Dokumenttype
Aktivitetsbeskrivelse

Dato
1. december 2023

FT07.05_2024 Data Validation and Data Spaces

IoT-drevet forretningsdesign – digitalisering af virksomheder og samfund

Indledende oplysninger

Indsatsområde	IoT-drevet forretningsdesign – digitalisering af virksomheder og samfund
Institut	FORCE Technology
Titel	Data Validation and Data Spaces
Nummerering	FT07.05_2024
Version	1.0
Periode	Januar 2024 – december 2024
Kontaktperson	Henrik Hassing (hnh@forcetechnology.com) Michael Vaa (miva@forcetechnology.com)

Revisions

This is the first version of the activity description for 2024. It builds on activities and results completed in 2023.

Description

Objectives

The ambition of this activity has since 2022 been to establish a centre for data validation and to develop the necessary methods, tools, framework, and processes for doing data validation. The work in 2024 will finalize the work from 2023. The Data Validation Centre is a virtual centre of competences created to help enterprises and organizations utilizing IoT applications to ensure fit-for-purpose data. Fit-for-purpose involves many characteristics of data but in general means the quality of data is high enough for a given application. In IoT systems the data quality has long been known as a factor of uncertainty because data comes from many sources, often sensors. The data validation centre, and the online data validation service developed as part of the data validation centre, aims to create a well-defined process for working with data quality to help companies and organizations improve data driven decision-making by increasing the reliability of data and mitigating data quality issues.

The Data Validation Centre and associated methodologies have been central to this activity in the first years, but another concept of data sharing has recently emerged and gained traction at the EU level. The concept of data spaces. The ongoing work has been followed through different means but will in 2024 have a significant role within this activity as it is related to data exchange where it is important to ensure data is of high enough quality, as well as being able to show that data is validated and being able to guarantee a certain quality of shared data sets through common principles and agreements.

The activity will also contribute to international standardization in relation to IoT data standards where the creation of a potential new standard for IoT data validation will be explored. This will promote data validation as a crucial part of any data related activity especially when it comes to automated systems and IoT systems. Similarly, standardization plays a role in relation to the implementation and application of data spaces.

In 2024 the aim is to finalize the development of the online data validation service and processes, while also following international standardization in this area to align with new development and increase trust in the developed processes. Furthermore, the aim is to get more hands-on experience in the development of data spaces, as this concept continues to evolve, to demonstrate the benefits which these enable. This is done through the 5 overall activities described in the section below:

Content

The activity plan consists of a number of activities to support the achievement of the objectives above. These are:

1. **Further development of data validation tools**

In 2022 a roadmap was defined for the development of an Online Data Validation Service (ODVS). Phase 2 of the roadmap was followed in 2023 where the service was established, and the basic functionality implemented. In 2024 this work will be finalized in Phase 3. The following activities are planned:

a. Final development of ODVS in Phase 3 comprising activities 13-20 seen in Development and evaluation of tools, templates, and algorithms for analysing and visualizing data.
. The development will focus on completing the generic functionality according to the roadmap and to automate functionality where possible.
b. Development of concrete metrics for evaluating and validating the quality of data.
c. Development and evaluation of tools, templates, and algorithms for analysing and visualizing data.

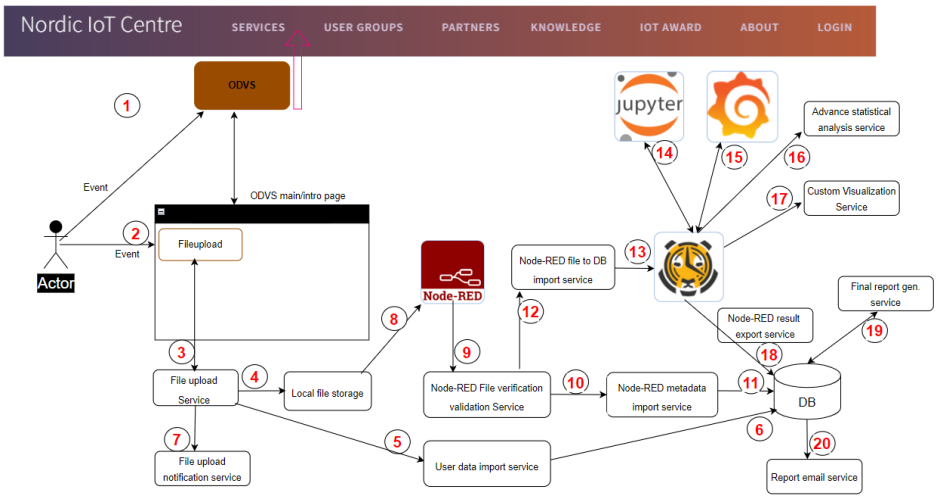

Figure 1: High-level ODVS architecture and development steps.

