, there is no one-size-fits-all solution. However, cyber-security practices are not new – they have existed in the IT space for many decades. OT should not be reinventing the wheel but rather adapting these practices as needed for the idiosyncrasies associated with OT and OT networks.

To understand the risks and costs, it is best to first consider the long-term and short-term strategy for a building – starting with the capabilities that the building requires now and in the future. Is it tenant comfort, energy efficiency and green ratings, or simply a reduction in opex and capex?

A deep and thorough roadmap of the building's strategy, along with a cost/benefit analysis, is required to evaluate the value of big data from an infrastructure perspective. Whilst considering the costs of the technological capabilities that are of interest, the costs to cover the gap of the building's current OT network infrastructure needs to be included to arrive at a true net-present value analysis of the investment.

4. Using your data

Once you are able to collect your data, store it and have placed a level of protection and security around it, it is time to use your data.

Using your data is often subjective and is generally dependent on the project's desired business outcomes. In the case of HVAC, there are three main drivers: to improve comfort, reduce costs and emissions, and minimise risk. To achieve these outcomes, we may employ some or all of the techniques discussed in this section.

4.1 – Energy modelling and monitoring

Throughout a building's operation, many changes and new additions are made and implemented. These could vary from new equipment, plant control upgrades, to new tenants and fitouts. Each change will produce resultant changes in the building's energy usage. A predictive energy model captures what could happen to future energy usage based a set of key parameters such as plant efficiency, plant set points, weather forecast and building conditions required.

Being able to predict or gauge what tomorrow or next week's energy consumption profile might look like enables further action such as peak demand optimisation and, to a certain extent, identification of when grid load-shedding windows could occur.

Peak demand optimisation and demand response management allow the user to either clip, fill or shift their energy consumption to a more favourable time where the cost of energy is cheaper, more available or self-generated. A monthly electricity bill can consist of between 30–50% demand charges for large users and optimising when and how electricity is consumed, while maintaining building conditions and processes, can yield favourable results.

For example, a building with multiple air handlers may have staggered start times for each unit, but often this spacing is only for a few minutes. Although this reduces the instantaneous in-rush current for the building, the utility provider generally measures demand in 30-minute windows. So even though the start times of the air handling units are staggered, what is reflected on the invoice will still be a high demand charge. The optimisation problem then becomes, how much consumption outside of the maximum demand window can we afford, before our actual consumption costs, outweigh the max demand costs imposed by the utility?

By collecting and understanding data from the building's historic performance, its energy consumption, demand profile and weather data, it is possible to answer this question and determine what plant should start when and how hard the plant start up for each day of the week and continually adjust everyday based on previous operation to yield the best results in terms of cost, risk and comfort.

4.2 – Automated fault detection and diagnosis

With the masses of data being generated by systems today it can be beneficial to use software for the analysis and automated detection of operational faults. These software platforms run data through a set of algorithms to automatically detect and diagnose faults within the HVAC&R systems.

The word "analytics" is used to convey various types of capabilities, so it is important to clarify what type of analytics is being discussed. Below is a summary of the types of analytics and their primary function.

4.2.1 – Threshold-based "fault detection" analytics

Fault detection has been around for many years and has typically been performed by the BMCS in the form of "alarms". The primary function of fault detection rules is to test data against known fault conditions to alert the end user. These rules are very basic and typically pick up items such as failed temperature sensors, equipment motor mismatches, no airflow readings, etc.

These rules give very little insight to the cause of these faults and only alert the user to a problem for further investigation. However, they can be important to highlight critical faults with critical pieces of equipment – e.g., if a chiller goes into "fault" mode or chilled water temperature is dangerously low, threatening freezing issues.

4.2.2 – Rules-based fault detection/diagnostics

Fault diagnosis algorithms expand further on fault detection rules and help pin-point the cause of an issue. These algorithms detect faults based on complex logic or rules that take into account multiple data points. These rules can be deployed efficiently across multiple facilities and can detect more complex faults and inefficiencies / deficiencies in operation. Advanced versions of rules-based fault detection first detect a fault, then interrogate data further to diagnose the issue. Fault diagnoses algorithms can help to reduce the amount of time spent investigating the cause of issues and streamlines the maintenance process.

