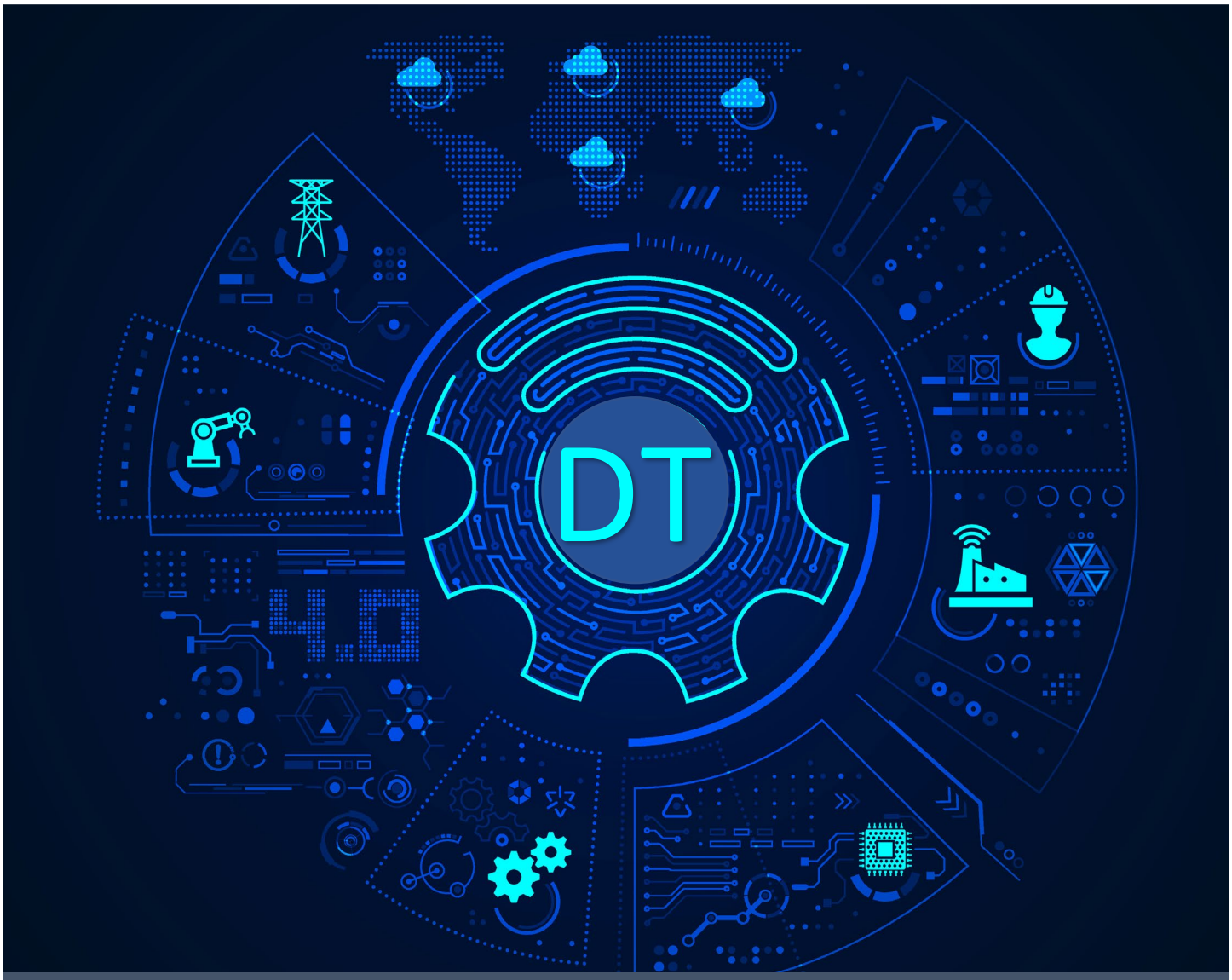
Matrikon®
Powering Interoperability.



Data Technology (DT): An Introduction

Bridging the IT/OT Gap for Enterprise-Wide Shop-floor Data Connectivity

Darek Kominek

Matrikon

Executive Summary

This whitepaper introduces Data Technology (DT), a new category of data interoperability software specialized for facilitating open and secure 3rd-party time-series data visibility across the enterprise. In a nutshell, Data Technology addresses three essential data connectivity and interoperability challenges at the edge:

- Data source federation
- Secure Network traversal
- Context preservation and enhancement (Information Modeling).

