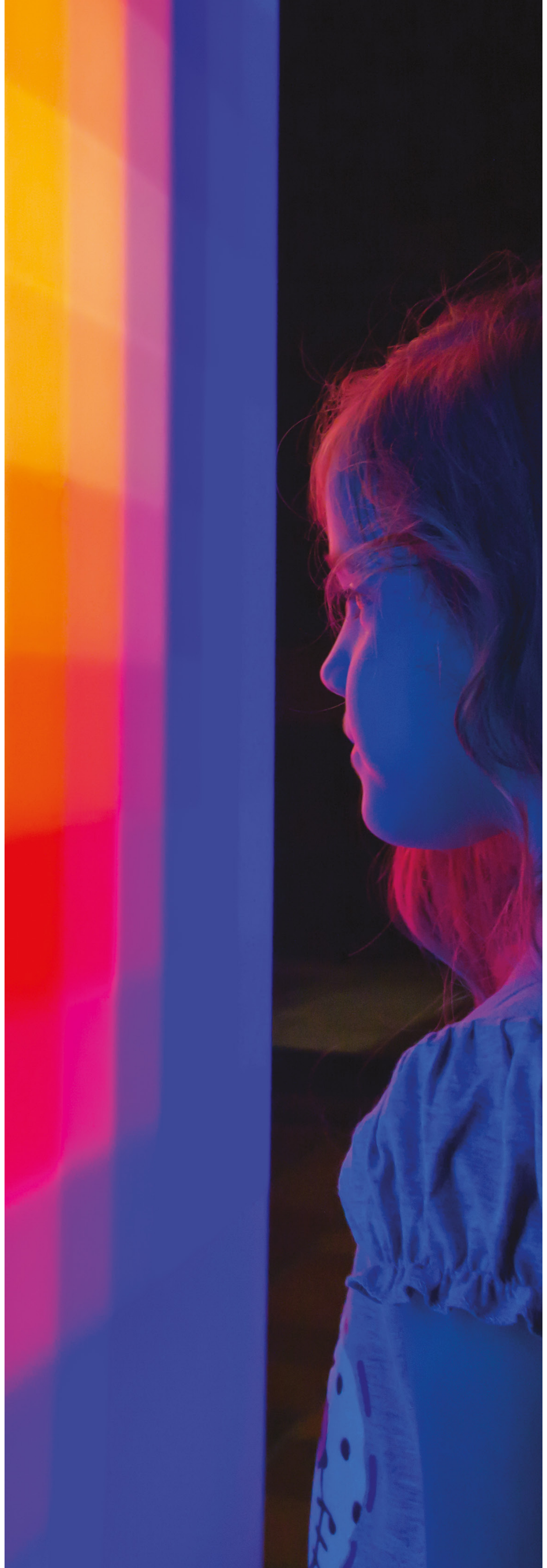
interact

A white paper

Data management, privacy, and security in connected systems

www.interact-lighting.com

Contents



01. Security, privacy, and connected systems

1.1 The Internet of Things and connected devices

The Internet of Things (IoT) has been called the next industrial revolution. It is already transforming the way businesses, governments, and consumers interact with the physical world. By blending the physical and digital realms, the IoT is profoundly changing the way we relate to our environment, to each other, and to information. It is revolutionizing the way we live, work, travel, heal, and relax.

According to Gartner, the IoT is "a network of physical objects that contain embedded technology to communicate and sense or interact with their internal states or the external environment."[1] The sensing is done by sensors of various types—whether these are motion sensors in the ceiling of an office space, noise-level sensors on a city street, or sensors that can detect physiological states and changes in an individual. The communication is handled by standard wired or wireless communications methods embedded in connected physical devices.

Connected physical devices range from thermostats to energy meters to tractors to wearables for monitoring personal fitness and vital signs. Any digital device that can collect or share meaningful data about itself, its usage, and its environment is a candidate for participation in the IoT.

Connected devices generate data of various kinds. In IoT applications, this data is often aggregated in the cloud. This aggregated data can be processed and analyzed to extract knowledge and actionable insights that businesses, municipalities, and individuals can use to achieve their goals.

The IoT is still a fairly young technology, and its infrastructure is still developing, but its potential impact is enormous. The McKinsey Global Institute (MGI) predicts that "the Internet of Things has a total potential economic impact of $3.9 trillion to $11.1 trillion per year in 2025...equivalent to about 11% of the world economy."[2] Business Insider predicts that 24 billion devices will be connected to the Internet by 2020—in other words, "four devices on average for every human on Earth."[3]

The sheer volume of data now being collected from these billions of connected devices requires a special combination of technologies, analytical approaches, software platforms, and computing power. This combination is known as Big Data. Big Data poses novel data management challenges that can't be resolved with traditional approaches. It therefore poses novel risks, especially while enterprises are still learning how to avoid pitfalls and adopt emerging best practices.

1.2 System security and data management

Some security experts have said that the only way to ensure that data is safe is not to save it in the first place. Others, like CTO of IBM Resilient Bruce Schneier, consider data a “toxic asset” that must be treated “as we would any other source of toxicity.”[4] While these might be extreme views, leading analysts and researchers agree that security is among the top challenges facing the IoT today. In a study that McKinsey conducted in 2015 in collaboration with the Global Semiconductor Alliance (GSA), respondents most frequently cited security as their greatest concern about the IoT.[5] Similarly, in Gartner’s 2016 *Internet of Things Backbone Survey*, which studied important geographic regions around the world, including China, Germany, India, Japan, the United Kingdom and the United States, “security emerged as the top concern” from a technology and administration perspective.[6]

With so much connectivity and data flowing through so many interconnected systems, how do you keep everything secure? How do you protect the enormous and growing range of invaluable digital assets—including company-confidential data, city infrastructure, information on private individuals, transactional data, and home devices?

The two main aspects of security are system security and data management. System security includes both physical security and cybersecurity—the physical network, connected devices, applications, communications operations, and so on. Data management comprises all the disciplines needed to manage data as a valuable resource. The Data Management Association (DAMA) Data Management Body of Knowledge defines data management as “the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets.”[7] Data governance, data security, and data privacy are all aspects of data management.

1.3 Connected lighting in the IoT

The LED lighting revolution began in the late 1990s, when companies such as Philips Color Kinetics designed and brought to market best-of-breed digital lighting systems and luminaires for professional and home use across the entire range of lighting applications. White-light LED luminaires with light output and quality equivalent to or better than comparable conventional luminaires began to appear in the mid-2000s, along with color-changing LED lighting for architectural applications, which optimized and improved upon approaches and control solutions adapted from entertainment and stage lighting. Recent innovations in the digital lighting space include spectrally tunable LED luminaires that support new human-centric lighting approaches in wellness and productivity applications.

Within the last five years, leading-edge companies have been applying their expertise in LED lighting to developing connected lighting systems, both in the consumer realm and in the professional lighting space. Connected lighting is the intersection of digital lighting and the IoT. In a connected lighting system, LED luminaires are enabled with two-way data communications, allowing them to participate in the IoT by sharing data about their own status and operations.

Since lighting is already installed, or has to be installed, everywhere that people work and live, and everywhere that they go in urban environments, it serves as a natural platform for sensor networks and other physically distributed systems. If the lighting system is connected, it can serve as an enabling platform for delivering IoT applications wherever lighting is used. In Cisco’s view, lighting is the first step in creating a single converged IP network that can integrate disparate networks—HVAC, metering, lighting, CCTV, physical security, scheduling—into one network. This converged and connected network provides building intelligence that delivers “new and innovative experiences for building occupants while providing granular energy management, control, analytics, and integration capabilities for building owners and operators.”[8] Similarly, smart cities increasingly rely on converged infrastructures to deliver better experiences and outcomes for citizens. Connected street lighting serves a similar role as the digital ceiling, creating a platform for distributing sensors, broadband communications equipment, and other connected devices throughout a municipality, and offering APIs for integrated monitoring and management of disparate city services networks.

Each of these applications represents new sources of data that must be managed. For example, LED luminaires can collect and share data collected from any sensors that may be integrated into the system. These typically include daylight, occupancy, motion, noise, and air quality sensors, but there’s no inherent limitation on the type of sensors that can be added to the system. Communications from the lighting system, via visible light communications or other means, enables location-based applications that typically make use of ubiquitous smartphone apps and wireless connectivity. These have many advantages for the users and managers of spaces, and represent additional data streams that a cloud-based system can store for processing and analysis.

Interact systems use the connected lighting infrastructure to offer IoT applications in several key application areas. Via partnerships with other leading technology and communications providers, such as Cisco and Ericsson, Interact systems offer end-to-end connected lighting solutions for smart cities, smart buildings, and smart retail. While specifics differ from system to system, all systems share a general architecture that includes:

- Lighting instrumentation, including connected luminaires, sensors, and lighting controls
- Networking hardware and software, including gateways, servers, switches, cabling, and data communications
- Management and monitoring software for the lighting system and IoT applications that the lighting system hosts
- Cloud services for hosting software applications and for gathering data from illuminated environments
- APIs for integrating Interact applications with other facilities and management applications in the digital ecosystem, and for building mobile apps and other software components, such as dashboards
- A data analytics platform for turning the raw data collected from connected systems into actionable knowledge and wisdom

Each of these aspects of an IoT system presents its own challenges to data privacy, data and system security, and data governance.

