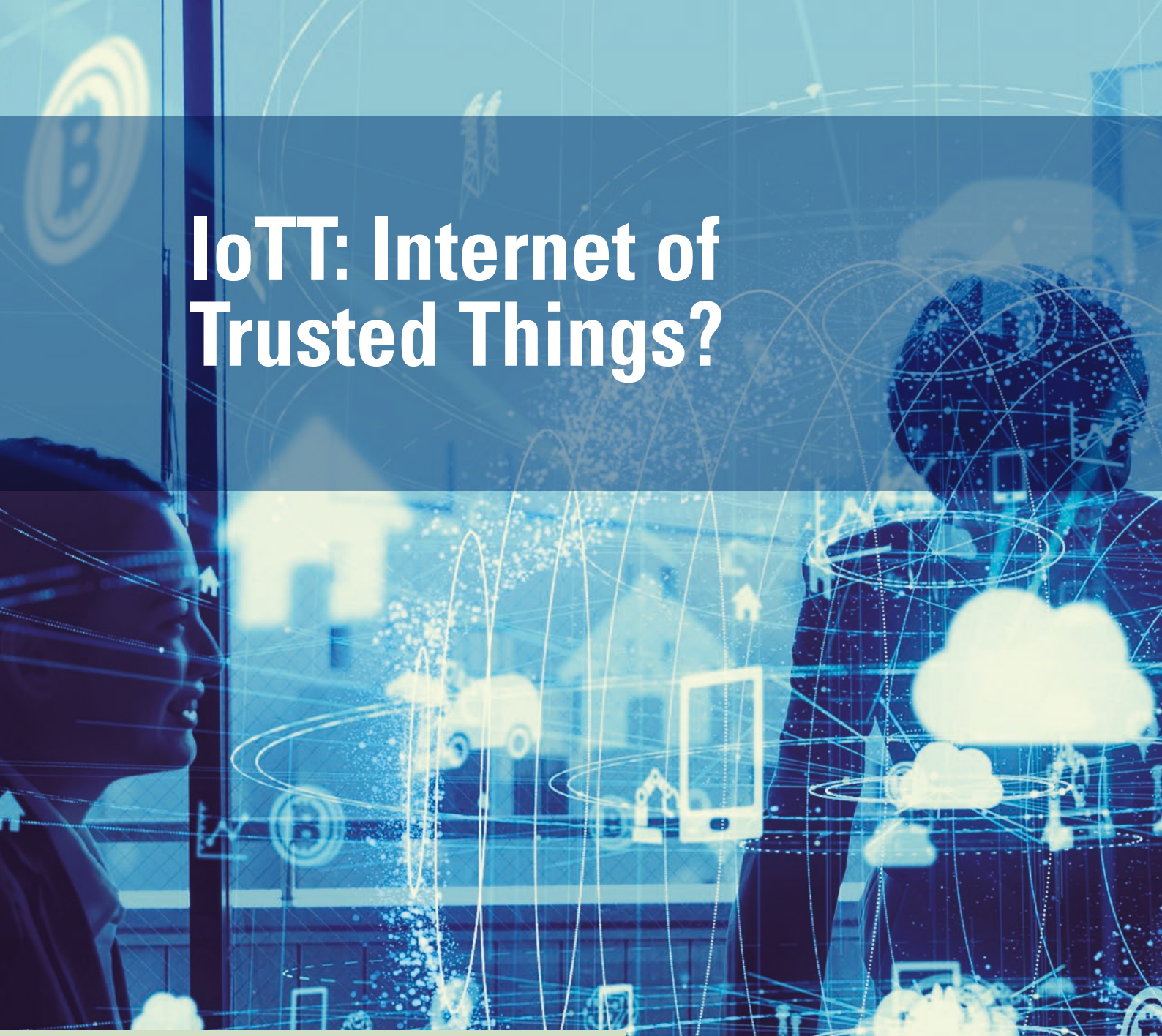
IoTT: Internet of Trusted Things?



The value of trust in data and information has increased with the adoption of new technologies such as the Internet of Things (IoT). In terms of control, experts often mention security measures of the devices, but forgo advising measures to control the data flowing between these devices. In this article we will highlight the importance of controlling data and give concrete examples along six dimensions of control on how to increase your trust in your IoT applications.

