CLOSER

Recommendations for standardisation of data exchange nodes along the entire chain

Authors: Edwin Mein, Tobias Beckman, Technolution;
Rodrigue Al Fahel, Felicia Hökars, CLOSER

Deliverable: D4.1

Published: December 2024

This report is part of the Smart Urban Traffic Zones project that seeks to develop intelligent solutions within the city to enhance flexibility in urban space utilisation, optimise transportation efficiency, and enhance traffic safety. This involves testing and evaluation of digital tools like geofencing, sensors, and digital signage within the city. The project is conducted in three steps, with the ongoing third step initiated in February 2023. It involves the participation of 24 project partners, including public entities such as cities, technology providers, OEMs, transport companies, and academia. The project is partially funded by the Swedish Innovation Agency, Vinnova.

VINNOVA
Sweden's Innovation Agency

Document elaborated with the support of the Smart Urban Traffic Zones project, 2022-02805, has received funding from Vinnova

Innehållsförteckning

Sammanfattning *3*

Summary *4*

Introduction *5*

Challenges **6**

Purpose **6**

Demarcation **6**

Method *7*

Results *8*

Data Producers **8**

Data Exchange Node (DEN) **12**

Data Processing Agreement (DPA) 12

Data Consumers **13**

Concluding remarks *17*

Appendix 1: Summary of the building blocks, risks and mitigations along the entire data chain. *18*

Sammanfattning

Denna rapport är en del av projektet Smarta Urbana Trafikzoner, som har fått finansiering från Vinnova. Syftet är att kartlägga och beskriva några av de viktigaste funktionella byggblocken (FBB) för datakedjan, inklusive aktiviteterna hos Dataproducenten, Datautbytesnoden och Datakonsumenten. Huvudsyftet med rapporten är att ge rekommendationer om standardisering av dataflöden som inkluderar datautbytesnoder.

Totalt identifierades 23 FBB: 13 som en del av dataproducentens dataflöde, 6 som en del av datautbytesnoden och 4 som en del av datakonsumentens dataflöde. Rapporten ger exempel på risker och åtgärder för varje FBB.

Följande rekommendationer, som anges i de avslutande kommentarerna, lyfter fram några av de viktigaste åtgärderna:

Dataproducenter

- Ha en övervaknings- och valideringsprocess samt tydligt definierade roller och ansvar, som överenskommits i exempelvis ett servicenivåavtal med datainsamlarna.

- Säkerställ interna kompetenser eller externa experter som är involverade tidigt i dataproduktionen för att säkerställa efterlevnad av dataskyddslagar, främst GDPR. Även om datainsamlarna förmodligen har processer för GDPR-efterlevnad måste Dataproducenter ha egna processer för att säkerställa detta. Fel kan leda till sanktionsavgifter och förtroendefrågor från användare, vilket kan äventyra användbarheten av datautbytesnoden (DEN) och datadelningen.

Datautbytesnod-leverantörer

- Inför en auktoriserings- och autentiseringsprocedur för att undvika obehörig åtkomst till systemet, vilket minimerar potentiella säkerhets- och integritetsproblem.

- Var medveten om vad som överenskommits i servicenivåavtalen gällande systemets drift, t.ex. avtalade drifttider och övervakningsprocesser. Missförstånd och orealistiska förväntningar kan leda till förtroendeproblem för både producenter och konsumenter. Tydlig kommunikation från DEN-leverantören rekommenderas för att säkerställa att Dataproducenter och Konsumenter är medvetna om förutsättningarna.

Datakonsumenter

- Ha en process för hur data ska användas och integreras i era system, samt hur olika datakällor kan kombineras och förfinas för att bättre passa ert ändamål. Läs metadatan innan du konsumerar data för att förstå om den är lämplig för ditt syfte, till exempel när det gäller dataformat.

Summary

This report is part of the Smart Urban Traffic Zone project, which has received funding from Vinnova. It aims to map and describe some of the main Functional Building Blocks (FBB) in the data chain, including the activities of the Data Producer, the Data Exchange Node, and the Data Consumer. The main purpose of the report is to provide recommendations on the standardisation of data chains that include data exchange nodes.

In total, 23 FBBs were identified: 13 as part of the Data Producer data chain, 6 as part of the Data Exchange Node, and 4 as part of the Data Consumer data chain. The report provides examples of risks and mitigation for each FBB.

The following recommendations, stated in the concluding remarks, highlight some of the main mitigations:

Data Producers

- To have a monitoring and validation process, as well as clearly defined roles and responsibilities, as agreed in, i.e., a service level agreement, between them and the data collectors.

- To have in-house capabilities, or external competencies, that are involved early in the data-producing process to ensure compliance with data protection laws, mainly GDPR. The data collectors will probably have processes in place to comply with GDPR, but ultimately, the Data Producers need to have processes as well to ensure compliance. If errors occur, this could lead to economic fines, and trust issues from users, which could risk the usefulness of the DEN and the data-sharing process.

Data Exchange Node Providers

- To have an authorisation and authentication procedure in place to avoid any unauthorised access to the system, therefore, minimising any security and integrity issues that could follow.

- To be aware of what is agreed upon in the service level agreements concerning the operation of the system, e.g., the agreed-upon uptimes, monitoring process, etc. Misunderstandings and unrealistic expectations could lead to a trust issue in the system for both Producers and Consumers. Clear communication from the DEN Provider is recommended to ensure that the Data Producers and Consumers are aware of the prerequisites.