1.4 Security and privacy challenges in connected systems

Among the most significant threats to security and privacy are malicious access, denial of service and theft of services, especially data. Attackers are increasingly using IoT hacks to stage other types of disruptions. This means infecting connected devices with malware and coordinating them to unleash a torrent of Internet traffic to bring down websites and other online resources in what's known as a distributed denial of service (DDoS) attack. In 2016, hackers used the massive IoT botnet Mirai to exploit IoT security shortcomings. Mirai was used "to attack and temporarily bring down individual websites, but was also turned on Internet Service Providers and internet-backbone companies, causing connectivity interruptions around the world."[9]

As the number of insufficiently secured connected devices increases, security experts expect the number of DDoS and other attacks to increase. And as new kinds of devices are connected, hackers are devising new kinds of attacks. A DDoS attack in Finland in November 2016 shut down the heating systems in several apartment blocks in Lappeenranta in sub-zero weather.[10]

While the heating system shutdowns in Finland appear to be inadvertent, it's not too hard to imagine attacks aimed directly at connected systems in cities, offices, and homes, such as heating, physical security, and lighting. "The hacking of baby monitors, smart fridges, thermostats, drug infusion pumps, cameras and even the radio in your car are signifying a security nightmare being caused by the future of IoT," writes IoT expert Ahmed Banafa. "Concerns will no longer be limited to the protection of sensitive information and assets. Our very lives and health can become the target of IoT hack attacks."[11]

Security experts have identified and addressed dozens of risks in the connected lighting ecosystem, taking the attitude that "it pays to be paranoid." Because connected lighting systems are complex systems, there are many ways to attack them. Security experts are developing system security and data management policies and procedures to mitigate the risks. These include attacks against connected devices, gateways, bridges, and other networking hardware, cloud interfaces and infrastructures, internal and external APIs, and mobile apps.

While many security risks apply across all connected lighting applications, there are some specific considerations in the different major application domains. In smart city applications, concerns around the security of municipal systems (traffic, street lighting, emergency response) are front and center, while security for smart building applications focuses more on protecting corporate assets and employee privacy. Smart retail applications have a special mandate to secure customers' financial, transactional, and personal data, while smart home applications must also focus on risks to the homeowners.

02.
All about data

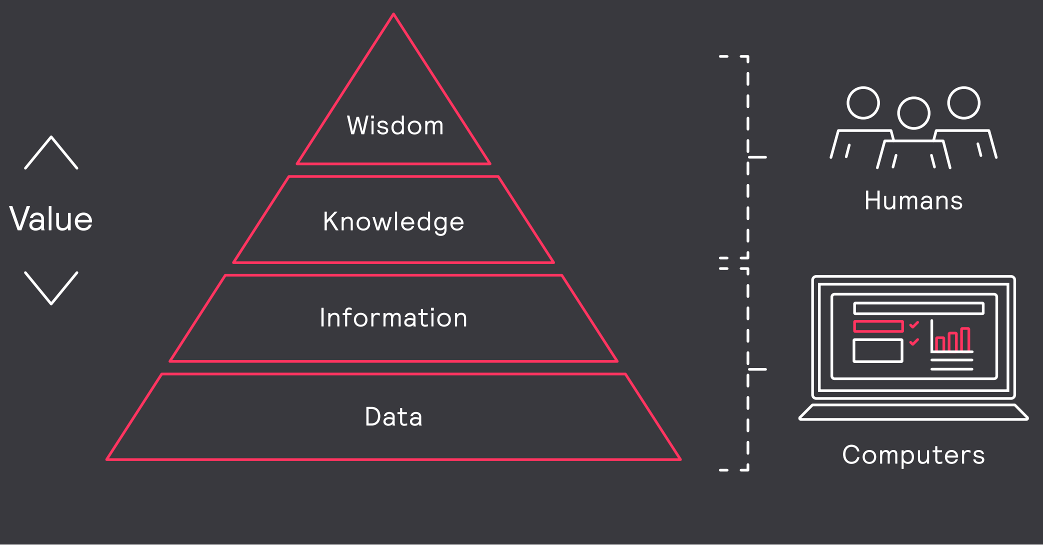
2.1 Data as a valuable smart system asset

Smart connected infrastructures, whether in cities, workplaces, retail environments, or homes, generate large amounts of data. Municipalities and privately-owned businesses can gain competitive advantage by reducing service delivery costs and streamlining operations; anticipating and satisfying the needs of citizens, employees, and customers; and creating new revenue streams through new data-driven products and services.

Organizations can combine data gathered through connected applications and infrastructures with relevant external data and domain models to gain a deeper understanding of people's behavior and interactions. For example, a city can save energy through sensors, controls, and software that gather and analyze data on when and where people are, allowing the city to deliver light when and where it's needed, giving people a better urban experience while minimizing wasted light and energy. Similarly, a business can gather and analyze data on the usage and activities in the workplace, allowing the organization to schedule lighting, HVAC, and other services to reduce costs, improve operations, and enhance the work environment for employees.

So what exactly is data? How is data transformed into actionable insights? What is the ideal approach to data lifecycle management, and how can the correct approach provide critical security and privacy protections for the users and owners of connected systems?

2.2 From data to wisdom



In the most general terms, *data* is a collection of numbers or characters. Data can be measured, collected, analyzed and presented using various formats such as tables, charts, graphs, and images. Conceptually, data refers to the fact that some existing information or knowledge is represented or coded in some form suitable for better usage or processing.[12]

Data, information, knowledge, and wisdom are closely related concepts, but each is distinct. Data alone has little value, as it must be processed and contextualized to yield actionable insights.

A widely used model to represent relationships among data, information, knowledge, and wisdom is known as the DIKW pyramid.

This model represents a hierarchy, and implies a series of transformations for ascending the hierarchy. Data is the basic constitutive element of information. To become information, data must be processed. Processed and interpreted information must be analyzed to yield knowledge. Principles must be applied to knowledge to result in wisdom.

without context

In the context of the DIKW pyramid, data is a set of symbols or signs that represent stimuli or signals. These signals have no meaning or value until they are put in a usable form and context. Examples of data are:

red

1466005743

-33.882816, 151.204150

with description

Information applies description to data to make it useful. If we add some description to the data examples above, we might end up with information like the following:

The traffic light turned to red on 15 Jun 2016 at 15:49:03 GMT on the corner of Pitt Street and George Street.

with knowledge

Knowledge brings additional context and rules to information:

I'm driving towards the traffic light that has just turned red. Rules says that I must stop my car when traffic light is red.

with wisdom

Wisdom builds on knowledge and experience accumulated over time. You could say that wisdom is "knowing the right things to do." Wisdom that builds on the knowledge in our example could be expressed something like this:

"Driving through a red light is illegal, dangerous, and potentially lethal. So I had better stop my car at this red light."

Because wisdom involves using knowledge and experience for the greater good or a high-level goal, it is deeper and more uniquely human than knowledge. It requires a sense of good and bad, right and wrong, ethical and unethical. It involves an understanding of people, objects, events, and situations, and the willingness as well as the ability to apply perception, judgement, and action in keeping with an understanding of the optimal course of action. As such, wisdom is intrinsically subjective.

As the DIKW pyramid shows, data derives meaning and value from processing with a purpose—with an end goal and a business objective in mind. For businesses, this requires an organizational focus on data management processes.

2.3 Data categories: structure, size, speed, and source

It is important to realize that there are different categories of data which have different management requirements. The main categories are *structure, size, speed*, and *source*. Each of these must be accounted for in an effective data management process.