Marinka Voorhout MSc
is a senior manager at KPMG Data & Analytics.
voorhout.marinka@kpmg.nl

Remi Verhoeven MSc
is a manager at KPMG Data & Analytics.
verhoeven.remi@kpmg.nl

INTRODUCTION

Until recently, most big data initiatives focused on combining large internal and external datasets. For instance, an organization that sees a reduction in sales to customers under thirty, but have difficulty pin-pointing the reasons for this decline. Insights distilled from the combination of their internal customer data with external sentiment analysis based on social media then shows that this specific customer group has a strong preference for sustainability when purchasing products. The organization can respond to this insight by launching a new product or specific marketing campaign. Such initiatives are typically born as proof-of-concepts, but are gradually developing into more frequently used analytical insights. Some organizations are already moving towards transforming these (ad-hoc) insights into more business-as-usual reporting. The transformation from proof-of-concept to business-as-usual leads to the necessity of processing controls, consistent quality and a solid understanding of the content, its potential use and the definitions used in both systems as well as reports. This means that for the above-mentioned analytics on customer data and social media data, it is necessary to be certain that the data is correct. Data might need to be anonymized (for example for GDPR – data privacy requirements). It has to be validated that the data is not outdated. And the meaning of the data must be consistent between systems and analysis. The need for control, quality and consistency of data & analytics is growing, both from a user perspective, wanting to be certain about the value of your report, as well as from a regulatory perspective. So, it's critical to demonstrate your data & analytics is in control, especially when the data is collected from and applied to highly scalable and automated systems, as is the case for the Internet of Things.

The need for control, quality and consistency of data & analytics is growing

AWARENESS OF THE VALUE OF DATA IN CONTROL

Whether it is a report owner, a user of Self-Service-BI, a data scientist or an external supervisory authority, all require insight into the trustworthiness of their data. As said, this process of bringing analytical efforts further under control is a recent development. Initially, organizations were more focused on the analytical part than on the controlling part. But more importantly, controlling data the entire journey from source to analysis is usually complex and requires a specific approach for acquiring, combining, processing and analyzing data. So, although companies are increasingly proven in control, this progress is typically rather slow. The primary reason is that obtaining this level of control is challenging due to the complexity of the system landscape, i.e. the amount of application systems, the built-in complexity in (legacy) systems and the extensive amount of non-documented interfaces between those systems. In most cases, the underlying data models as well as the ingestion (input) and exgestion (output) interfaces are not based on (international) standards. This makes data exchange and processing from source to report complex and increases the time it takes to achieve desired levels of control. Organizations are currently crawling towards these desired levels of control, although we expect this pace to pick up soon: all because of the Internet of Things.

Wikipedia [WIKI18] defines the Internet of Things (or IoT) as: "the network of devices, vehicles, and home appliances that contain electronics, software, actuators, and connectivity which allows these things to connect, interact and exchange data." Or simply: IoT connects physical objects to the digital world.

IoT seems as much a buzzword as big data was a few years ago ([Corl15]). The amount of publications on the topic of IoT and IoT-related pilots and proof-of-concept projects is rapidly increasing. What is it about? An often-used example is the smart fridge, the physical fridge that places a replacement order via the internet at an online grocery store when the owner of the fridge takes out the last bottle of milk. While the example of the refrigerator is recognizable and (maybe) appealing, most of these sensors is far simpler and has much higher potential due to its scale for organizations than merely automating grocery shopping.

A practical IoT example of sensor data used in a very practical manner is developed in the agricultural sector. Dairy farmers have large herds that roam grasslands. Nowadays, cows in these herds are being fitted with sensors to track their movement patterns, temperature and other health-related indicators. These sensors enable the dairy farmers to pin-point cows in heat within the optimal 8-30-hour window, increasing the chance the cow will become pregnant and therefore optimize the milk production.

For organizations, IoT provides the opportunity to significantly increase operating efficiency and effectiveness. It can mitigate costs, for instance when used to enable preventive maintenance which reduces the downtime of machines – sometimes even by days. Sensor data can be derived from smart (electricity) meters and smart thermostats at your homes, the fitness tracker around your wrist. But similarly, also from the connected switches within railroads, smart grid power breakers or humidity sensors within large agricultural projects to fine-tune irrigation. All these devices and sensors collect and analyze data continuously to improve customer response, process efficiency and product quality.