For brevity, this introductory Data Technology paper covers the core characteristics applications must exhibit to be characterized as DT. Topics covered include:

- Federation of data sources into a common access point
- Multi-level federation with data source context retention at each level in the federation hierarchy
- Establishing secure data connections with data sources across OT side firewalls without opening inbound ports
- Facilitating data connectivity using new and existing (legacy) components in new (greenfield) and existing (brownfield) installations
- mandatory support for OPC UA and optional support for other data connectivity standards

Of the many connectivity standards and technologies that could facilitate aspects of data connectivity in specific settings, OPC UA was chosen as the standard DT is based on. The reasons for this are also covered.

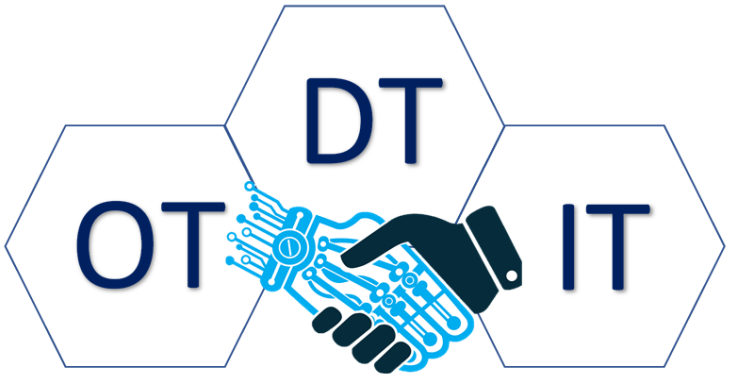
Examples of DT in this paper use Matrikon Dispatch Data Broker because this software is the first implementation of this exciting new technology.

Introduction

Currently, Industrial Internet of Things (IIoT) era solutions promise to deliver game-changing business benefits like real-time data-driven decision making, maximized operational safety, uptime and efficiency, new revenue stream creation, and much more. To realize such tantalizing benefits, it is necessary to address an essential requirement: the need for shop floor and field data and its context to be readily accessible to authorized entities throughout the enterprise.

Given how often "open, secure, and reliable enterprise wide shop floor data visibility" is touted as being the "foundation for all things digital", one might assume that this essential requirement is the first thing that would be addressed in today's control automation and IIoT related projects, but experience shows otherwise.

Instead, the challenge of establishing enterprise-wide data connectivity is often underestimated early on in IIoT and non-IoT projects leading to delays, unexpected expenses, and underwhelming results. The assumption often made by vendors and end-users alike is that the shop-floor data connectivity challenges (due to existing legacy or proprietary communications or common IT/OT gaps) are somehow solved by new IIoT technologies or are simply not considered until they pop up. While many new controllers, sensors, and applications using OPC UA and other open communications are available, businesses are reluctant to use them. Reasons for this include the high cost of replacement and because what they already have has worked so far.



Data Technology (DT)

This paper introduces a new category of software called Data Technology (DT). The main purpose of DT is to let companies start to use the 3rd-party data they have locked away within their infrastructures in a straight-forward, secure, and sustainable manner by helping bridge key IT/OT gaps. Matrikon Data Broker is used as an example of software categorized as true Data Technology.

Why DT? What's Different?

While many IT and OT focused applications overcome common data connectivity obstacles, the scope of the IT/OT gap challenges they address is quite narrow. Using such single-function applications often requires the use of a patchwork of components to 'build-up' a solution. The quality of such solutions often relies heavily on an integrator's experience. DT takes a different approach.

DT based solutions eliminate a diverse range of data connectivity challenges in an integrated manner "under the hood". As such, DT based solutions make the experience of setting up data connectivity straight forward while delivering standardized solutions that are mostly independent of the skill of the installers (some base knowledge is of course always necessary). This holistic DT approach to facilitating enterprise-wide shop-floor data visibility starts to help industry move beyond the complex and often inflexible architectures in use today.

Engineers and system integrators using DT based solutions should expect to drop DT into new and existing infrastructures, easily configure it, and start working with the data without having to work around the many IT/OT gaps commonly seen today.

What DT Addresses

At a high level, Data Technology covers three core aspects of data connectivity and interoperability:

1. Data Source Federation
2. Network Traversal and Open Data Connectivity
3. Context Preservation and Enhancement (Information Modeling)

Each of these is covered below.