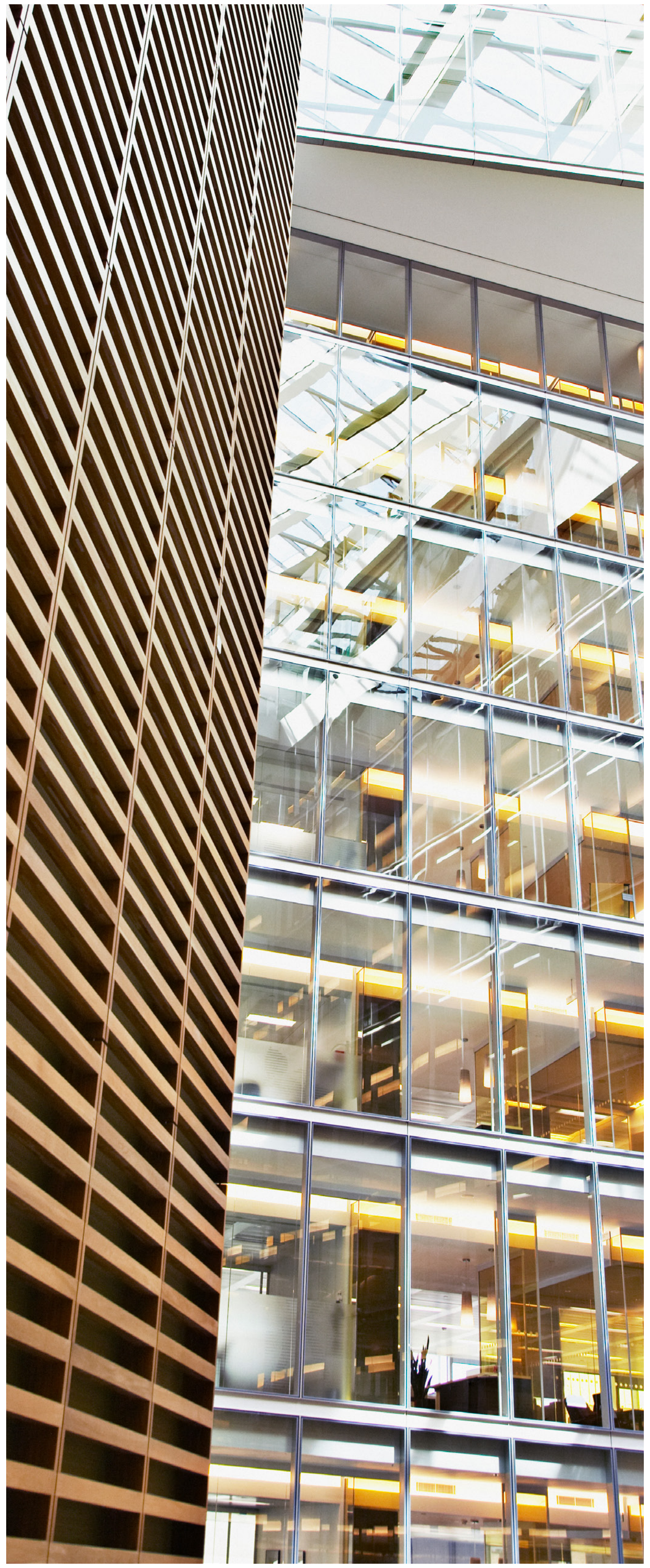
2.3.1. Structure: structured, semi-structured, and unstructured data

Structured data refers to data with a high-level organization. Structured data has a pre-defined data model or a *schema* – a model of the types of data that will be recorded and how they will be stored, processed, and accessed. The schema includes defining what data will be stored and how it will be stored, including data type (numeric, currency, alphabetic, name, date, address) and any data input restrictions (number of characters, specific terms, numeric ranges).

Examples of structured data are customer relationship management, enterprise resource planning, transaction management, supply chain management, employee information, and system logs. One advantage of structured data is that it can be easily queried and analyzed. Historically, because of the high cost and performance limitations of storage, memory and processing, relational databases and spreadsheets were the only way to effectively manage data.

Semi-structured data is a type of structured data that lacks a strict data model structure. With semi-structured data, tags or other types of markers are used to identify certain elements within the data, but the data itself does not have a rigid structure. For example, emails have the sender, recipient, date, time and other fixed fields added to the unstructured data of the email message content and any attachments. Photos or other graphics can be tagged with keywords such as the creator, date, location and keywords, making it possible to organize and locate graphics. Filesystems and various file formats are often used to manage semi-structured data.

Unstructured data refers to data that either does not have a pre-defined data model or is not organized in a pre-defined manner. Unstructured data is typically text-heavy, but may contain other types of data such as dates and numbers. This results in irregularities and ambiguities that make it difficult to understand using traditional programs as compared to structured data. Techniques such as data mining, natural language processing (NLP), and text analytics provide different methods to find patterns in, or otherwise interpret, unstructured data. Examples of unstructured data are social media, web content, and call center logs.



2.3.2. Size: small data and big data

Another way to categorize data is by its size. Data that can fit in the memory of a computer and can be managed by traditional data processing applications is called small, while data that is so large or complex that it cannot be dealt with using traditional data processing applications is called big data.

Traditional relational database management systems and desktop statistics and visualization packages often have difficulty handling big data. Processing big data may require massively parallel software running on tens, hundreds, or even thousands of servers. What counts as “big data” varies depending on the capabilities of the users and their tools and expanding capabilities, making defining big data a moving target.

2.3.3. Speed: data at rest, data in motion, slow data, and fast data

Another way to categorize data is by its dynamic characteristic. Data that is static in nature, i.e. stored in persistent storage (disk, type) in any digital form (e.g. database, data warehouse, spreadsheet, files) is often called data at rest. Data in motion is the term used for data as it is in transit. Data in motion involves processing of data on the fly without storing it.

A typical characteristic of data in motion is velocity. The velocity is the rate of flow at which the data is created, stored, analyzed, and visualized. Fast data velocity means data is being processed in a short amount of time. In the fast big data era, data is created and passed on in real time or near real time. Increasing data flow rates create new challenges to enable real- or near real-time data usage. Traditionally, this concept has been described as streaming data.

The second characteristic for data in motion is variability, which refers to any change in data over time, including the flow rate, the format, or the composition. Given that many data processes generate a surge in the amount of data arriving in a given amount of time, new techniques are needed to efficiently handle this data.

2.3.4. Source: internal data and external data

From the data source or origin perspective, data can be categorized as internal or external. Data that comes from internal systems (for example, company systems, IT applications) is called internal, while data that comes from third party services is called external. Examples of external data are social media, weather or traffic data from third party web services.

Traditional enterprise information management systems deal with structured or at most semi-structured data, small and slow data, data at rest and internal data. IoT data, which are often characterized by unstructured, big data streams and external data, demands new architectures for managing such data.

2.4. Data value principles

There are several key principles of value that drive how data assets should be managed for value.

The value of data increases with use
The more that data is used, the more valuable it is. The costs associated with data are primarily determined by acquisition, storage, and maintenance, which are fixed costs independent of usage. The costs of *using* the data are negligible. Because data is infinitely shareable and non-depletable, its value can be multiplied many times over if used to its fullest potential. Conversely, data that is not used is a liability because acquisition, storage, and maintenance costs are incurred for no reason. To be valuable, data must be easy to discover, easy to use, and shared to the maximum extent possible.

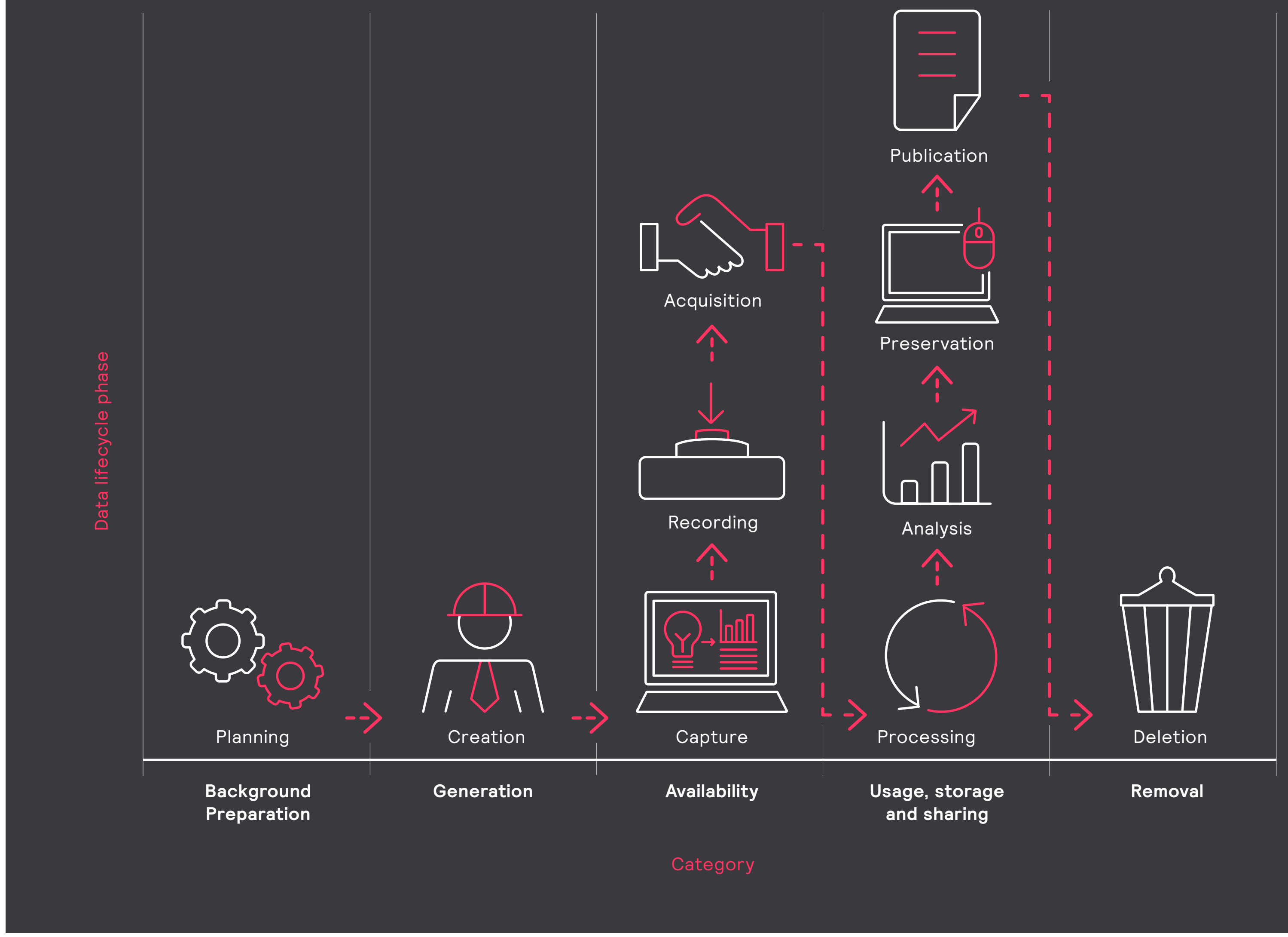
The value of data decreases over time
Although highly application-dependent, in most cases the more current data is the more valuable it is. Over time, data tends to become less relevant and less valuable, and eventually obsolete.