Given this potential, it is expected that more and more companies are setting up initiatives to understand how IoT can benefit their business. We predict that the IoT will be commonplace within the next five years. The effect is that due to the number of sensors and continuous monitoring, the data volume will grow exponentially, much faster than the current growth rate. This means that the level of control, quality and consistency required will grow at least at the same rate. At the same time, IoT data requires more control than 'traditional' data. Why? IoT has its owns specifics, best illustrated by the two examples.

Understanding where risks lie, how reliable insights are and what impact false negatives or false positives have is therefore essential to embedding IIoT in the organization in a sustainable manner.

THE PLATFORM ECONOMY[1]

We see the sheer volume of IoT data, the fact that captured data needs to be processed (near) real-time and the amount of controls as the main drivers for the development and growth of so-called 'IoT platforms'. An IoT platform is the combination of software and hardware that connects everything within an IoT ecosystem – such an ecosystem enables an entity (smartphone, tablet, etc.) that functions as a remote to send a command or request for information over the network to an IoT device. In this way, it provides an environment that connects all types of devices. It also can gather, process and store the device data. To be able to do that in a proven and controlled manner, the platform should contain the required controls. Examples include having an anonymization function, the ability to set up access controls and having data quality checks when data is captured by the platform. And lastly, the platform allows the data to be either used for analytical insights or transferred to another platform or server. In some cases, the data generates so much value by itself that it is not shared but sold ([Verh17]). The platform than act as a market space where data can be traded. This is called 'data monetization' and its growth mirrors the IoT platform growth.

1 For the sake of this article, we limit our consideration of the IoT platforms to their data management functionalities.

CONTROLLING YOUR DATA – A CONTINUOUS EFFORT

Being in control of your data from source to analysis is not an easy effort. As mentioned, controlling data is complex due to differences within and between (IoT) devices or systems that capture data. The fact that data is exchanged within organizations, where there is a consistent use of data, is usually a challenge. But also, with external parties, which usually leads to even bigger differences between data, data quality and data definitions. Both internal and external data exchange therefore increase the need for e.g. data quality insights, data delivery agreements and SLAs). This is further increased by the growing regulatory requirement for data & analytics contributes. GDPR has been mentioned earlier in the article. Yet there are other less well-known regulations, such as specific financial regulations like Anacredit or PSD2. Yet, complex doesn't automatically mean that it is impossible. The solution is having a standard set of controls in place. This set needs to be consistently used within and between systems – including the IoT platform. To illustrate: when the data enters the IoT platform, the data quality must be clear and verified, the owner of the data must be identified and the potential (restrictions for) usage of the data must be validated. The continuous monitoring of and adhering to these controls means that organizations are perfectly capable of being in control.

Example 1: smart home & fitness trackers

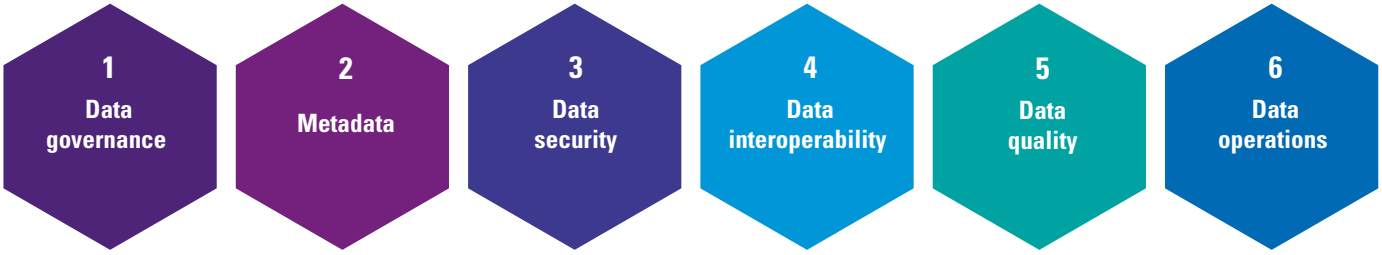
For both smart home devices and fitness trackers, it's typically the case that if the data stays on the device, controlling the data is mostly limited to the coding of the device itself. If the device is connected to an internal corporate system, control measures such as understanding where the device is located (e.g. is the device in an office or in a laboratory) must be added. And once the data is then exchanged with external servers, additional technical controls need to be in place to receive and process the data. Examples include security controls such as regular security keys rotation, penetration testing and access management. Furthermore, when tracking information on consumers that either reside in a house or wear a fitness tracker, privacy regulation increases the level control required for using data from these devices. This requires additional anonymization measures for example.