4.2.3 – Machine-learning-based fault detection/diagnostics

As buildings have become more data-intensive, analytics based on statistics and machine learning approaches have been developed to identify faults and optimisation opportunities. Machine learning approaches do not rely on prior knowledge of building physics or control logic to detect faults – instead, they utilise data science to explore relevant insights hidden behind the data set.

The benefit of this is that the data does not necessarily need to have the high levels of consistency and structure that is required for rules-based algorithms. Commonly applied machine learning approaches include statistical regression models, artificial neural networks, and pattern recognition.

4.2.4 – Automated system optimisation

A separate genre of capability that often overlaps with fault detection/diagnosis-based analytics are "automated system optimisation" tools. These tools aim to actively analyse and modify BMCS settings to optimise HVAC energy consumption while maintaining occupant comfort, all without human intervention [4].

[1] Characterization and Survey of Automated Fault Detection and Diagnostics Tools - https://eta.lbl.gov/publications/characterization-survey-automated
[2] Unsupervised learning for fault detection and diagnosis of air handling units - https://www.sciencedirect.com/science/article/abs/pii/S0378778819320134
[3] A Review of Fault Detection and Diagnostics Methods for Building Systems - https://www.tandfonline.com/doi/abs/10.1080/23744731.2017.1318008
[4] Proving the Business Case for Building Analytics - https://buildings.lbl.gov/publications/proving-business-case-building

4.3 – Data-driven maintenance

Traditional maintenance typically involved arriving to site, checking the system operation by looking at the graphical headend and maybe manually manipulating the system to test control strategies. All too often this is only done on a single day a month (or a quarter) and only problems that are apparent on the day are rectified.

The introduction of automated fault detection and diagnostic systems means the mundane task of staring at a screen all day wasting time looking for problems is a thing of the past. Automated fault detection systems will functionally test equipment throughout its normal operation and identify issues as they occur. This allows the maintenance team to fix problems instead of looking for them.

Preventative maintenance models are the industry norm and still dominate the market. These models see maintenance staff perform time-intensive tasks that don't actually improve the buildings comfort or performance. These tasks typically involve the interrogation of graphics and trend logs which becomes repetitive and often issues are missed, not only this but typically only a couple of days' worth of data is checked at most.

Let's take a simple VAV box maintenance procedure for which maintenance staff would perform the following:

1. Look at BMCS graphic to check the following:
    a. Airflow reading – Is it reading zero? How close is it to set point?
    b. Zone temperature – Is it reading open or closed circuit? How close is it to set point?
    c. Damper position – Is it modulating to maintain airflow set point?
2. Change room temperature set point to check if airflow set point responds.
3. Drive damper actuator open to check if airflow increases
4. Drive damper actuator closed to check if airflow decreases and reads zero (or close to)

The above tasks are typically the minimum performed and only catch operation at that point in time, the following tasks are also often performed.

1. Room temperature trend logs interrogated to check how close to set point they operate.
2. Airflow trend logs interrogated to check how close to set point they operate.

Once any issues have been identified, it is only then that the VAV is physically checked and any issues rectified. This manual interrogation if system operation is very time-consuming, the above checks would take a technician 10–15 minutes per VAV to perform, this starts to add up very quickly when there are hundreds of units to check.

Because of this time-consuming process, maintenance procedures have adapted to typically only check any equipment operation once a year, with only certain pieces of equipment checked every month. This means that problems could potentially go unchecked until month 12 of the contract.

Data-driven maintenance is relatively new to the industry and has become more prevalent with the introduction of analytical systems. By using analytics to functionally check equipment and raise issues as they occur, maintenance time onsite can be focused to fixing the worst performing items of equipment.

This removes the need for maintenance time to be used checking the BMCS graphics for control issues, and allows more time to rectify problems, thus reducing tenant comfort issues and costly faults.