The value of data increases with quality
The concept of data quality includes characteristics intrinsic to the data itself, including accuracy, precision, consistency, completeness, trustworthiness (traceable lineage), and timeliness. True value can be derived from data only when the data is accurate, complete, in-time, and trusted. Poor quality data that leads to poor decisions or operational errors can be very costly.

The value of data increases when integrated with other data
There are multiple concepts related to integration. Increased value is often achieved when multiple data sources are combined together to create new data or key insights. Processing of data can include improving and ensuring the quality of data, adding metadata and indexing to improve context and accessibility, and transforming and enriching data to increase utility and value.

2.5. Data lifecycle

Across its lifecycle, data goes through several phases of processing, storage, conversion, and analysis. The overall sequence of phases constitutes *data lifecycle management*. While there are many different models for the data lifecycle, the following figure presents a comprehensive overview.



Planning
The first phase in the data lifecycle is planning, which identifies the type of data needed, data acquisition strategies, and the allocation of human, computational, and system resources.[13] This phase also includes development and acquisition of required resources (for example, software).

Data creation
The second step of the data lifecycle involves the actual creation of data. *Creation* means the direct, indirect, or automatic generation of data that represents characteristics of an underlying data *source*. The source can be an entity such as a physical environment, an organization, a device, or even a computer program that generates data relevant to a process or an objective.

Data capture
Once data is created, data capture comes into play, in which the created data is directly or indirectly measured. For example, imagine a camera that records vehicles driving along a street. The data creation process—vehicles driving along the street—happens automatically, and the camera captures the data. In this case, the data is not directly generated by the entity which is interested in it (the camera).

Data recording
Once data is captured, it is recorded. Recording allows the data to be available for use beyond the time window required for its capture.

Data acquisition
Data acquisition refers to the acquisition of data that has been measured or recorded by organizations outside the enterprise. This phase may or may not occur with the data lifecycle. When data acquisition from third parties is required, the creation, capturing, and recording phases could have taken place prior to or during the planning phase, since it is managed by a data supplier identified during the planning phase. Data acquisition takes place after data measurement or after data recording, and can be governed by a contract that defines how third parties are allowed to use the acquired data.

Data processing
All of the phases up to this point cover the availability of the data in its basic form. Once data becomes available, processing takes place. During processing, data is transformed, enriched, and combined with other data in such a way that it can be used to derive meaningful results and conclusions. Data processing is about formatting and data fusion, and does not directly result in assets that serve organizational objectives. That occurs in the next phase—data analysis.

Data analysis
In the data analysis phase, data is studied to identify patterns, make inferences, and reach conclusions that are meaningful to the final goals of the organization. Techniques from machine learning, big data, signal processing, image processing, and statistical analysis are applied in this phase. Referring back to the DIKW pyramid, data analysis turns information into knowledge.

Data preservation
Data preservation ensures that the data as well as the intermediate and final results derived from it are securely maintained, according to specified guidelines and rules, for a specific amount of time, either as backup or for potential future use. Data backups, additional metadata generation, documentation, and data archiving are all aspects of data preservation.

Data publication
Access to data and results derived from it can be granted to interested local or third-parties, within the scope defined by contracts and copyright law. Different users might be granted different access rights to all or a subset of the data/results.

Data disposal
At the end of the journey, the data must be deleted or archived from all locations where it is stored. No copies of the data are supposed to be kept for usage at any location once this phase is completed.

03.
Data management and governance

Data management is key to ensuring security and privacy in the growing IoT and smart systems space, which includes connected lighting systems for cities, workplaces, retail environments, and homes. Data management comprises all the disciplines needed to securely manage data as a valuable asset. Successful data management requires a data management architecture that enables a unified real-time view of all data assets, which are typically derived from disparate sources.

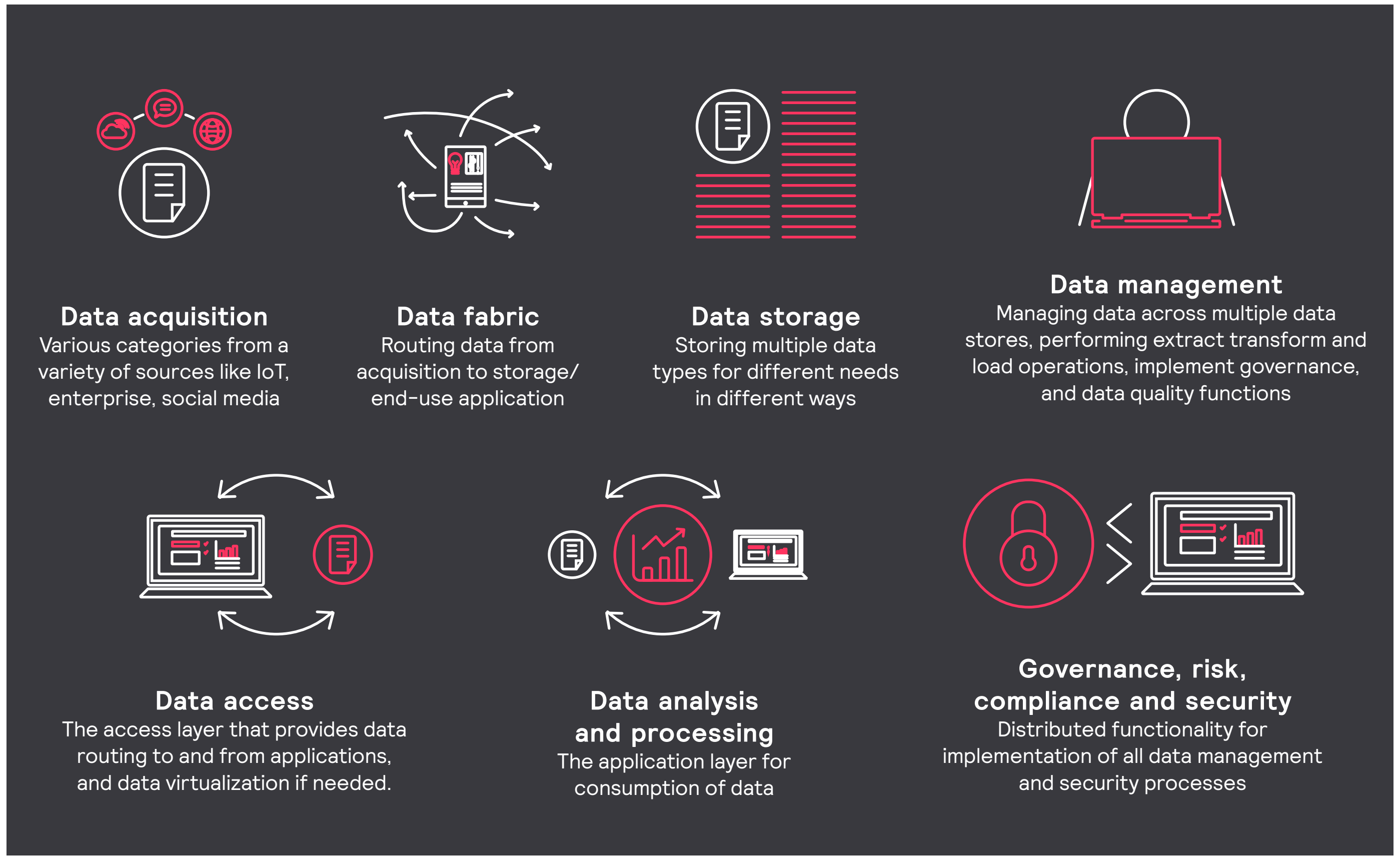
Ideally, a data management architecture manages an organization's data assets throughout its lifecycle, from the point of acquisition to consumption in end-user applications and on to deletion. Central governance and a company-wide data management strategy is needed to align data-related activities and maximize the value of data assets.