Data Consumers

- To have a process in place for how to use and integrate the data into your systems and how to combine different data sources and refine the data to fit your purpose better. Read the metadata before consuming to understand whether the data fits your purpose and whether it is fit for you to use in your system, in regard to data formats, etc.

Introduction

A Data Exchange Node (DEN) is a critical component in the infrastructure for data sharing and communication. It acts as a central hub for secure and efficient data sharing between various entities such as systems, organisations, and devices. It ensures interoperability, data encryption, and robust authentication, supporting high throughput and scalability.

Many smart traffic zone applications will require several data sources and data users. These applications would need a DEN to collect and integrate data from sensors and traffic management systems, facilitating seamless communication and data sharing.

In the Smart Urban Traffic Zone project, DENs have been adapted and used to suit different pilots, such as to create situational awareness of loading zone usage in central parts of Gothenburg, Sweden. In this case, two DENs have been used: the City of Gothenburg and Technolution, a service provider.

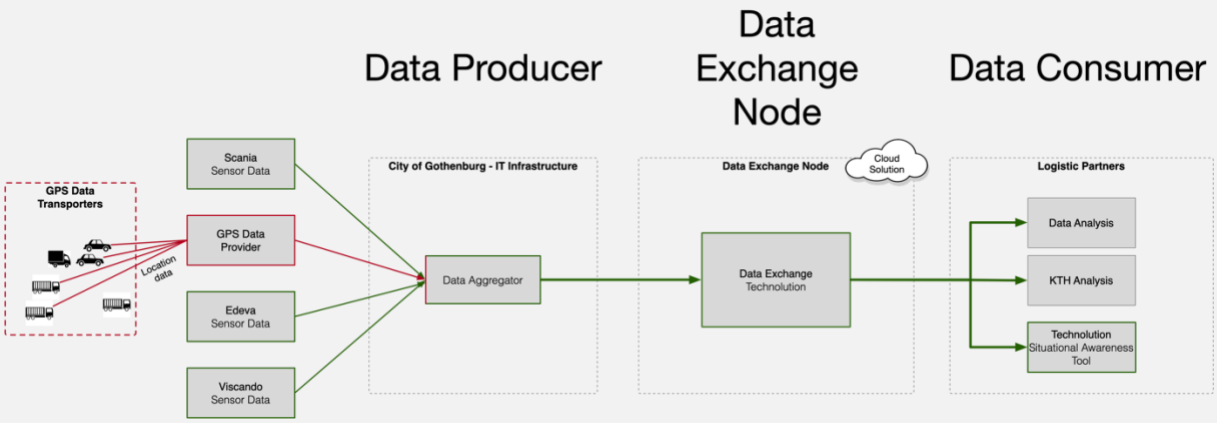
Data Producers are entities or devices that generate and provide data to the DEN. **Data Consumers** are entities or systems that receive and use data from the DEN for decision-making and operational purposes. Both data producers and data consumers play a central part in how well the data chain and DENs can function.

Pilot in Gothenburg

As a continuous story throughout this document, we will use the Pilot in Gothenburg as a frame for the different aspects of the recommendations in this document. We mark these references with a border around the text, as follows:

The lack of loading zones in Gothenburg leads to inefficient logistics, queues, double parking and possibly even unnecessary driving for the goods transported in the city. There are also temporary parking and shared areas that can be used for loading and unloading. The need for loading areas depends on the kind of goods to be delivered. Today we do not know what is needed for each delivery and we do not know how the available areas are used. We believe that we could improve the physical planning for logistics and make the transports through the city more efficient with a higher availability of data.

The figure below shows how the Data Producer, DEN and Data Consumer are applicable in the Gothenburg pilot case:



The City of Gothenburg acts as the Data Producer by collecting all relevant data from Scania, Edeva, Viscando, and GPS-tracks and processes this data internally with a data aggregator. The resulting aggregated data is shared with the DEN for further publication to the Data Consumers, such as KTH for analysis purposes and Technolution for the Situational Awareness Tool.

Challenges

Deploying DENs in traffic management faces several challenges including ensuring interoperability between diverse systems and integrating with legacy infrastructure, protecting sensitive data, and complying with regulations, managing large data volumes, and scaling infrastructure effectively, maintaining low latency and high reliability. Additionally, it requires technical expertise to manage complex systems, ensure consistent and accurate data, handle errors, and gain stakeholder buy-in and public trust.

Other challenges are related to governance and policies, including aligning the interests of diverse stakeholders, navigating bureaucratic processes, and fostering public-private partnerships. Additionally, cities must ensure that the deployment aligns with urban planning goals and addresses concerns around data ownership, privacy, and governance. There are also challenges related to the actual value of the DENs and the value of the services being provided through them.

Purpose

The purpose of this activity has been to identify and provide cities and relevant stakeholders with recommendations on common practices and how to standardise the entire data chain, including the production of data, the Data Exchange Node (DEN) and the consumption of data.

Demarcation

The main focus of this report is on the technological aspects of the data chain rather than recommending certain business model structures.

Method

The report is based on input from a series of workshops with project partners conducted by Technolution and Closer during the spring of 2024.

During the first workshop, the project partners were divided into different stakeholder groups: *1. Cities/road authorities, 2. Technology providers, and 3. Transport companies.* The purpose of the workshops was to get a better understanding of what systems the stakeholders were using, their experiences of Data Exchange Nodes (DENs), both as data providers and data consumers, and identify the main opportunities and barriers to use and scale up DENs.