If we take the previous VAV example and apply a data-driven maintenance model to it, the maintenance staff would perform the following:

1. Log in to the analytics system to check for any VAV faults.
2. Check VAV KPI reports to identify the worst controlled faults, based on room temperature and airflow.

The above tasks would take roughly 5–10 minutes to perform for ALL VAVs onsite, compared to the 10–15 minutes per VAV using the traditional maintenance model – this considerably reduces time to check systems for faults. It also means that issues are identified and fixed as they occur, not as the system is checked (potentially 12 months).

A data-driven model now sees an allocation of time onsite to check poor-performing equipment and to perform some physical checks. For example, a time allocation to check the "X" worst performing VAVs and AHUs, calibrate all sensors, and physically check damper actuators. Onsite checks still need to be performed to visually verify everything is OK.

5. Asking for a quote

Embarking on the journey of implementing a data and analytics platform for your business is a strategic decision that demands clarity and foresight. Before diving into the intricacies of selecting a data and analytics contractor, it's imperative to lay a solid foundation by defining your goals, needs, stakeholders, and usage patterns. This initial groundwork not only streamlines the procurement process but also ensures that the solution aligns seamlessly with your organisational objectives.

By addressing these fundamental questions and factors, you will be better equipped to navigate the complexities of selecting a data and analytics contractor and laying the groundwork for a robust and tailored solution.

5.1 – What are your goals?

Before requesting a quote from a data and analytics contractor, it's important to clearly understand your business goals. What are you seeking to achieve by implementing a data and analytics platform, and what are your success criteria? Having a clear understanding of this will not only allow you to create and communicate clear business cases, but it will also allow data and analytics providers to provide a quote tailored to your specific use case.

Note that the initial deployment costs of this kind of project often represent a significant portion of the lifecycle costs.

5.2 – What do you need?

Prior to issuing a quote, a data and analytics contractor will need some technical information on the systems, level of service and assets being onboarded. Before starting conversations you should collate the following information:

- Breadth of scope – e.g., which systems are to be included in the scope?
- Magnitude of application – How many buildings form this scope of works? How large are the buildings?
- System OEM and software version – Be ready with the brand and version of the system(s) you would like integrated. This can impact initial cost.
- Complexity – Points list and/or asset registers allow vendors to provide accurate pricing.

5.3 – What stakeholders are involved?

Stakeholder involvement varies based on the scope of the project. The solution provider will be able to advise who will need to be involved for a streamlined experience, but the core stakeholders are as follows:

- End users – The most important stakeholders, end users may vary depending on the organisation. It's Important to engage with them during the procurement process and final sign-off of the selected solution. Other considerations for end user requirements are discussed in 6.1.6.
- Corporate IT – Any use of cloud software or connection of OT systems to the internet will likely require involvement of IT to ensure the solution meets the security requirements of the organisation. Passing security tests may take months, so starting this journey early is advisable.

- OT system vendor – This is the vendor responsible for the equipment that produces the data points to be gathered, and there may be multiple vendors depending on the breadth of data collection. Typical vendors are BMCS, mechanical, lighting, and fire contractors; however, this list is rapidly expanding with the installation of newer technologies.

For projects with large, ambitious scopes, the following stakeholders may also be involved:

- Network manager – The vendor responsible for managing the network infrastructure onsite. Usually only involved if a site has a converged network.
- MSI – Typically involved in new works or large-scale network upgrades. This is a specialised vendor for ensuring network communications and data flows are built and commissioned as specified. It a role akin to project management.
- Consultant – Specialised consultants may be engaged by an organisation to provide advice or create packages for tendering.
- Procurement – Required to be involved if the scope of a contract exceeds a certain cost.

5.4 – How is it being used?