Successful data management must:

- Manage the data lifecycle
- Provide accessibility to all acquired data
- Ensure data quality
- Provide a unified view of all data assets to enable integration across system verticals and enterprise silos
- Include processing, transformation, and enrichment of data
- Specify data governance policies and procedures to ensure the highest possible level of data security and privacy

3.1. Data management in concept

A data management architecture manages data assets from acquisition to consumption. As the following figure illustrates, a conceptual data management architecture at the organizational level—for a city or an enterprise—can be complex, comprising multiple functions and capabilities. Data governance, the function primarily responsible for mitigating risk, enforcing compliance, and ensuring security, is a single vertical function that touches all layers of the architecture.



Data acquisition
Various categories from a variety of sources like IoT, enterprise, social media

Data fabric
Routing data from acquisition to storage/ end-use application

Data storage
Storing multiple data types for different needs in different ways

Data management
Managing data across multiple data stores, performing extract transform and load operations, implement governance, and data quality functions

Data access
The access layer that provides data routing to and from applications, and data virtualization if needed.

Data analysis and processing
The application layer for consumption of data

Governance, risk, compliance and security
Distributed functionality for implementation of all data management and security processes

Functions of a complete conceptual data management architecture include:

Data acquisition
The ability to acquire data of various categories (structure, size, speed) from a variety of sources (IoT, enterprise, social media).

Data fabric
The routing of data from acquisition to storage, end-use application, or both. This function can also include a data enrichment pipeline to enhance data quality.

Data storage
The ability to store multiple data types for different needs, including a storage sink or staging area for initial temporary storage, a data lake or reservoir for economically storing large-scale, lightly structured data, a data warehouse for primary structured data storage, and an enterprise data warehouse for critical enterprise data storage.

Data management
The ability to manage data across multiple data stores, perform traditional and nontraditional extract transform and load (ETL) operations, and implement governance, including data provisioning, lifecycle management, and data quality functions.

Data access
The access layer that provides data routing to and from applications, and data virtualization if needed.

Data analysis and processing
The application layer for consumption of data (big data sandbox discovery, business analytics, data modeling and transformation).

Governance, risk, compliance, and security
Distributed functionality for implementation of all data management and security processes.

This conceptual data management architecture is not implementation-specific, and it does not address issues related to private cloud, public cloud, on-premise, or hybrid implementations.

3.2. Data governance in concept

Governance, risk, compliance, and security is shown as a single vertical function across all layers in the architecture, as one unified architecture makes governance more straightforward.

Governance requires policies, processes, and procedures to function effectively, including:

- Policies to ensure legal, regulatory, and privacy compliance
- Security measures for accounts, networks, and data
- Policies and processes that ensure the value of data, as determined by the data value principles

Data governance practices should implement the data value principles outlined in section 2.4:

- The value of data increases with use: Ensuring the value of data includes providing access to data through provisioning, user access, and user authorization
- The value of data decreases over time: Data must be managed consciously through its lifecycle to maximize its value
- The value of data increases with quality: Because data must be complete, accurate and trustworthy to have value, it also involves data cleansing, enrichment, lineage, metadata, and related data quality functions
- The value of data increases when integrated with other data: Enterprises must use a combination of policy and technology solutions

Appropriate metrics must be used to monitor the effectiveness of the entire data management system, ensuring continuous improvement.

Actual system implementation will have an impact on security and governance. For example, privacy and regional laws and regulations depend on the physical location of cloud servers. Governance must therefore be geographically aware, and organizational policies must be aligned appropriately.

For governance, policies drive from the top down and eventually reach technical implementation in the data architecture. But just as technical solutions need to be policy driven, so policies need to be organizationally driven. Explicitly defined organizational roles and responsibilities are necessary to develop, deploy, and enforce effective policies.

Key roles for data governance include:

Governance body
A multidisciplinary team responsible for developing policies, directives, and strategies. The team should have executive authority, with representation from leadership (for example, a CIO); responsible personnel from security, privacy, and compliance teams; and technical resources from IT management, data management, and related functions. The governance body should report to the information owner and provide policy direction to the custodian.

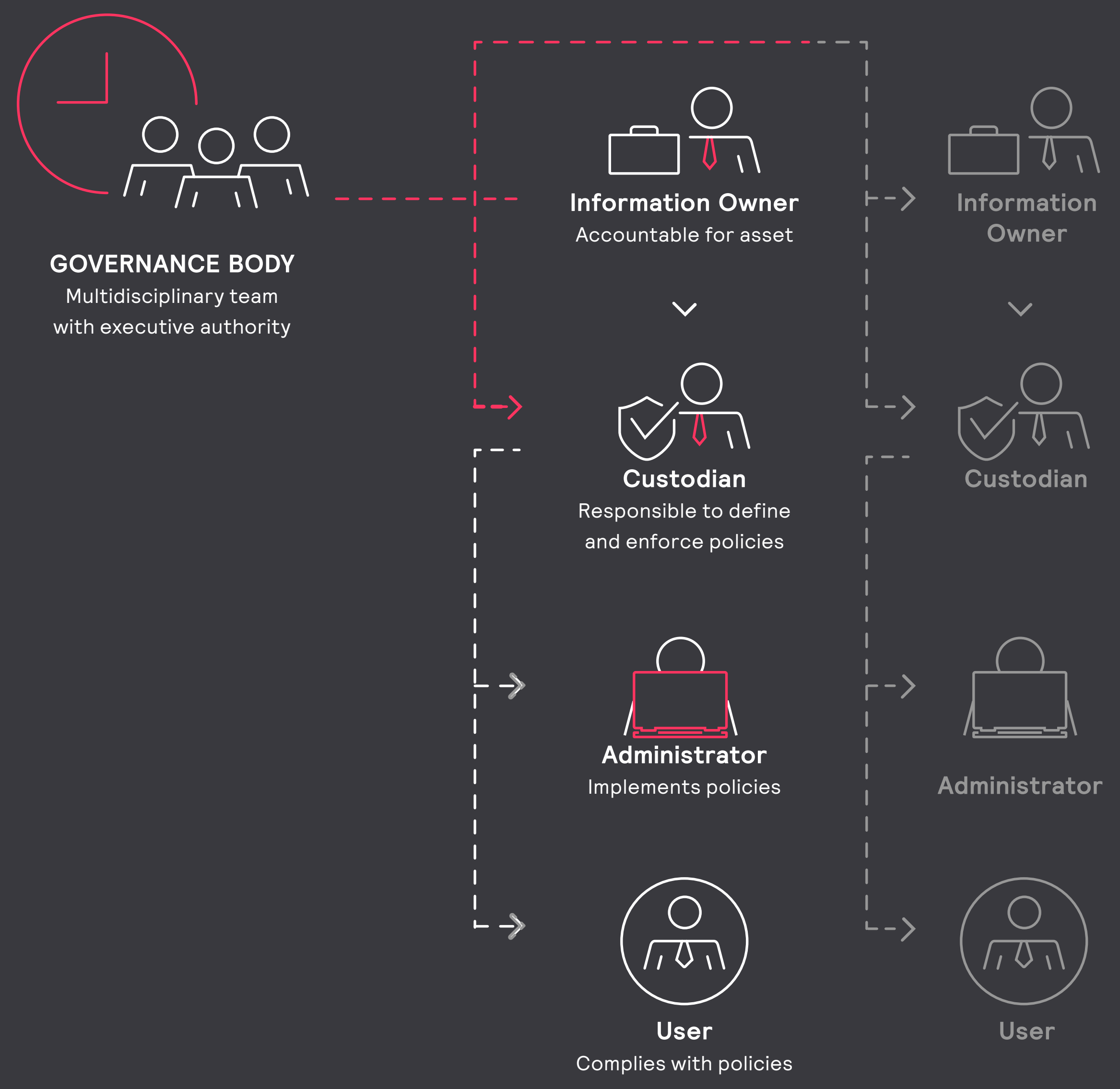
Information owner
An individual with accountability for data assets and the authority to approve rules put forth by the governance body. The information owner delegates implementation to the custodian.

Custodian
Receives policy direction from the governance body and defines and enforces the rules for assets on behalf of the information owner. The custodian delegates implementation to the administrator.

Administrator
Implements the rules for an asset and reports to the custodian.

User
Follows the implemented rules and provides requirements and feedback to the custodian.

The figure below shows the vertical hierarchy from information owner to user, which is often a business group or a sub-organization within a business group. There can be multiple information owners within a business group, and the custodian and administrator roles can cross business groups. Individual business groups can put governance policies in place, but doing so does not address the problem of siloing. Breaking down silos requires an organization-wide or cross-business group governance body with executive authority to set policies that encourage information sharing. Successful data governance also requires education and training for all employees, and a culture that recognizes and supports data governance principles and values.