Hybrid meetings with the stakeholder groups were held during the same day, starting with the cities/road authorities (appr. 2 ½h) and following with technology providers and transport companies (appr. 1 ½ h), each in separate rooms. Workshop questions were prepared by CLOSER and Technolution and sent in advance to the stakeholders, who filled in the answers and sent them back to the coordinators. Additional notes were taken during the workshop. The coordinators summarised the main results and findings.

The second workshop was held online, using the MIRO software, an interactive online whiteboard. A summary of workshop 1 was provided. The purpose of the workshop was to identify gaps between project pilot system solutions and what is needed to achieve scalable solutions after project completion, for DEN-related applications.

The workshop was divided into three iterations. Two groups were formed, one group involved in Pilot 1 in the SUTZ project, and the second group involved in Pilot 2. A similar exercise was done on a later occasion with Pilot 3 in SUTZ.

The first iteration focused on identifying the main building blocks needed for scaling up the pilot solution, focusing on the stakeholder role as data producers and data consumers in applications in which a Data Exchange Node was used. Building blocks were provided by the coordinators but stakeholders were able to add/remove building blocks. The second iteration focused on identifying the role of each stakeholder in this process and understanding who might be responsible for owning and/or executing each building block. The third and last iteration, asked each group to reflect, write down and discuss current pains/risks in the process and propose measures to mitigate the risks. Finally, the groups presented their findings to each other.

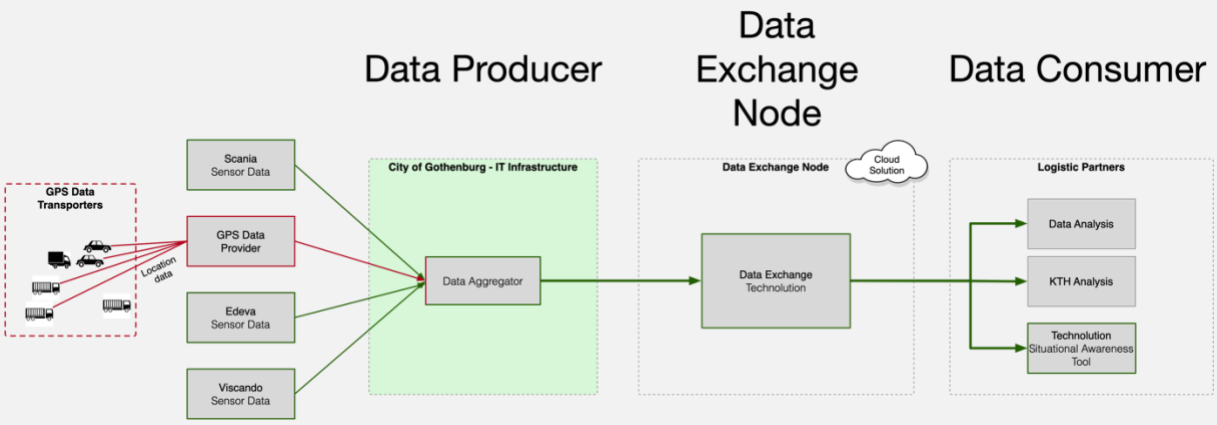
Results

This section will provide an overview of building blocks, risks and mitigations along the data chain, from Data Producers, Data Exchange Node Providers, and Data consumers. Appendix 1 summarises the findings in one table.

Data Producers

Data producers are in this report described as entities with agreements with the DEN that allow them to provide data to the DEN. Examples include the provision of traffic sensors, cameras, GNSS (GPS) devices, mobile apps, and public transportation systems data. The primary role of a data producer is to continuously send raw or processed data to the DEN, which is then used e.g., real-time traffic monitoring, analysis, and decision-making.

In the Gothenburg Pilot, the Data Producer is established within the IT environment of the City of Gothenburg.



This Data Producer in particular gathers the information from sensor providers, like Edeva and Viscando, from OEMs, like Scania and from their service contracts with GPS data providers. The Gothenburg Data Producer aggregates the raw GPS data samples into an GDPR valid data stream, removing all privacy related elements. The data from the Data Producer is safe to be shared with other stakeholders.

The following Functional Building Blocks (FBB) in the data chain were identified as part of the Data Producers´ responsibilities:

- Connect to data sources
- Collect data from sensors
- Quality assurance of data
- Validate data (e.g., standards, formats, etc.)
- Transform data into a standardised format
- Create data (e.g., geofenced parking zones)
- Send data into an internal database
- Clean data
- Combine data from several sources
- Aggregate data

- Make data GDPR safe
- Create historical data
- Publish data to the Data Exchange Node (DEN)

The Functional Building Blocks are interconnected. Below are the pains and gains identified for each FBB:

1. *Connect to data sources*

Description: Establish secure and reliable connections to various data sources (e.g., sensors, databases).

Risk: Connection instability or failure, leading to data loss or delays.

Mitigation: Set up backup connections, monitor them regularly, and ensure automatic reconnection if something goes wrong. Use the same secure communication methods throughout.

2. *Collect data from sensors*

Description: The DP manages and maintains sensors that produce the required data.

Risk: Sensor malfunction, data corruption, inconsistent data formats, lack of a sensor management plan from the DP.

Mitigation: Start with a sensor management plan. Implement regular sensor maintenance, calibrate sensors, and use standardised protocols for data collection. Introduce redundancy where critical data is concerned.