Data Technology: Data Source Federation

Motivation

Once the need to make connections with individual data sources via their native protocols is met, it is essential to balance maximizing access to their data while managing the number of connections made directly to them.

Key considerations behind this include:

- **Security:** minimizing how many components directly connect to shop-floor systems, either on the shop floor or higher company levels to reduce the risk that one or more of those connections intentionally or unintentionally ends up disrupting the normal operation of the data sources.
- **Performance:** Shop floor data sources have varying amounts of computing resources like memory, processor speed and power. For economic reasons, the computer power engineered into these components is kept to a minimum and is used for both the operation of the data source and for communications with the outside world. Making too many connections to such data sources can impact the reliable operation of the data source.
- **Cost:** Some client-side applications include fees on a per-connection basis, while others only support a single connection. In both cases, the ability to expose multiple underlying data sources via a single connection is advantageous.
- **Sustainability:** With a growing number of data sources that require ever more connections to be made and maintained by a shrinking workforce (caused by various factors including an aging workforce and for economic reasons driven by competition and unprecedented events like the COVID19 pandemic) - it is prudent to minimize the number of individual connections made to each data source by creating a single access point that serves the needs of most, if not all future data consumers.

Given the universal nature of these considerations and the high frequency they crop up in projects makes 3rd-party data source federation a core functional requirement for DT applications.

Functionality

On the surface, the concept of data source federation is simple: expose data from multiple data sources into one location (many-to-one relationship) and make it available to multiple clients (one-to-many). In practice, however, there are additional options to consider:

Centralized Configuration

For ease of use and flexibility, and the growing need for performing configurations remotely – it is essential that the DT software throughout the enterprise can be

managed remotely. Whether data sources need to be added or removed – this must be done “on-the-fly" without disturbing existing connections.

While centralized configuration is convenient for installations on-premises, it is a requirement for use-cases where the DT applications are containerized. This is particularly important when integrating DT applications with popular cloud solutions like Microsoft Azure, AWS, Google Cloud, and others.

Multi-level Federation

Each DT federation node must be able to seamlessly federate individual data-sources along with other federation nodes so a common address space can be built. This is essential when your architecture involves multiple federation levels. For example, bringing together different data sources that are dispersed throughout the plant-floor that are sitting behind different firewalls into a common address space.

Scalability

With growing demand for greater operational visibility, the number of data sources on the shop floor continues to climb along with the number of systems requiring access to it. Therefore, a DT node must be able to handle high volume of data and connections.

Unlike most of the OPC applications from the OPC Classic era, which typically handled twenty to fifty thousand items, modern DT applications like Matrikon Data Broker must be scalable to work as easily with a few hundred values as they do with hundreds of thousands of them and sub second rates.

Data Technology: Open Connectivity & Network Traversal

Motivation

With the popularity of the IIoT movement in full swing, a lot of emphasis is placed on how to get data to the cloud and increasingly how to pre-process it at the edge before sending it to the cloud. This is natural given these are the early days of the IIoT era. However, it is equally important not to lose sight of the need to address data connectivity obstacles present in the majority of today's industrial infrastructure. While it is built on a combination of legacy components that use proprietary and older communications protocols along with modern components - this 'non-IoT era' industrial infrastructure is what keeps most plants and factories running.

To be useful in both brownfield and greenfield environments, DT software like Matrikon Data Broker must address the IT/OT gaps spanning the multi-generational spectrum of infrastructure components.

Open Standards & OPC UA

While many of the underlying infrastructure components comprise of proprietary communications, modern data connectivity calls for the use of open standard technology. This helps minimize interoperability, security, and vendor lock-in issues and helps keep architectures simple.

To create a level playing field across today's complex data connectivity architectures, DT classified applications will need to support various connectivity standards in general but will have to support one data interoperability standard in particular: OPC UA.

OPC UA was chosen as a core DT standard include:

- **Technology:** from a technology perspective, OPC UA addresses the full spectrum of data connectivity scenarios in a consistent and integrated fashion.
- **Adoption:** the OPC UA standard is already well recognized and adopted by companies and standards bodies around the world across industrial verticals and increasingly commercial applications too.
- **Community**: Led by the OPC Foundation (www.opcfoundation.org), an independent, not for profit, international organization, the OPC UA community is active, well-structured and continuously invests in growing and improving the OPC UA standard to best address the needs and threats that arise over time.