3.3. Data governance considerations across enterprises

The data management architecture and governance structure outlined above describes the “awesome state” for an organization, whether that organization is a city, an enterprise, or an ecosystem partner. Data management across cities, enterprises, and ecosystem partners presents additional challenges:

- Standards for metadata and ontologies (the formal naming and definition of the types, properties, and interrelationships of data assets) are required to ensure semantic interoperability across organizations.
- Open architectures and APIs are required to enable data exchange among different systems in different organizations.
- Governance across organizations is required to ensure data integrity, value, and availability throughout an ecosystem. Governance could follow a centralized, top-down approach, or a de-centralized approach supported by automated transactions and governance management through blockchain-like technologies.

04. Privacy

Privacy is related to security and compliance but has its own definition, risks, and mitigation strategies.

Privacy often applies to a consumer's or user's right to safeguard personal information from use by others. Potentially vulnerable data includes, but is not limited to, data shared on social media, and any kind of demographic or personal data. In a smart system, vulnerable data could include data about a worker's habits (in and out time, salary and other HR data); a citizen's preferences, engagement in crowdsourcing applications, or movements around a city; or a shopper's buying habits, bank account, and credit information. In general, privacy is an individual's right to keep this and other sorts of personal data to herself.

One main goal of security is the protection of an enterprise, organization, or agency which may or may not store and manage vulnerable personal data. While privacy and security objectives can sometimes coincide, security policies and procedures may not address all privacy concerns. For example, a business or municipality may secure the personal data it stores from cyberattacks, but employees or officials inside the network may be able to review this data. A very common scenario is one in which an online retailer has top-of-the-line system security measures in place, but it freely sells personal data to realize secondary revenue streams.

4.1. Private data, confidential data, open data

Privacy and confidentiality are closely related. Privacy is usually understood to refer to an individual's data (private to the person), while confidentiality is usually understood to refer to a company's data (private to the organization). Very generally, we can talk about the relationship of collected data to privacy and confidentiality:

Privacy-related data
Personally identifiable and personally sensitive data that contains privacy invasive information. Restricted by law and ethical conditions.

Confidential data and trade secrets
Business-related data such as strategies, blueprints, formulas/recipes, and operational processes. May or may not be fully or partially restricted depending on a company's policies and agreements with partners and other third parties.

Open data
Data with no privacy or confidentiality issues. Such data may be collected and shared without restrictions.

4.2. Legal and ethical risks

Legal and ethical risks accompany the collecting, monitoring, processing, and storing of data derived from smart systems. At all times, there should be an actionable assessment of the potential benefits of collected data versus the potential harms that the collection, storage, and use of such data could have on the privacy of individuals and groups, as well as on the ethical norms and standards of society. These concerns hold for all affected constituencies, whether citizens in a municipality, employees in a workplace, customers in a retail environment, or residents at home.

Privacy concerns in smart cities are especially acute, given that misuse of collected data may lead to abuses of power and discrimination that could undermine the basic principle of human equality and lead to unwelcome dependency on those who control the data.

The UN Declaration of Human Rights, the International Covenant on Civil and Political Rights, and many other international and regional treaties recognize that all humans have the right to privacy so that they are protected against governmental or private parties intruding on their personal lives via surveillance or other types of monitoring. Only when there are serious societal reasons—public safety, national security, and so on—may individuals' right to privacy overruled by law, and then only under strict conditions. The infringement on personal privacy must be proportional to the importance of such action to society.

4.3. Privacy protection mechanisms

There are a number of privacy protection mechanisms specifically intended to protect private and confidential data, independent of security mechanisms that may also afford protections.

Anonymization is a protection measure where personal identifiable information is removed or masked from data sets. Pseudo-anonymization is a protection measure where the identifiable elements are replaced by artificial identifiers known as *pseudonyms*. In some cases, when re-identification of anonymized data is needed at a later stage, a key to reverse the measure is preserved in a separate secure location.

k-anonymity is a process in which data points are suppressed (removed) and/or generalized (replaced by a broader category) in such a way that each person contained in a data set cannot be distinguished from at least k-1 individuals whose information also appears in the data-set. An example of suppression is replacing data in a field with asterisks. An example of generalization is replacing the specific ages of individuals with an age range into which multiple individuals fall (between 35 and 50 years of age, for example). The value of k determines how many indistinguishable records the k-anonymity process will produce. For example, 2-anonymity would ensure that an individual's data set cannot be distinguished from at least one other individual in the data set (k=2, and 2-1 = 1), while 3-anonymity would ensure that an individual's data set cannot be distinguished from at least two other individuals in the data set (k equals 3, and 3-1 = 2).

Data perturbation is a data security technique that adds "noise" to databases, preserving individual record confidentiality. The techniques for adding "noise" may be value distortion approaches, which use some sort of randomization procedure directly on the values in a data set, or probability distribution approaches, which use some sort of algorithm for transforming the data. Data perturbation allows users to ascertain key summary information that is not distorted. This method is commonly used to protect the privacy of electronic health records.

Differential privacy, another promising technique, incorporates mechanisms that ensures that outputs of queries to a differentially private database yield the same conclusions irrespective of whether a particular individual's data is present in the database. That is, it addresses the problem of reaching conclusions about a population while learning nothing about an individual.

Even if privacy protection mechanisms are in place, there is still potential for data leaks. One situation that must be considered is the combination of data sets that have privacy preserving mechanisms built into them. For instance, a dataset may be anonymized, but by combining it with other information, anonymity may be lost. Mechanisms must be put in place to limit or prohibit leaking information by combining or further processing data sets.

05.
System security

Lighting systems are ever more connected and software-driven, becoming an important enabler for the Internet of Things. When connected, lighting systems have a vastly increased attack surface. Connectivity also opens the door for attacks that originate from a remote location. To protect company-confidential data, customer data, and other corporate assets secure, companies must design their connected systems with proper security measures in place, and ensure that they are properly deployed by installers and customers. As discussed in section 4, connected systems also collect and process potentially privacy-sensitive data. Regulations mandate that such data must be sufficiently protected.

Security is more than secure communications. In fact, secure communications are only a relatively small piece of the puzzle. Securing communications or other individual system features individually does not provide effective protections. Security organizations must consider the complete system holistically—including people, and processes to assess whether the system is secure as a whole.

A holistic approach to systems security has three main areas of focus, the *what*, the *how*, and the *when*:

Secure architecture for lighting systems
Describes what system architects need to do to design for a secure system (the *what*)

Secure development processes
Ensures that security is embedded in all phases of design, implementation, and deployment (the *how*)

Security lifecycle
Embeds security in all phases of a system's lifetime from manufacturing to deployment, during maintenance and during decommissioning (the *when*)

5.1. Secure architecture for lighting systems

One of the first steps in architecting a secure system is identifying the system's *attack surface*. The attack surface is the collection of entry points that can be used to attack the system. A clear view of the attack surface of a system helps to identify which components need to be secured and to decide which security controls to implement, aids in assessing the risks associated with using certain components and technologies, and helps to decide how security testers and (external) penetration testers can most effectively direct their security verification efforts.

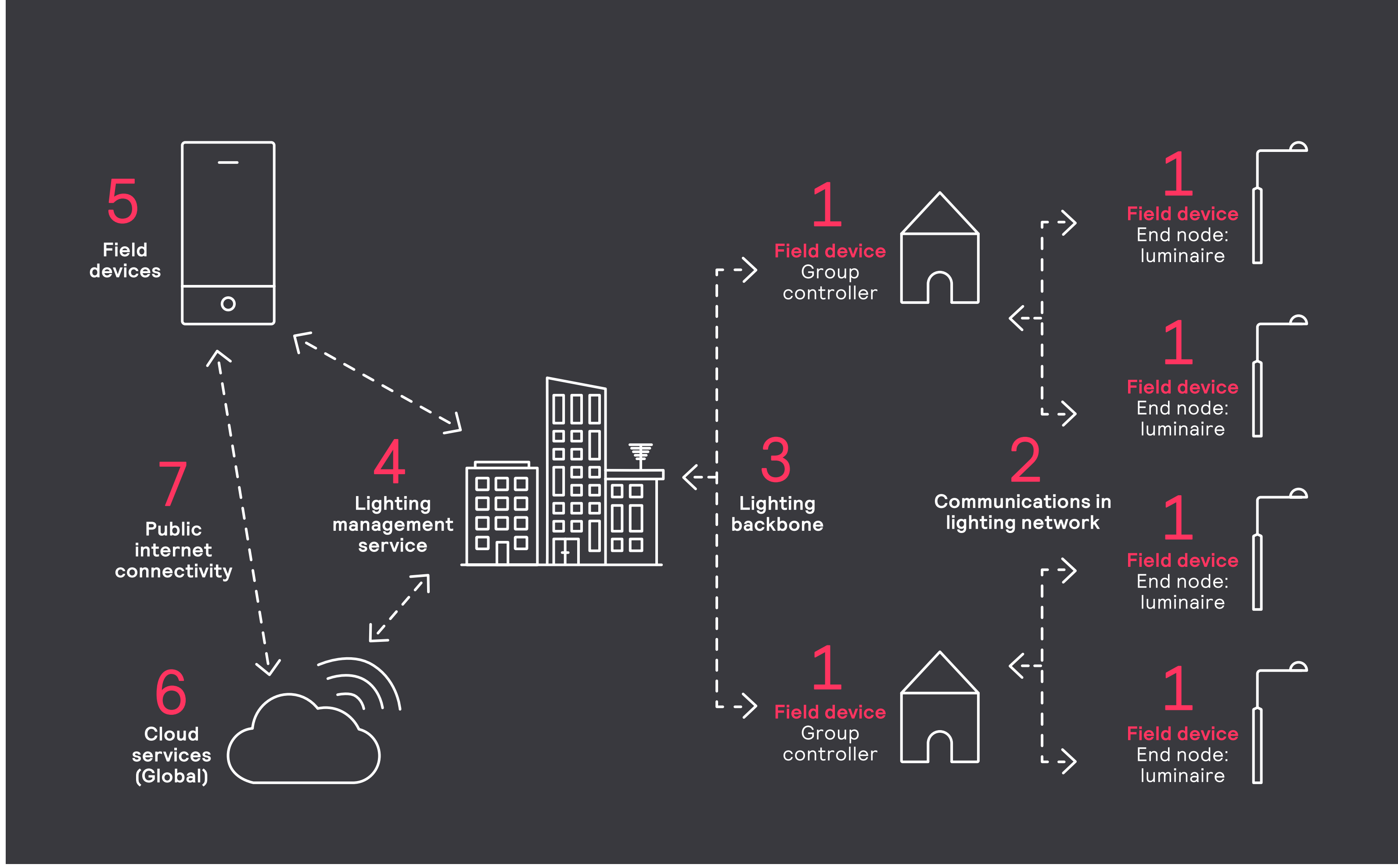
Formally, the attack surface of a connected system is the collection of the attack surfaces of each system component and the connections between them, along with all attack vectors. Each attack vector presents potential risks and security pitfalls, and suggests mitigation strategies and security controls for risk reduction. The following diagram shows a typical, somewhat simplified architecture for an indoor connected lighting system, and identifies seven main attack surfaces. Other connected lighting systems, such as street lighting and façade lighting systems, differ in the details, but many of the attack surfaces and the principles are similar across all connected lighting systems.