3. *Quality assurance of data*

Description: Ensuring a solid governance of the processes that produce the data (e.g., ISO9001) and includes monitoring of the process.

Risk: The data provider can break down, and there is no process or knowledge to repair an issue.

Mitigation: Document the processes that define and maintain the data provider's activities.

4. *Validate data (e.g., standards, formats, etc.)*

Description: Check that the data adheres to the required standards and formats.

Risk: Data not conforming to required standards could disrupt downstream processes.

Mitigation: Use validation rules that are aligned with industry standards and employ automated validation tools. If needed, establish a standardised validation framework to ensure consistency.

5. *Transform data into a standardised format*

Description: Convert raw data into a standardised format suitable for downstream processing, e.g., taking traffic counts in a CSV and transforming these into a European DATEX2 message.

Risk: Incorrect transformation can result in data loss or misinterpretation, which can make it impracticable to share data among stakeholders.

Mitigation: Implement a transformation process that follows a standardised format and includes error-handling procedures. Regularly review and update the transformation rules to align with industry standards.

6. *Create data*

Description: Generate new data (e.g., geofenced parking zones or digital permits with geographical zones) complementary to the data collected from sensors.

Risk: Errors in data creation could lead to incorrect or insufficient outcomes.

Mitigation: Automate the creation process where possible and validate new data against existing datasets for consistency. Use predefined templates and rules for data creation.

7. *Send/receive data into an internal database*

Description: Insert validated and transformed data into an internal database for storage. Raw data can also be inserted into the internal database, allowing for a reprocessing of the original data.

Risk: Incorrectly inserted data may lead to data loss or data corruption.

Mitigation: Use database features that allow you to undo errors if something goes wrong. Automate the data insertion process with error checks built in. Make sure the database structure is standardised, and all data follows this structure before being added.

8. *Clean data*

Description: Correct or remove inaccurate, incomplete or irrelevant data.

Risk: Cleaning too much might remove important data, while not cleaning enough could leave errors behind.

Mitigation: Use clear and consistent rules for cleaning data. Keep records of what you remove so you can recover it if needed.

9. *Combine data from several sources*

Description: Merge datasets from the same or different sources into one dataset.

Risk: Different formats or structures might cause problems when combining data.

Mitigation: Use the same formats and labels for data from all sources. Use tools that help merge data and resolve any conflicts.

10. *Aggregate data*

Description: Summarise or group data to make it easier to manage or analyse.

Risk: Summarising might exclude important details or introduce mistakes.

Mitigation: Use clear and consistent methods for summarising data. Offer and store both summarised and processed data when possible.

11. *Make data GDPR safe*

Description: Anonymise or pseudonymise personal data to comply with GDPR and protect privacy.

Risk: If not done properly, it could lead to privacy issues or legal problems.

Mitigation: Use reliable methods to anonymise data and regularly check that they work. Create GDPR procedures to ensure the rules are followed strictly.

12. Create historical data
Description: Save old data or snapshots for future use or analysis.

Risk: Poor management could lead to data loss or difficulty in finding old data.

Mitigation: Use consistent methods to store and organise historical data. Regularly back up the data and have a clear plan for how long to keep it.

13. Publish data to the Data Exchange Node (DEN)
Description: Share data with other systems or users by sending/publishing the data through a data exchange platform.

Risk: Data could be intercepted or accessed by unauthorised users during the transfer.

Mitigation: Use secure methods to send the data and ensure only authorised people can access it. Use standardised formats and secure protocols for sharing data.

Data Exchange Node (DEN)

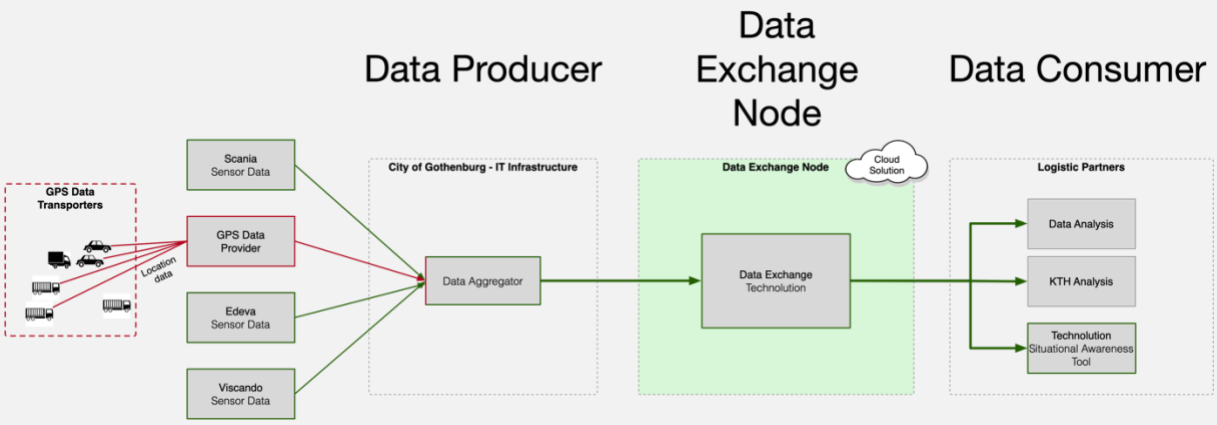
The following Functional Building Blocks (FBB) in the data chain were identified as part of the Data Exchange Nodes´ responsibilities:

- Publish data to (technical) stakeholders
- Check authorisation
- Maintain system/solution
- Improve the system/solution
- Send data into an internal database
- Follow up on compliance

Data Processing Agreement (DPA)

A Data Processing Agreement is required between any stakeholders operating under the General Data Protection Regulation. Under DPA agreements, all stakeholders' legal obligations and rights are stated, ensuring all security measures to protect vital information concerning the future of all organizations.