Functionality

Firewall and DMZ Traversal

The Digitalization trend is deepening dependency on data connectivity, which makes cyber security more important than ever. Firewalls are a key tool in the cyber security toolbox and are used to control communications between more-secure and less-secure areas. Used as part of the defense-in-depth strategy, firewalls are as common in OT environments as they are on the business IT side. Deploying DT

components in new and existing architectures enables people to work with firewalls in an integrated fashion instead of trying to work around them.

Bi-directional Communication Option

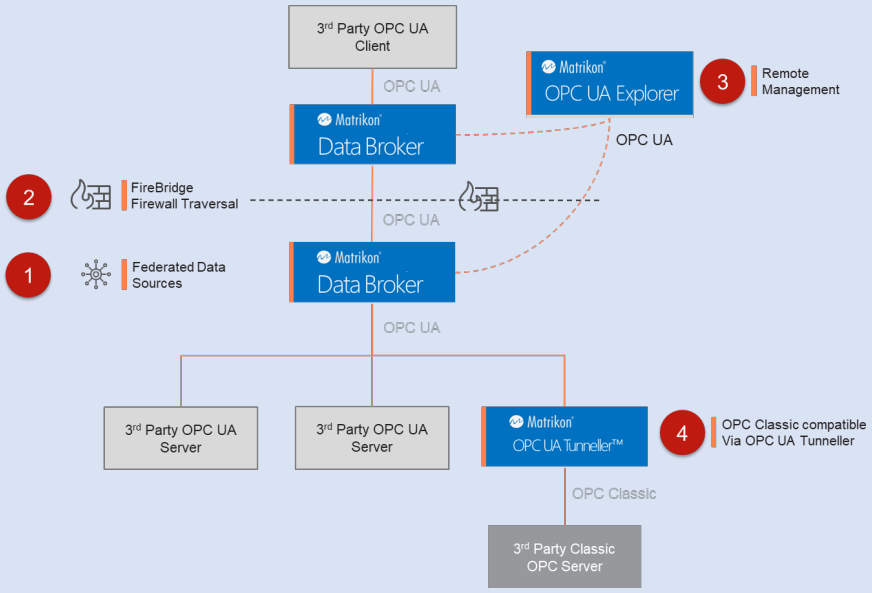
DT can use the OPC UA Client/Server communication model combined with the OPC UA ReverseConnect functionality. This gives administrators the option to choose whether they want to allow authorized clients to only read data from underlying data sources or perform both reads and writes.

Example Scenario:

Consider a scenario where data is needed from shop floor data sources operating behind a firewall that must keep its in-bound ports closed for compliance with the company's IT security policy. Instead of requesting a firewall exception from IT to get an in-bound port open, an administrator uses a DT application like Matrikon Data Broker to resolve the problem directly.

Here are the high-level steps:

1. Data Source Side (More Secure Side):
    a. Deploy an instance of the Matrikon Data Broker on the firewall protected subnet and federate the OPC UA servers. Matrikon Data Broker runs on Windows, Linux, and as a container maximizing the chances it would work directly in the environment in question.
    b. Configure the Matrikon Data Broker within a few mouse clicks to have it call out to specific OPC UA client(s) outside the firewalled subnet. These 'shoulder tap' type calls to the OPC UA Clients open a channel which the OPC UA Clients then use to establish a regular OPC UA Client/Server session.
2. Data Consumer Side (Less Secure Side):
    a. Configure the OPC UA client(s) outside the firewall that natively support OPC UA ReverseConnect calls to listen for the call configured in step 2.
    b. For OPC UA Clients that do not support OPC UA ReverseConnect functionality – deploy a Matrikon Data Broker DT node outside the firewall and configure it to accept the shoulder tap from behind the firewall. Once this is done, point all the OPC UA clients to the Matrikon Data Broker node to get all the data coming from behind the firewall.



Matrikon Dispatch Data Broker Data Technology solves multiple data connectivity challenges: 1. Federator: consolidates OPC sources for single-point access. 2. FireBridge: enables secure connectivity while keeping in-bound firewall ports closed.3. Remote Management makes it easy to manage all Data Broker nodes. 4. Matrikon OPC UA Tunneller enables access to OPC Classic components.

3rd Party Interchangeability & Multifunctionality

As seen in the Firewall example, a useful feature of DT components is the freedom to choose what aspects of their broad functionality to use in different infrastructure settings. If a particular 3rd-party component supports OPC UA ReverseConnect, it can be used directly – if not, the DT component can take care of it. However, if enhanced Information Modeling is required but the data source does not support it, the same real-time (OPC UA DA) component can be used to supplement that functionality as well, without the need for additional components.