2. **Testing data validation methodology with 5 datasets from companies/organizations**

Data validation is a concept getting increased attention within IoT as more and more systems are installed and data quality issues arise. Therefore, it is necessary to get feedback from the ecosystem through the entire development, also when finalizing development of the tools and algorithms for data validation. This is to make sure the final tools and algorithms accounts for the needs and meets the expectations of IoT system users. The following activities are planned:

a. The validation process should be carried out on 5 different data sets from companies and/or public institutions, e.g. municipalities. This will help create new or better algorithms to identify errors in different kinds of IoT data and could give opportunities to apply them on more generic kinds of data to explore the capabilities.
b. The data quality metrics and algorithms will be tested with 5 different companies or organizations to evaluate them.
c. Knowledge will be gathered from the cases in this activity to support further improvements of algorithms and tools.

3. **Data Spaces international development**

Dataspaces specifications and framework are continuously evolving. This activity is focused on following and contributing to the current development through close collaboration with Alexandra Institute, International Data Space Association (IDSA), Fraunhofer, and other partners. Membership in IDSA facilitates this, as it allows for participation in various workshops, seminars, and conferences. Furthermore, it enables knowledge gathering on an international level as well as exchanging learnings and experience with other organizations working with data spaces. The activity will create services to be offered to Danish companies in exploring and actively implementing and using data spaces and to address challenges and help them resolve potential issues related to data spaces. Figure 2 shows the architecture and building blocks of data spaces which are being updated and extended by IDSA.

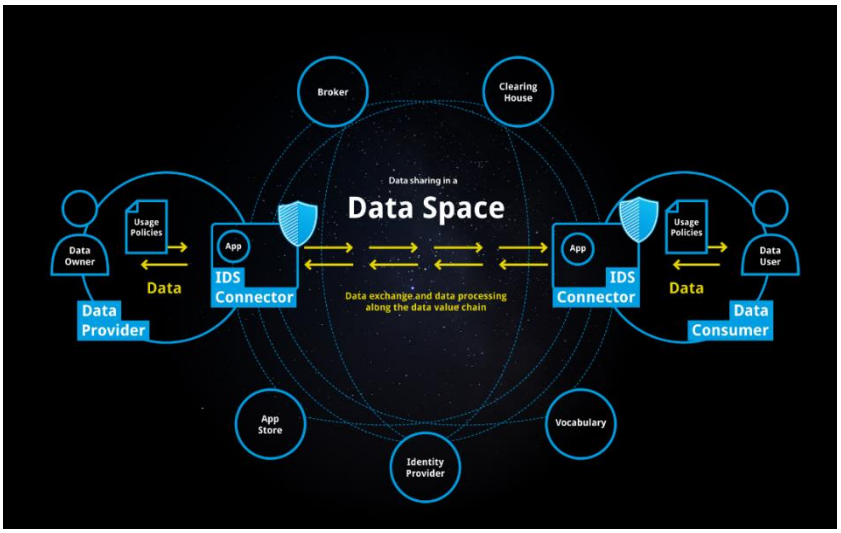

Figure 2: The IDSA Data Space architecture and building blocks for data sharing.

4. **Tracking the international development of data spaces:**

   a. Participation in 3-4 seminars, webinars, workshops on data spaces through IDSA or other entities to align with international development and gather knowledge.
   b. Participation in IDSA related working groups as a Danish RTO, e.g. the Architecture Working Group or Eclipse data space working group.
   c. Presenting the work from the activity below at 2+ occasions such as webinars, workshops, or other events to relevant Danish stakeholders and publishing and presenting relevant knowledge on suitable forums on development of the emerging trends and standards for data spaces with present and potential future use cases.

5. **Data Spaces tool development and demonstration projects**

Data Spaces is the underlying technology for enabling the EU Data Act and Product Pass initiatives. It's also the coming global ISO standard for trusted data exchange over the Internet. It will make transborder interoperability more agile, and it will soon be an industrial demand as seen in the automotive industry driven by the Catena-X network. In addition, data spaces are seen as a necessary vehicle for calculating CO_2 emissions and driving supply chain cooperation.

It has always been slow for IoT to be fully accepted in OT and SCADA environments because of lack of trust to supplier, security and quality of data and to some extend to SCADA vendor lockout. At the operational level it is expected that Data Spaces will be the glue that opens up for a more professional adaptation of efficient IoT in strategic sectors; an example could be Energinet. They need to have all small energy producers, users and car batteries communication into a common energy-sector data space to be able to manage all the different kind of datatypes. Data space is set to be the mediator, which can open up a more equal market competition between big and small enterprises at an EU level. At a broader level, data spaces will foster service innovation and cooperation as many problems today like GDPR, Cyber Security, Trust, Interoperability, will be standardized through the application of data

spaces and as such be implemented by design.