The Gothenburg Pilot uses the same Data Exchange Node as the other pilots as provided by Technolution in the Smarta Urbana project.



Here the Data Exchange Node receives the different data streams from the City of Gothenburg and makes them available to the Data Consumers. Each Producer and Consumer uses a unique entry point to discriminate what system is sending what data and what other systems acquire the data.

14. Publish data to (technical) stakeholders
Description: This is the Node's main function, publish the received data from external sources to authorised stakeholders, meaning forwarding the data to whom it concerns.

Risk: Publication of data fails due to technical problems.

Mitigation: Monitor the Data Exchange Node's functioning and verify that the publication of a known data stream is operative continuously.

15. Check authorisation
Description: Access to the Data Exchange Node is controlled by API keys (i.e., usernames and passwords). The Data Exchange Node checks this authorisation at the receiving part of

data from providers and the publication part to subscribers. Only authorised users can use the Data Exchange Node.

Risk: Unauthorised users may access the Data Exchange Node is not set up properly.

Mitigation: Validate and check the authorisation mechanism regularly, with more data also cyber security tests like Penetrations tests (PEN tests) can be implemented.

16. Maintain system/solution
Description: The service provider maintains the Data Exchange Node's server and software environment ensuring that it has the latest patches installed.

Risk: Failing to install patches will result in security risks or an unstable system.

Mitigation: Service and maintenance of the Data Exchange Node should be implemented at a regular interval with automated systems to check installed software versions.

17. Improve the system/solution
Description: The service provider improves the Data Exchange Node functionality by adding new functionalities and updating existing functionalities.

Risk: Without improvements, the system might become slow, unresponsive or fail to implement the desired function.

Mitigation: Regular maintenance including the development of new features and updating existing features should be part of the maintenance plan.

18. Send data into an internal database
Description: Data received from data providers shall be stored in an internal database – if allowed - for historical retrieval and analysis.

Risk: Data can be lost if the original provider has no backup of the data and trusts the Data Exchange Node to make the backup.

Mitigation: Implement an automated process that collects the data from the data providers and stores this – if allowed – in the internal database.

19. Follow up on compliance
Description: A special function that focuses on sending data from data users back to the providers, e.g., a truck shares its real-time location so the city can check if the truck follows the right route. At this moment a highly debated function, but technically possible.

Risk: Without a compliance check, the users of the data can use it for their purpose without consequences if they interpret the data differently.

Mitigation: Arrange agreements between service providers and users on how follow-up compliance shall be implemented.

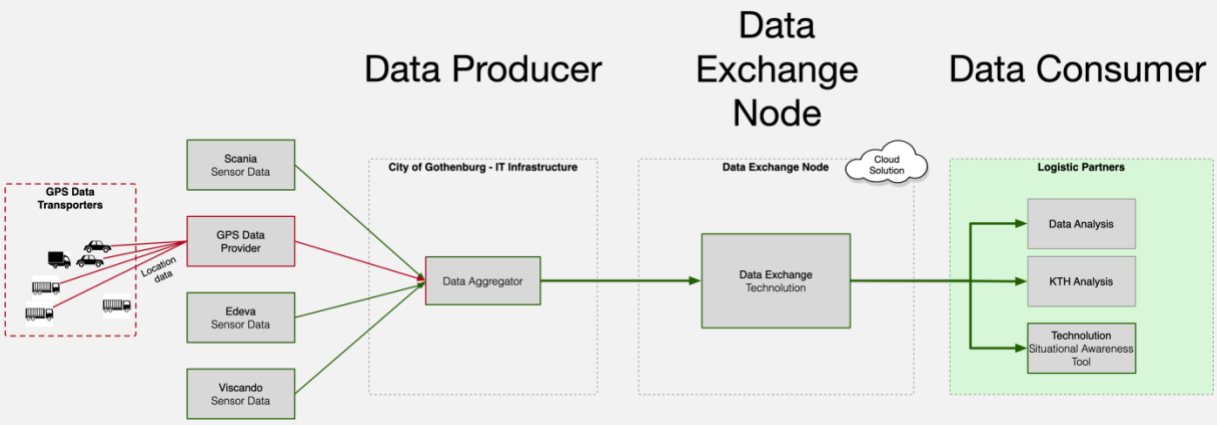
Data Consumers

Data Consumers are entities or systems that receive and use data from the DEN for decision-making and operational purposes.