Just as the OPC UA standard addresses a wide range of use-cases which are typically not required simultaneously in a given scenario, DT class applications include a broad range functionality to overcome IT/OT gaps and data connectivity challenges which may not need to be used at the same time. When such functions are needed however, the DT components are there ready to address them using the same unified environment without additional installations.

Data Technology: Data Context Preservation & Enhancement (Information Modeling)

Motivation

Today, there is an unprecedented abundance of data being generated on the shop floor and the volume of that data is only going to grow. While such data can be fed into massive data lakes and Data Scientists will come up with ever more complex algorithms to extract meaningful insights from Big Data, the IIoT era is breathing new life into the old GIGO principle (Garbage-In-Garbage-Out) and hence, the need to make data as meaningful as possible before feeding it into the cloud. For example, while accurate sensor readings are essential for generating useful insights about the state of what is being measured - so are the units and other attributes needed to put the raw values into context. While new components typically expose ample meta-data, this does not solve the challenge of maximizing contextual information because:

- Legacy systems often do not provide semantic data, but are either cost prohibitive to replace with newer components or modern replacements are not available
- Modern systems may provide semantic data but use an information model not compatible with what the rest of the system is using

DT can help alleviate such issues by enabling users to add contextual information to data from such data sources. Such functionality is based on OPC UA information modeling.

Functionality

Semantics and Information Modeling (IM)

The contextual information today's data sources offer ranges from none (only values are generated) to robust (with hundreds of attributes per item). DT applications help normalize how context for related data sources is expressed by enabling users to map original data-source data to new contexts (information models). Such models can be either generated by the users or pulled in standardized models based on OPC UA Companion Specifications.

Hence, DT should include the following functionality:

1. Must: Preserve the original data context from the OPC UA servers it federates
2. May: Enable data context enrichment as close to the data-source as possible by letting users dynamically associate existing data source data with one or more information models.

In both scenarios, such mapping enables data consumers (OPC UA clients) accessing the data via a DT application to work with the data according to the most appropriate views. This is especially important because it allows users to continue using even the simplest data sources (ex. Modbus based PLCs) they currently have, while presenting their data

Drill-Down Context Resolution

Just as it is vital to be able to build up a multi-level federation hierarchy, it is also essential to be able to freely drill back down through that hierarchy level-by-level, to see what information model (context) mapping was used at that level along with its current status. This functionality enables authorized OPC UA clients to:

1. See what the live status of underlying OPC UA sources is (AS BUILT) as they work their way down the federation hierarchy to troubleshoot where connectivity issues originated along the layered data sources. Seeing how this information comes from the original data sources, potential discrepancies between what the original configuration was and what is current are eliminated.
2. Discover what the underlying tree structure of the data sources is regardless of what part of the hierarchy they connect into.

Conclusion

The newly introduced Data Technology (DT) software category in this paper delivers vendor agnostic solutions for enabling secure, enterprise-wide 3rd-party data connectivity to and across the shop floor. As a comprehensive data connectivity and interoperability solution, DT applications like Matrikon Dispatch Data Broker resolve multiple connectivity challenges arising from longstanding IT/OT gaps and new connectivity challenges alike. Unlike typical solutions that solve specific parts of the overall data connectivity puzzle, applications in the DT category are multi-functional components that help keep architectures as simple to setup and maintain as possible, while taking care of the data connectivity obstacles 'under the hood'.

Products Discussed in this Whitepaper:

- Matrikon® Dispatch Data Broker (Download free trial)
- Matrikon® OPC UA Tunneller (Download free trial)
- Matrikon® OPC UA Explorer (Download Free Utility Client)

About the Author

Darek Kominek

Marketing Director at Matrikon (Honeywell)

Darek helps vendors, system integrators, and end-users best leverage OPC UA technology as the data connectivity foundation for their products, projects, and infrastructures respectively.
As a member of the OPC Foundation Marketing Control Board (MCB), he works with leadership on strategic marketing activities to drive OPC UA adoption, presents globally on behalf of the OPC Foundation, fosters collaboration with other standards organizations, and publishes articles about the advantages of using OPC UA in the IIoT/Industrie4.0/M2M space.
Darek holds a bachelor's degree in computer engineering from the University of Alberta.