Water Data Space

Note on governance for Water Data Space

The content of this note is developed by
Chiara Farné Fratini & Peter Steen Mikkelsen, DTU Sustain, Technical University of Denmark
Henrik Dissing, The Danish Environmental Portal
Ulla Sparre, Water Valley Denmark

June 25th , 2023

DEN EUROPÆISKE UNION
Den Europæiske Socialfond

DEN EUROPÆISKE UNION
Den Europæiske Fond for Regionaludvikling

Finansieret som et led i EU's reaktion på COVID-19-pandemien

Vi investerer i din fremtid

What will you find in this note on governance?

- Overall: this slidedeck aims at contributing to operationalization of the concept of Data Spaces in the water sector with a particular focus on governance
- Starting point for a data space: who takes the initiative, who designs the data space and who ensures long-term financing of operational costs?

- Setting the Scene – What do we need to know to define governance?
- Framing a Water Data Space – What is it? What are the characteristiques? What is the "new thing" in a nutshell? How to provide a systematic overview of roles in a data space?
- Which types of data-sharing and data flows can be foreseen?
- Governance elements including governance of existing Data Spaces for common public environmental data in Denmark
- Suggestions for further reading

Content

- Section 1 – Getting started: what kickstarts a dataspace?
- Section 2 – Definitions: comprehending a Data Space.
- Section 3 – Governance: what is encompassed and why is this the difficult part?
- Section 4 – Conclusions in short

Section 1

- Getting started: what kickstarts a dataspace?

  - Partners agree on opportunities and challenges to co-operate about, and where data sharing is required;
    consequently, partners agree on financing the proces, design and operations of the data space

  - Partners agree on governance details;

Starting Point – what kickstarts a data space?

Who takes the initiative?

- A number of stakeholders agrees on a Value Proposition for a new Data Space
- These stakeholders will form the first constituency base for the Data Space and decided on key principles including interoperability with other data spaces and data usage policies

Who designs the data space?

- The stakeholders agrees on main purpose and principles as well as scope for the data space
- Either by themselves or by involving consultants, the organizational including governance and operational including technical design is developed, tested and launched

Who ensures long-term financing of operational costs?

- A Data Spaces comes with costs for the the design phase as well as (and especially) the operational phase, how will these costs be covered?
- A sustainable business model is an important part of designing and implementind a data space

Data Spaces – what do we need to know to define governance?

- What is an agreed definition of a water data space? Who will take part? Which type of stakeholders? Chategorizing roles, show illustrative examples.

- What is the overall vision and value creation of data sharing? Who are
- possible beneficiaries?

- What type of data-sharing will be encompassed? Can we identify types of data which can be placed in groups with common characteristiques? – eg. commercial versus altruistic

- Which types of roles will be part of a data space? – and which of these are new and thus impacting the organizational structure, and adjoined mandates in decision-making and financing of new operational tasks?

➢ Altogether following the key question: "Why should we set up a data space"?

Checklist when setting up a data space – cf DSSC Starter Kit

The EU Data Spaces Support Centre (DSSC) **the Data Spaces Blueprint (v0.5)**

Success criteria for a Water Data Space

- Research and innovation stakeholders express that it has become much easier for them to find and access data, as well as new partners;

- Utilities express that they have improved their performance due to joint
- analysis based on smooth access to more data from different sources

- Service providers express that it has become much easier to set up new business based on the data space;

- Service providers express that their products have improved due to more and better data sources;

- Authorities express that they can base water management on more advanced tools (eg AI) due to more and better and more timely data available;

Section 2

- Definitions: what are basics of the technical framework?
- Examples of data sharing.
  - Currently, in Denmark within water data only one, altruistic, close-to-being data space exists (in this section, examples of comprehensive data-sharing are shown, but fully-fledged data spaces meeting all requirements * are still be to developed

* According to definitions by e.g. The International Data Space Association IDSA

Overall vision and value creation

Overall vision: Seamless sharing of data involving an increased number of data owners and data users leading to more and better data as a means for a digital transformation of the water sector, hereby enabling a digitally-based approach to addressing water challenges.

Value Creation: Increased efficiency; increased precision; enabling use of digital twins; enabling AI application in the planning and decision-making processes; improving search functions for data and partner organisations; new and better services; new business models; allowing for increased transparency and involvement of the public.

The purpose and value creation of the project Water Data Space:

Value creation is the main target for the Water Data Space. Through the data space companies, utilities and knowledge institutions will develop solutions that enable efficiency, innovation and sustainability.The output is measured in new business models, new sustainable solutions/products and optimizations.