The following Functional Building Blocks (FBB) in the data chain were identified as part of the Data Consumers´ responsibilities:

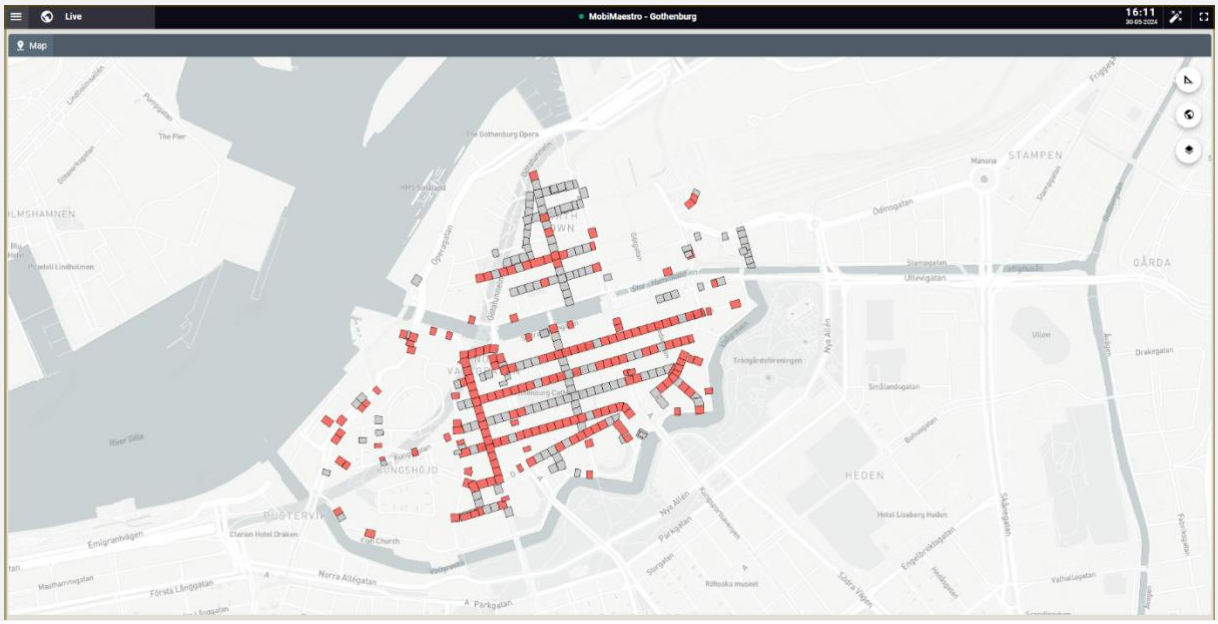
- Consume data from the Data Exchange Node
- Quality assurance and validation of data from the consumer of data
- Use data in internal processes (e.g., planning or automatic speed adaptations)
- Visualise data

In the Gothenburg Pilot, the Data Consumers take the data for different reasons from the Data Exchange Node. KTH uses the data for historical analysis of the service and how logistics partners could use the data. Technolution collects the data to provide a real-time and historical situational awareness view of the loading zone status in the city.

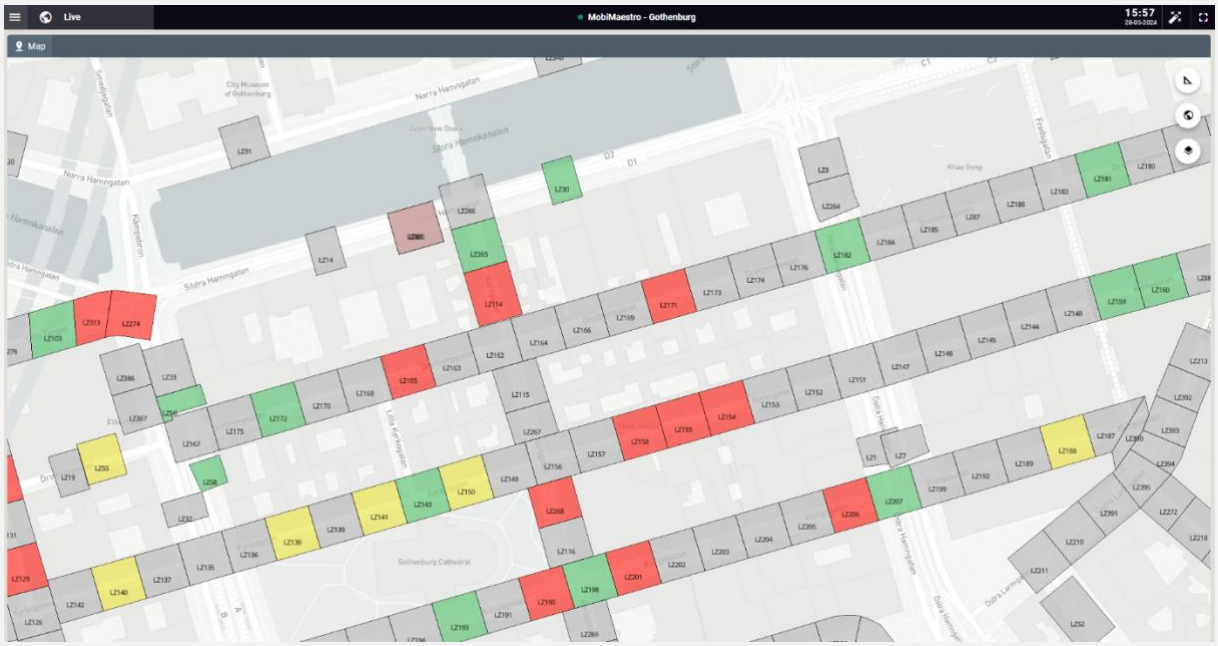


The real-time situational awareness tool is made available to the logistics partners directly, so they can use it in their day-to-day planning objectives.

An example of the real-time situational awareness tool is shown below with different colours for the parking zones in the city.



And when zoomed in at a particular area:



(red means that the zone is blocked by a vehicle, yellow means that a vehicle just left the zone, green means that the zone is available for parking and grey means that the status is unknown).

20. Consume data from the Data Exchange Node

Description: Connects and collects the data from the Data Exchange Node based on a subscription model, meaning, if new data is available from a provider (e.g., sensor) the consumer will be notified.