Example 2: Industrial IoT (IIoT)

Although, consumers are gaining understanding of the value of their data and require organizations to take good care of it, the industrial application of IoT is also growing. Companies in the oil and gas, utilities and agricultural industries are applying IIoT in their operations;We do believe there is an important role for the industry (manufacturers, platform operators, trade associations, etc.) to ensure that their products and services offer security by design and would come 'out of the box' with security measures in terms of encryption, random passwords, etc. ([Luca16]).

For example, imagine a hacker targeting a switch on the point of failure in an attempt to derail a train.

Figure 1. The six dimensions considered of most influence to managing IoT data.



In short, controlling the data from the device to usage means that different measures need to be in place of the data flow. These measures are related to different data management topics,[2] as visualized in Figure 1.

Ad 1
To ensure that changes to the infrastructure, requirements from sensors or processing and changes to the application of the data are adopted within the processing pipeline and throughout the organization, decent *data governance* measures should be in place. For instance, a data owner needs to be identified to ensure consistent data quality. This will facilitate reaching agreements and involve the people required to address changes in a structured manner.

Ad 2
The consistency of data is to be ensured by *metadata*: the information providing meaning or context to the data. Relevant metadata types such as data definitions, a consistent data model, consent to use the data as well as corresponding metadata management processes need to be in place. The need for robust and reliable metadata about IoT data in terms of defining its *applicability* in data analysis became painfully clear in a case we recently observed at an organization where data from multiple versions of an industrial appliance was blended without sufficiently understanding the difference between these versions. In this case, the manner of which an electricity metering value was stored in the previous version was with a 16-bit integer (a maximum value of 6553,5 kWh), while the metering value in the most recent version was stored with a 32-bit integer (a maximum value of 429.496.729,5 kWh). Since the values observed easily exceed 6553,5 kWh, the organization had implemented a solution to count the number of times the meter had hit 6553,5 and returned to 0 kWh. Their solution was simple: a mere addition of 6553,5 kWh to a separately tracked total for each of their devices. This however, had caused spikes in the results that seemed unexplainable to business users and caused confusion with their end customers.

Ad 3
Data security measures should be in place such as access and authentication management, documented consent of the data owner to let the data be used for a specific purpose, regular penetration testing and a complete audit-trail for traceability of the data ([Luca16], [Verh18]). Awareness of this topic is growing due to a stream of recent examples of breached security through IoT devices, such as the hacking of an SUV and a casino's aquarium ([Will18]).

2 The topics mentioned are part of the KPMG Advanced Data Management framework that embodies key data management dimensions that are important for an organization. For the sake of this article, we have limited the scope of our considerations to the topics most applicable for practically managing IoT data. A comprehensive overview of data management topics can be found here: https://home.kpmg.com/nl/nl/home/services/advisory/technology/data-and-analytics/enterprise-data-management.html

Being in control of your IoT data from source to analysis is not an easy effort

We do believe there is an important role for the industry (manufacturers, platform operators, trade associations, etc.) to ensure that their products and services offer security by design and would come 'out of the box' with security measures in terms of encryption, random passwords, etc. ([Luca16]).

Ad 4
For decent *interoperability* of data between sensor, processing nodes and the end user, exchange protocols to move the data need to be specified and documented, preferably based on international standards when available, such as ISO 20022 (standard for exchanging financial information between financial institutions, such as payment and settlement transactions). Important to consider are the physical constraints that traditional data processing don't often pose. In the case of the dairy farm, the farmer places a limited number of communication nodes on his fields. This means that the cows won't be in range of these nodes continuously. Furthermore, these field nodes are connected wirelessly to a processing node on the farm, which, in turn, is connected to the cloud infrastructure in which information of all cows worldwide is processed.