In collaboration with 2-3 Danish companies the aim of this activity is to work on an industry related use case that can be solved using a Data Space to demonstrate practical and operational applications of data spaces and to demonstrate the mutual benefits of companies sharing data in a data space. The activity will help benefit the industry at a broader level by providing a more agile way to use combined data to create new services but also to better collect the right data, process the data and feed them automated into a machine / robot. The activity will also provide input to Digitaliseringsstyrelsen who has requested operational demonstration cases to showcase the technology and what it can do and thus enabling Digitaliseringsstyrelsen to define their position and formulate any government initiatives on data spaces.

Data spaces are focused on interoperability and data sovereignty, and as such data validation will be an integral part of the activity. Activities include:

- **a.** Identify and engage 2-3 companies in a collaboration project to develop an industry related use case that can be solved using a data space to demonstrate a practical and operational application of data spaces.
- **b.** Development of a toolbox for practical implementation of data spaces
- **c.** Development of dedicated information sections/pages on data spaces, on Nordic IoT Centre's website and ODVS website to advertise data space services that FORCE Technology and Alexandra Institute can provide to Danish industry.
- **d.** Plan and execute online/offline seminars for Danish industry for knowledge dissemination and discuss the potentials, challenges and benefits of data spaces.

6. **Standardization development contribution and knowledge gathering**.
   It will be beneficial to make data validation processes uniform across industry, i.e. standardized, to ensure comparability between validations, and it will benefit industry since it can be shown that the validation is done according to commonly agreed methods and processes. The work on standardization of data validation will continue in 2024 and build on the work done in 2023:

   - **a.** Participation in Danish Standard "S-441 Cyber- og informationssikkerhed" standardization committee through which also the work in "JTC 1/SC 41 – Internet of Things and Digital Twins" is followed.
   - **b.** Contribution to standards development related to data validation and data quality.
   - **c.** Research on other relevant standards within data quality of which there are currently some in development.
   - **d.** Knowledge gathering related to standardization on data spaces within the framework of IDSA
   - **e.** Participate in the standardization work in the standardization committee under IDSI
   - **f.** Knowledge sharing to specific sector target group: operational managers in the Danish water sector participating in the yearly event Water Days ultimo 2024 by DANVA.

Stakeholders and collaboration partners

The activity plan particularly involves FORCE Technology's and Alexandra Institute's competencies within IoT, software development, cloud computing, data pipeline, data analytics, data visualization and data spaces. It will be carried out in close collaboration between FORCE Technology and Alexandra Institute. Alexandra Institute and FORCE Technology have also already been collaborating on data spaces through other means.

Synergies/collaboration with other projects

As part of the activity plan, coordination is ensured with the following other efforts and projects, so that knowledge and services developed under these become available to the target group.

RK:

- "Fremtidens hybride testbeds", FORCE Technology: about the use and validation of data from tests and from digital twins

- “Metrologi for digitalisering and datasikkerhed”, DFM (lead) and FORCE Technology: The activity contributes knowledge about metrology within data validation of IoT data in particular.
- “Digitale vandløsninger til den grønne omstilling”, DHI (lead) and FORCE Technology: The activity contributes knowledge about IoT environmental sensors and IoT systems in harsh environments as well as knowledge on data spaces since work on this also has been carried out in this previously.
- “Digital sikkerhed, tillid og dataetik”, Alexandra Institute (Lead) and FORCE Technology: Joint interest to find anomalies in datasets, more specifically for cybersecurity.
- “Digitale teknologier til datadrevet, bæredygtig vækst”, Alexandra Institute (Lead) and FORCE Technology: Joint interests to improve the processes around data collection and transformation including application of data spaces.
- “IoT-drevet forretningsdesign”, FORCE Technology: Two other activities in the same performance contract as this activity. Collaboration on the AI aspect will take place across these three for knowledge sharing and development. The focus area across these is “AI for Healthy Smart Cities” from the EDIH of the same name. The activities are: FT07.03 FT07.08 and this (FT07.05).

Other FoU projects:

- Data validation project with Energy Cluster Denmark.
- EDIH project AI-Boost, which is a large 3-year collaboration project in the Capital Region and Zealand Region, aims to increase the quality of life for people in Greater Copenhagen and on Zealand by pushing out advanced digital technologies in the Danish SMEs. The focus is on creating an increased degree of digitization for companies working in the fields of life science, health tech and the built environment - "Healthy Living". The project has connections to other performance contract areas and is expected to contribute in particular with numerous knowledge dissemination activities, including workshops, webinars, test cases, etc.
- Industriens Fond project focusing on Data Space awareness among the SME segment. The project is carried out in cooperation with Alexandra Institute and the goal is to launch “Danish Data Space Hub”.

Advisory Board

The activity plan has been shared by email with the Advisory Board on November 28, 2023.

Knowledge dissemination

Results developed under the activity plan are disseminated via the Nordic IoT Centre (nordiciot.dk) and associated stakeholders. The specific activities for knowledge dissemination are described in FT07.09_2023 Vidensprednin og økosystem.