Risk: Impossible to implement the desired service based on the data from the service providers.

Mitigation: Make sure to read the metadata, i.e., the data about the data, and evaluate the relevance, quality, format, rights, and relationships of datasets, ensuring they align with your goals, are compatible with your tools, and comply with usage guidelines. Setup an automated mechanism that collects the data from the Data Exchange Node and reconnects if the connections fail.

21. Quality assurance and validation of data from the consumer of data

Description: Check the incoming data if it upholds the consumer's quality assurance standards as these standards may differ per consumer and reason.

Risk: Without quality assurance, it will be unclear how the desired service will operate.

Mitigation: Define KPI that can be used in the quality assurance process to check the incoming data.

22. Use data in internal processes (e.g., planning or automatic speed adaptations)

Description: This is the consumer's main task, using data in its internal process, e.g. collecting the loading zone occupancy values and using this in the internal planning process to align routes with loading zones.

Risk: Without the required data, the data consumer's process will fail.

Mitigation: Define how the internal process operates and what data (and quality) is needed from the data providers and implement the data consumption process.

23. Visualise data

Description: Create visual representations of the data, such as charts or dashboards, to make it easier to understand and use. This includes also in-car visualisations.

Risk: Poor visualisations could lead to misunderstandings or miscommunication.

Mitigation: Use standardised and clear visualisation techniques and ensure that visualisations are tested for clarity, ease of use and accuracy before being shared.

Concluding remarks

This report identifies building blocks and highlights some risks and mitigations for Data Producers, Data Exchange Node Providers and Data Consumers in the data producing, sharing and consuming process. The focus is mainly on how the system can be scaled up, be efficient and of high, or at least of satisfactory quality.

Data Producers are not necessarily those who collect the data but rather those entities that are responsible for the data that will be sent to the Data Exchange Node. From the examples for the pilots in SUTZ, data can be collected from multiple sensor providers and other organisations.

With multiple actors involved, problems will occur that could increase the risk of errors in data-producing and data-sharing activities. **Two key recommendations to Data Producers** are to:

- Have a monitoring and validation process, as well as clearly defined roles and responsibilities, as agreed in, i.e., a service level agreement, between them and the data collectors.

- Have in-house capabilities, or external competencies, that are involved early in the data-producing process to ensure compliance with data protection laws, mainly GDPR. The data collectors will probably have processes in place to comply with GDPR, but ultimately, the Data Producers need to have processes as well to ensure compliance. If errors occur, this could lead to economic fines, and trust issues from users, which could risk the usefulness of the DEN and the data-sharing process.

Data Exchange Node Providers usually have experience in creating secure systems that allow for seamless data exchange. In any case, Data Producers and Data Consumers should be aware, assess and only use DEN with reliable processes. **Two of the main things to look for are:**

- For the DEN to have an authorisation and authentication procedure in place to avoid any unauthorised access to the system, therefore, minimising any security and integrity issues that could follow.

- Be aware of what is agreed upon in the service level agreements concerning the operation of the system, e.g., the agreed-upon uptimes, monitoring process, etc. Misunderstandings and unrealistic expectations could lead to a trust issue in the system for both Producers and Consumers. Clear communication from the DEN Provider is recommended to ensure that the Data Producers and Consumers are well aware of the prerequisites.

Data Consumers are a cornerstone in this data chain. They are the recipients of the data and hopefully use the data to create value for their businesses, citizens, etc. To optimise the usefulness of the data, one **recommendation is to**:

- Have a process in place for how to use and integrate the data into your systems and how to combine different data sources and refine the data to fit your purpose better. Read the metadata before consuming to understand whether the data fits your purpose and whether it is fit for you to use in your system, in regard to data formats, etc.

Appendix 1: Summary of the building blocks, risks and mitigations along the entire data chain.