5.1.1. Field devices

Field devices (1 in the diagram) include end nodes (luminaires that contain controllers for communications with other devices) and group controllers (devices that provide an interface between the lighting network, which contains the end nodes, and the IP network). Field devices must be easy to install, must be replaceable when necessary, and typically remain in operation for several decades.

A prudent approach to securing a field device is to assume that hackers are able to physically breach its security. The effects of such a security breach must be limited to that device only—that is, compromising one device should not result in a compromise of the entire system. This approach is known as *containment*, and one of its principles holds that storing global secrets in a field device is a bad idea.

The likelihood that a hacker will compromise a device increases as that device type becomes more widely deployed. The hacker community is increasingly attracted to hardware attacks, as the tools to mount such attacks are becoming less expensive and therefore available to a larger audience. To mitigate the implications of a hardware attack, device debug interfaces should be disabled or password-protected. Encryption keys should not be transferred in the clear from one hardware component to another. Key material should not be stored in memory that can be easily read out by attackers. Whenever keys are used in memory, the memory should be erased as quickly as possible after the cryptographic operation finishes. Attackers with sufficient budget may be able to carry out side-channel attacks by injecting faults, measuring power consumption, or adjusting timing. Devices must carry software and hardware protections to prevent leaking cryptographic keys as the result of such side-channel attacks.



Since field devices are deployed for an extended period of time, they need to be firmware updatable. The firmware update process must be sufficiently protected to guarantee the integrity of the firmware update. For example, a device must be able to verify that a firmware update is genuine. If such a verification method is not in place, attackers can soft-brick a field device, causing it to error or making it completely unavailable.

5.1.2. Lighting network

Communications in a lighting network (2 in the diagram) may include communications between luminaires and group controllers, communications among luminaires, or both. Attackers may be able compromise the lighting network if they have a local presence.

The main security requirements for the lighting network are availability and integrity. Availability can be improved by ensuring that there is no single point of failure, and by increasing the number of communication paths between field devices. Traditionally, availability is a very difficult requirement to satisfy and to test.

Integrity can be ensured by signing messages using an encryption method that also incorporates a message authentication code (MAC). Devices must be configured to discard messages that have been replayed. Most protocols for communications between field devices already support integrity requirements. For example, Zigbee provides replay protection and ensures the integrity of messages using an authenticated encryption algorithm.

5.1.3. Lighting backbone

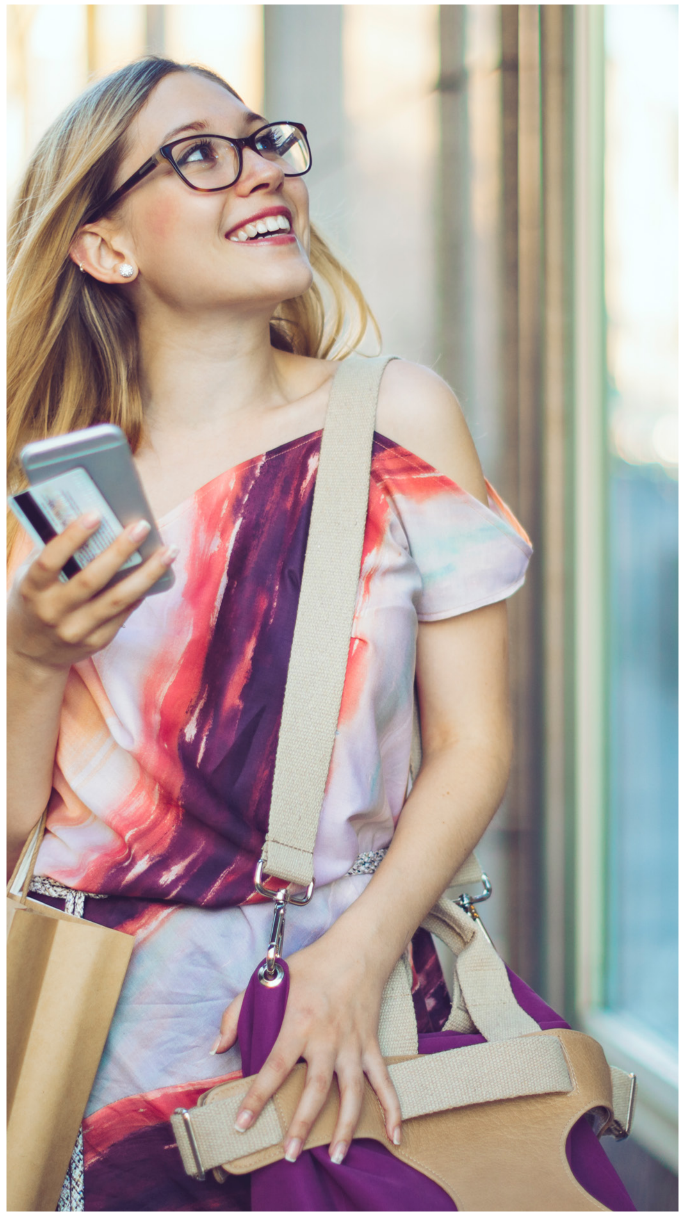
The lighting backbone (3 in the diagram) consists of IP-based communication between controllers and the local lighting management service. The devices that connect to the lighting backbone are less restricted in terms of hardware resources than the luminaires. Therefore, they can use standards and protocols for protecting Internet communication. These include protocols such as Transport Layer Security (TLS), which is an upgrade of Secure Sockets Layer (SSL), and virtual private networks (VPNs), which can verify the authenticity of communication partners and ensure data privacy and authenticity.

Network segregation, such as a virtual local area network (VLAN), can be used to separate traffic flows, reducing the exposure of the lighting backbone to the rest of an internal network. Network access control measures can further reduce the likelihood of network attacks on the lighting backbone.

5.1.4. Local lighting management service

An on-premise lighting management service (LMS; 4 in the diagram) can be deployed to provide an interface to access the lighting network from the local IT network. An LMS provides management and diagnostic services for local site managers to fine-tune and perform maintenance on the lighting system. In some cases, the LMS provides an interface to connect to building management services. These services usually run on a local web server, and attackers can exploit the typical web server vulnerabilities to gain access. Proper authentication and authorization of entities accessing the system is required to ensure the integrity of the system and system logs. Secure programming and testing against well-known web attacks is required to mitigate identified threats. Risk-based threat and vulnerability assessment must be performed regularly, and continuous patching is a must.[14]

5.1.5. Mobile apps and web-based applications
Mobile and browser-based interfaces to lighting management systems (5 in the diagram) are becoming more common. This brings new vulnerabilities when the level of trust on the device running the apps is unknown. Security credentials and sensitive data cannot be stored on such devices without proper countermeasures. Mobile apps should be tested for specific vulnerabilities (such as OWASP Mobile Top 10) before release.



Browser access from the web should be tested for common vulnerabilities such as cross-site scripting, cross-site request forgery, and so on. Mobile apps should be treated as a separate product and tested accordingly.