Ad 5
Even if the sensors, processing nodes and infrastructure are reliable, a good deal of attention should be paid to identifying to which *data quality* criteria these components should be measured against. In the case of IoT, the question is very much focused on what is important for a specific use case. For cases in which the information of interest is dependent on averages, such as body temperature or dimensions, such as distance travelled, missing out on 5 to 10% of potential measurements doesn't pose an enormous risk. On the other hand, in a scenario in which anomalies are to be detected obtaining complete data is essential. Examples include response times of train switches and security sensors. In other cases, the currency (or: timeliness) of the measurements is much more important when immediate action is required, such as in the case of dairy cows showing signs of heat stress. Determining which quality dimensions should be monitored and prioritized must be decided on a use case by use case basis.

Ad 6
Examples of data operations include storage replication, purging redundant, obsolete and trivial data, enforcing data retention policy requirements, archiving data, etc. Like data quality, organizations should start by identifying which specific *data operations* aspects should be considered. The best method to address this is through use cases, as these aspects are important for use cases that, for example, rely on time series analysis, (historic) pattern detection or other retrospective analyses.

The consistency of data is to be ensured by *metadata*

Better safe than sorry is never the best idea due to its complexity and volume

CONCLUSION

Increasing control of Internet of Things applications is necessary to apply trusted insights in (automated) decision-making. In practice, trusted insights derived by Internet of Things applications data often turns out to be a challenge. This challenge is best faced by not only focusing this control from a system or application point of view. Controlling secure access and usage, or the application of the insights from a privacy point of view is a good start for trusted IoT insights. But it also requires a fundamental reliance on the insights received, quality of data and the applicability of data per defined use case. This means that the total set of required measures and controls is extensive. When you increase the controls and measures, the trustworthiness of IoT insights increases. But it also important not to drown in unnecessary measures and controls. Better safe than sorry is never the best idea due to its complexity and volume. By using a sufficient framework, such as the KPMG Advanced Data Management framework, organizations know the total amount of required measures and controls – which mitigates the impulse to be over complete. And at the same time by having a complete framework, an implementation timeline for controls and measures can be derived based on a risk-based approach.

References

[Corl15] G. Corlis and V. Duvvuri, *Unleasing the internet of everything*, Compact 15/2, https://www.compact.nl/en/articles/unleashing-the-internet-of-everything/, 2015.

[Luca16] O. Lucas, *What are you doing to keep my data safe?*, Compact 16/3, https://www.compact.nl/en/articles/what-are-you-doing-to-keep-my-data-safe/, 2016.

[Verh17] R. Verhoeven, *Capitalizing on external data is not only an outside in concept*, Compact 17/1, https://www.compact.nl/articles/capitalizing-on-external-data-is-not-only-an-outside-in-concept/, 2017.

[Verh18] R. Verhoeven, M. Voorhout and R. van der Ham, *Trusted analytics is more than trust in algorithms and data quality*, Compact 18/3, http://www.compact.nl/articles/trusted-analytics-is-more-than-trust-in-algorithms-and-data-quality/, 2018.

[WIKI18] Wikipedia, *Internet of things*, Wikipedia.org, https://en.wikipedia.org/wiki/Internet_of_things, accessed on 01-12-2018.

[Will18] O. Williams-Grut, *Hackers once stole a casino's high-roller database through a thermometer in the lobby fish tank*, Business Insider, https://www.businessinsider.com/hackers-stole-a-casinos-database-through-a-thermometer-in-the-lobby-fish-tank-2018-4?international=true&r=US&IR=T, April 15, 2018.

About the authors

M.A. Voorhout MSc is a senior manager at KPMG Data & Analytics. She has been involved in many international data management projects. She has a background in trusted data analytics, master data management, data quality, as well as enterprise content management, and focuses primarily on data-related issues regarding regulatory compliance within various sectors.

R.S. Verhoeven MSc is a manager at KPMG Data & Analytics. He is an expert in the fields of data architecture, (asset) big data management and data quality management. He advises large and medium sized companies within several different industries on tactical and operational data management topics, the development of data monetization initiatives and regulatory compliance.