	Description of Building Block	Risk	Mitigation
Data Producers			
1. Connect to data sources	Establish secure and reliable connections to various data sources (e.g., sensors, databases).	Connection instability or failure, leading to data loss or delays.	Set up backup connections, monitor them regularly, and ensure automatic reconnection if something goes wrong. Use the same secure communication methods throughout.
2. Collect data from sensors	The DP manages and maintains sensors that produce the required data.	Sensor malfunction, data corruption, inconsistent data formats, lack of a sensor management plan from the DP.	Start with a sensor management plan. Implement regular sensor maintenance, calibrate sensors, and use standardised protocols for data collection. Introduce redundancy where critical data is concerned.
3. Quality assurance of data	Ensuring a solid governance of the processes that produce the data (e.g., ISO9001) and includes monitoring of the process.	The data provider can break down and there is no process or knowledge to repair an issue.	Document the processes that define and maintain the data provider's activities.
4. Validate data (e.g., to standards, formats, etc.)	Check that the data adheres to the required standards and formats.	Data not conforming to required standards could disrupt downstream processes.	Use validation rules that are aligned with industry standards, and employ automated validation tools. If needed, establish a standardised validation framework to ensure consistency.
5. Transform data into standardised format	Convert raw data into a standardised format suitable for downstream processing e.g., taking traffic counts in a CSV and transforming these into a European DATEX2 message.	Incorrect transformation can result in data loss or misinterpretation, which can make it impracticable to share data among stakeholders.	Implement a transformation process that follows a standardised format and includes error-handling procedures. Regularly review and update the transformation rules to align with industry standards.
6. Create data (e.g., geofenced parking zones)	Generate new data (e.g., geofenced parking zones or digital permits with geographical zones), complementary to the data collected from sensors.	Errors in data creation could lead to incorrect or insufficient outcomes.	Automate the creation process where possible and validate new data against existing datasets for consistency. Use predefined templates and rules for data creation.
7. Send data into internal database	Insert validated and transformed data into an internal database for storage. Raw data can also be inserted into the internal database, allowing for a reprocessing of the original data.	Incorrectly inserted data may lead to data loss or data corruption.	Use database features that allow you to undo errors if something goes wrong. Automate the data insertion process with error checks built in. Make sure the database structure is standardised, and all data follows this structure before being added.
8.Clean data	Correct or remove inaccurate, incomplete or irrelevant data.	Cleaning too much might remove important data, while not cleaning enough could leave errors behind.	Use clear and consistent rules for cleaning data. Keep records of what you remove so you can recover it if needed.
9. Combine data from several sources	Merge datasets from the same or different sources into one dataset.	Different formats or structures might cause problems when combining data	Use the same formats and labels for data from all sources. Use tools that help merge data and resolve any conflicts.
10. Aggregate data	Summarise or group data to make it easier to manage or analyse.	Summarising might exclude important details or introduce mistakes.	Use clear and consistent methods for summarising data. Offer and store both summarised and processed data when possible.
11. Make data GDPR safe	Anonymise or mask personal data to comply with GDPR and protect privacy.	If not done properly, it could lead to privacy issues or legal problems.	Use reliable methods to anonymise data and regularly check that they work. Create GDPR procedures to ensure the rules are followed strictly.
12. Create historical data	Save old data or snapshots for future use or analysis.	Poor management could lead to data loss or difficulty in finding old data.	Use consistent methods to store and organise historical data. Regularly back up the data and have a clear plan for how long to keep it.
13. Publish data to the Data Exchange Node (DEN	Share data with other systems or users by sending/publishing the data through a data exchange platform.	Data could be intercepted or accessed by unauthorised users during the transfer.	Use secure methods to send the data and ensure only authorised people can access it. Use standardised formats and secure protocols for sharing data.
	Description of Building Block	**Risk**	**Mitigation**
Data Exchange Node (DEN)			
14. Publish data to (technical) stakeholders	This is the Node's main function, publish the received data from external sources to authorised stakeholders, meaning forwarding the data to whom it concerns.	Publication of data fails due to technical problems.	Monitor the Data Exchange Node's functioning and verify that the publication of a known data stream is operative continuously.
15.Check authorisation	Access to the Data Exchange Node is controlled by API keys (i.e., user names and passwords). The Data Exchange Node checks this authorisation at the receiving part of data from providers and the publication part to subscribers. Only authorised users can use the Data Exchange Node.	Unauthorised users may access the Data Exchange Node is not set up properly.	Validate and check the authorisation mechanism regularly, with more data also cyber security tests like PEN tests can be implemented.
16. Maintain system/solution	The service provider maintains the Data Exchange Node's server and software environment ensuring that it has the latest patches installed.	Failing to install patches will result in security risks or an unstable system.	Service and maintenance of the Data Exchange Node should be implemented at a regular interval with automated systems to check installed software versions.
17. Improve the system/solution	The service provider improves the Data Exchange Node functionality by adding new functionalities and updating existing functionalities.	Without improvements, the system might become slow, unresponsive or fail to implement the desired function.	Regular maintenance including the development of new features and updating existing features should be part of the maintenance plan.
18. Send data into an internal database	Data received from data providers shall be stored in an internal database – if allowed - for historical retrieval and analysis.	Data can be lost if the original provider has no backup of the data and trusts the Data Exchange Node to make the backup.	Implement an automated process that collects the data from the data providers and stores this – if allowed – in the internal database.
19. Follow up on compliance	A special function that focuses on sending data from data users back to the providers, e.g., a truck shares its real-time location so the city can check if the truck follows the right route. At this moment a highly debated function, but technically possible	Without a compliance check, the users of the data can use it for their purpose without consequences if they interpret the data differently.	Arrange agreements between service providers and users on how follow-up compliance shall be implemented.
	Description of Building Block	**Risk**	**Mitigation**
Data Consumers			
20. Consume data from the Data Exchange Node	Connects and collects the data from the Data Exchange Node based on a subscription model, meaning, if new data is available from a provider (e.g., sensor) the consumer will be notified.	Impossible to implement the desired service based on the data from the service providers.	Set up an automated mechanism that collects the data from the Data Exchange Node and reconnects if the connections fail.
21. Quality assurance and validation of data from the consumer of data	Check the incoming data if it upholds the consumer's quality assurance standards as these standards may differ per consumer and reason.	Without quality assurance, it will be unclear how the desired service will operate.	Define KPI that can be used in the quality assurance process to check the incoming data.
22. Use data in internal processes (e.g., planning or automatic speed adaptations)	This is the consumer's main task, using data in its internal process, e.g. collecting the loading zone occupancy values and using this in the internal planning process to align routes with loading zones.	Without the required data, the data consumer's process will fail.	Define how the internal process operates and what data (and quality) is needed from the data providers and implement the data consumption process.
23. Visualise data	Create visual representations of the data, such as charts or dashboards, to make it easier to understand and use. This includes also in-car visualisations.	Poor visualisations could lead to misunderstandings or miscommunication.	Use standardised and clear visualisation techniques and ensure that visualisations are tested for clarity, ease of use and accuracy before being shared.