5.1.6. Cloud-based lighting management services
Increasingly, connected lighting systems in cities and buildings are shifting management from on-premise lighting management services to cloud-based lighting management services (6 in the diagram). This shift has both advantages and disadvantages in terms of security.

A cloud-based service provides an attack surface that can be accessed and attacked from the entire Internet. On the other hand, it is much easier and faster to perform security updates on a centralized cloud-based service than on individual on-premise management systems. Cloud-based systems can be used to ensure a single source of truth (SSOT) across multiple deployments—for example, for detecting cloned devices across multiple deployments.

Cloud-based lighting management services require extreme care in terms of security design, since they are very high-risk components that can affect multiple deployments. The security design of a cloud service should be scalable to ensure maximum availability. External penetration testing and auditing to established standards is extremely important.

5.1.7. Public Internet-based connectivity
Whenever public Internet connectivity (7 in the diagram) is used to interface lighting components (for example, from the cloud to a mobile app), a risk-based selection of applicable security mechanisms should be used. The security mechanisms used on the web have evolved from multiple failures and lessons learned, and reinventing the wheel is not recommended. Protocols such as TLS and Internet Protocol Security (IPSec) are suitable to bridge communication interfaces across the Internet. Additionally, the security design should provide the correct means for establishing trust on both sides of the interface, including key management and secure bootstrapping.

5.2. Secure development process

Lighting organizations need a set of processes to ensure that lighting systems are architected, designed, implemented, and deployed securely. In secure engineering practice, these processes are detailed in a Security Development Lifecycle (SDL) process or the Security Software Development Life Cycle (SDLC) methodology, such as the one published by OWASP. These methodologies offer a complete set of development processes, milestone deliverables, ways of working, and trainings.

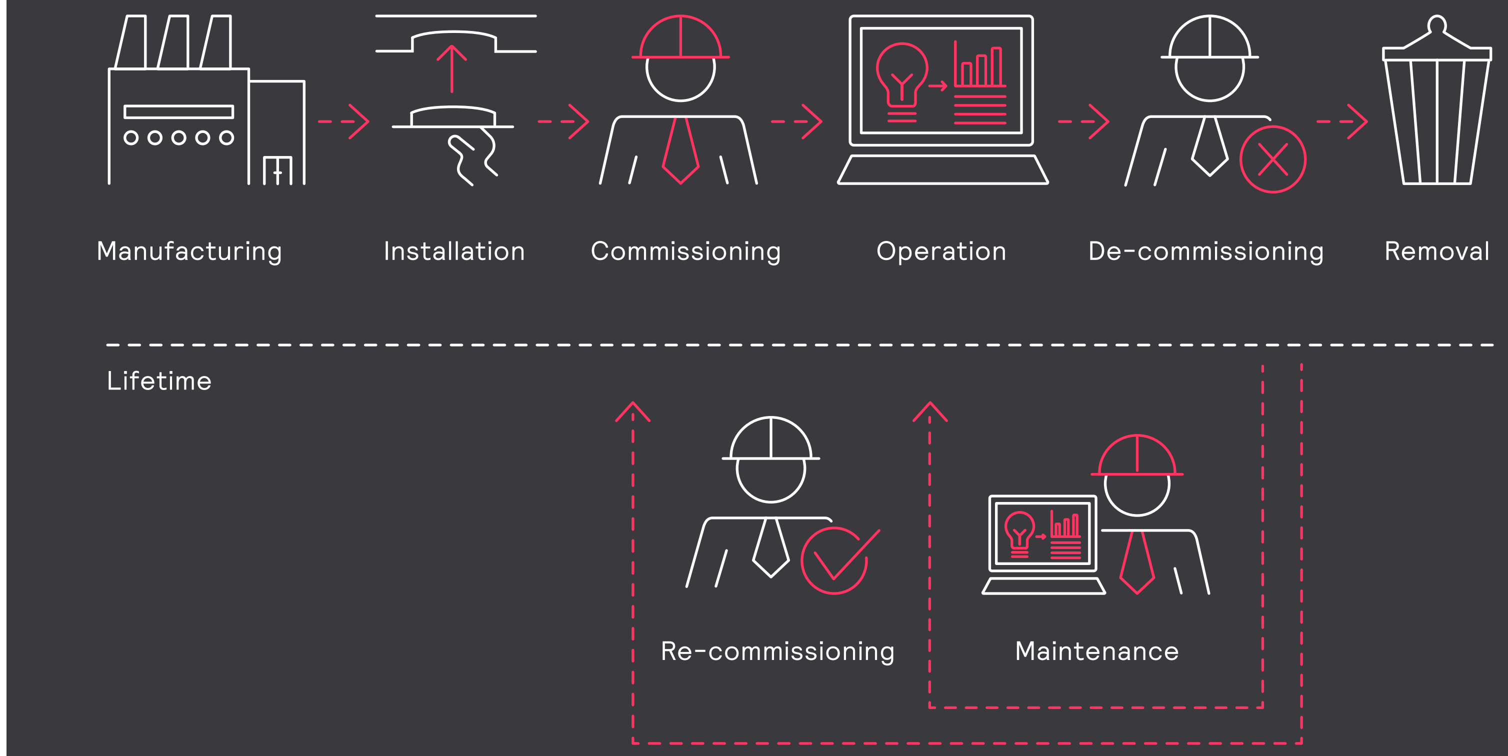
5.3. Security lifecycle

Every connected system follows a lifecycle that consists of several phases, typically manufacturing, deployment, maintenance, decommissioning, and final disposal. Security must be an integral part of each lifecycle phase, and the phases must be chained together to ensure that the system is secure continuously from cradle to grave.

The security of the whole system is only as strong as the weakest link. The secure manufacturing of devices with unique credentials is the basis for the security of a system over its lifetime. These long-term credentials need to be stored securely on the device along with proper key management in the back end. The commissioning or recommissioning of a system to define operational behavior depends on these original manufacturing keys. The operational security of the system is based on the commissioned keys, therefore proper authentication and authorization of the commissioning steps are essential.

Operational security is where the system spends most of its lifetime and needs to withstand attempts against breaches with secure communication protocols, authentication and authorization of peers, and proper key updates. Maintenance operations like software updates and device replacements should be performed with the proper security to maintain the operational security of the system. Decommissioning can be based on a combination of operational and manufacturing keys to enable secure removal of devices from existing systems. Proper operational key management is required when disposing of devices to prevent leaking keys and inadvertently admitting access to the operational environment.

Lighting systems in different domains have different security lifecycle requirements dictated by workflow requirements, leading to different architectural designs.



Notes

1. "The Internet of Things." *IT Glossary*, Gartner: http://www.gartner.com/it-glossary/internet-of-things/

2. Manyika, James, et al. "Unlocking the potential of the Internet of Things." McKinsey Global Institute, June 2015: http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/the-internet-of-things-the-value-of-digitizing-the-physical-world

3. "There will be 24 billion IoT devices installed on Earth by 2020." *Business Insider*, 9 June 2016: http://www.businessinsider.com/there-will-be-34-billion-iot-devices-installed-on-earth-by-2020-2016-5

4. Schneier, Bruce. "Data Is a Toxic Asset." *Schneier on Security*, 4 March 2016: http://www.schneier.com/blog/archives/2016/03/data_is_a_toxic.html

5. Bauer, Harald, Burkacky, Ondrej, and Knochenhauer, Christian. "Security in the Internet of Things." McKinsey & Company, May 2017: http://www.mckinsey.com/industries/semiconductors/our-insights/security-in-the-internet-of-things

6. Nuttall, Nathan, Goodness, Eric, Hung, Mark, and Geschickter, Chet. "Survey Analysis: 2016 Internet of Things Backbone Survey." Gartner, 5 January 2017: http://www.gartner.com/doc/3563218/survey-analysis--internet-things

7. *DAMA-DMBOK: Data Management Body of Knowledge (2nd Edition)*. Technics Publications, 2017: p. 17.

8. Huijbregts, Rick. "Re-imagining business value in a digital world." Cisco, May 2016.

9. "The biggest security threats coming in 2017." *Wired*, 2 January 2017: http://www.wired.com/2017/01/biggest-security-threats-coming-2017/

10. Doctorow, Cory. "Winter Denial of Service attack knocks out heating in Finnish homes." BoingBoing, 8 November 2016: http://boingboing.net/2016/11/08/winter-denial-of-service-attac.html

11. Banafa, Ahmed. "Three Major Challenges Facing IoT." SemiWiki.com, 25 May 2017: http://www.semiwiki.com/forum/content/6796-three-major-challenges-facing-iot.html

12. Definitions based in part on the Wikipedia entry on *data*, at http://en.wikipedia.org/wiki/Data

13. "Data Lifecycle Overview." USGS website: http://www2.usgs.gov/datamanagement/why-dm/lifecycleoverview.php

14. Rubens, Arden. "A Closer Look: OWASP Top 10 Application Security Risks." CheckMarx, 22 May 2017: http://www.checkmarx.com/2017/05/22/closer-look-owasp-top-10-application-security-risks/

Learn more about Interact
www.interact-lighting.com

interact