Data and analytics service providers specialise in different areas. Having a clear picture of how you intend to use a platform, who will be using it, and how many users you anticipate will help you find the right provider for your needs. For example, the platform functionality required for executives to view dashboards of performance against environmental rating systems is very different to what a maintenance technician would need to perform data driven tasking. Similarly, platforms are designed to work well for certain scales of client, and the platform which works well and is cost effective for a single site with a small number of users, may not be the best application for an enterprise level application.

Consider the content and the user experience when making your decision. Try to match your own ambitions with the strengths of the platform, and make sure the features you're really interested in are user friendly.

5.5 – Who is using it?

The intended users, and the frequency of use should influence the choice of provider. As a rule of thumb, "decision maker" users, e.g., executives, are most engaged with simple KPIs, which are presented simply and with little optionality. Decision maker users will rarely want to self-service large quantities of raw data.

"Operator" users, e.g., technicians, are typically technical; however, they will usually want to only use the platform for very specific tasks, and prefer not to learn a complex platform. More technical users, e.g., consultants and data contractors, will often want to be able to heavily customise their experience and will frequently seek to access raw data.

Consider asking for a demonstration from shortlisted providers. Providing all of your proposed users with an opportunity to do test runs and provide feedback will assist you in choosing a platform which is easiest to use for your staff and contractors. Ease of use will drive engagement, which will maximise the value you can derive.

5.6 – Independent or existing services provider?

Many BMCS and mechanical contractors offer data and analytics solutions, so it is easy to simply discuss your digital transformation goals with your incumbent contractors. However, you may prefer an independent contractor, particularly if one of your motivating factors is transparency, or a fresh set of eyes. There are pros and cons for both approaches. The use of independent providers can create an adversarial environment if not well managed, and it's worth noting that any independent provider will rarely have knowledge of site history that an incumbent has. A counterpoint is that incumbents may have developed inertia with habits and methods which deviate from industry best practices and introducing an independent data contractor could allow for some new ways of working to be identified.

5.7 – Big data services checklist

The checklist here is presented as a series of questions, the answers to which ensure every aspect of your big data journey receives due consideration. Covering essential elements across the gamut of issue, from network security to asking for a quote, it provides a holistic assessment to determine readiness for the challenges and opportunities ahead.

Network security	
Is site connected to internet?	
Is the BMCS network residing on the sites IT network or physically segregated?	
Does your organisation have IT security policy and procedures in place?	
Do your prospective data and analytics vendors meet your policy requirements?	

Data capture	
What are your project objectives?	
Do you have a list of data points you expect the project to capture?	
Do you have an established naming convention and/or ontology?	
What frequency of data capture do you expect (e.g., 5-minute, 15-minute intervals? Change of value?)	
Will there be transparency around data integrity? Are there service level agreements (SLAs) in place to provide you contractual assurance of uptime?	
What processes are available when data loss occurs? What costs will be associated with restoring data loss caused by a third party?	
What is the process to expand data capture if the project evolves?	

Data management	
How will your data be stored? Is this consistent with your data policy?	
How will the stored data be structured? (e.g., data schema)	
How do you want to be able to access your data? (e.g., dashboards, downloadable histories, API)	

Data management (continued)	
How long will histories be retained?	
Will you be able to access your historical data on termination of an agreement, in what format and is there a cost associated with this?	
How will your data be protected if it leaves your premises and do you own the data?	
Do you have storage redundancy?	
What would be the process and impact of transitioning data storage (i.e., Azure cloud to AWS cloud)?	

How will it be used	
Will you use the interface provided with the platform?	
Will you incorporate the data into your existing platforms?	

Who will use it	
What roles need to consume the data in your organisation?	
Do you need different access levels for different roles?	
Will external third parties need to access the data and platform?	
Are there any costs associated with adding and creating new users? Can you easily delete unused logins (e.g., with churn of staff)?	

Asking for a quote Minimum recommended information to provide vendors in request for quote:	
Size and scale of the building (by square meters)	
BMCS points list and HVAC asset list	
BMCS type and internet connectivity	
Utility demand history either from an EMS or utility provider	
What systems intended to be connected for analysis	


AIRAH
HVAC&R FOR A BETTER WORLD