The political context

- Data Spaces are considered as key element of an EU Data Economy but are also having a solid geo-political starting point: the ambition to create European independence of large, US and CN commercial players
- Cornerstone of EU Data Policy #1: data ownership are in the hands of the citizens (US: companies own the data; CN: the state own the data)
- Cornerstone of EU Data Policy #2: all publicly financed data must be FAIR – Findable, Accessible, Interoperable and Reuseable, this encompasses also utilities and research (EU Data Act in prep, negotiations expected to be concluded 2023)
- Existing legal framework: PSI Directive; INSPIRE Directive

For more detailed information please read the Policy Analysis from the project (Chiara Fratini & Peter Steen Mikkelsen DTU, 2023)

Data Spaces – overall, generic definitions

- The International Data Spaces Association (IDSA): "*A data space is defined as a decentralized infrastructure for trustworthy data sharing and exchange in data ecosystems based on commonly agreed principles.*"

- Nordic IoT Centre: "*Data Spaces are collections of data sources, including definitions of the data, shared between multiple parties, typically publicly available, and with a description of the terms of use including possible financial payments.*"

A Data Space illustrated – other illustrations can be found

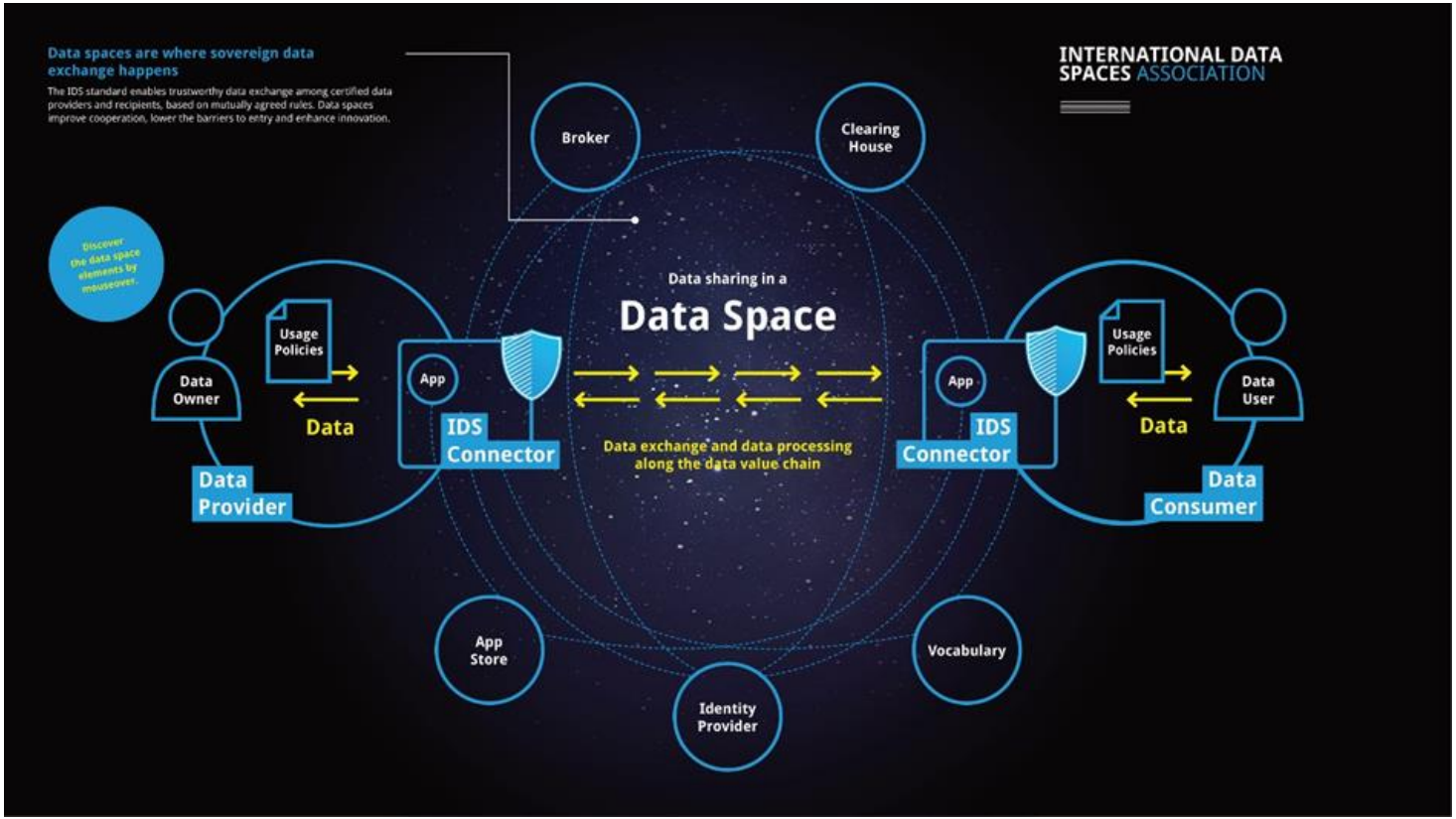


Illustration by International Data Space Association

Data Spaces in a nutshell – what is new?

- More Data for More Usages
  - Data Management is carried out in a way, which allow for data sharing in a broader context than the original purpose of each data sampling. A data space entails a “level up” compared to a platform, which serves as a data repository.
- Harmonizing Data Management to smoothen Data Sharing
  - Stakeholders are expected to contribute to a larger common good by subscribing to governance principles, standards, classifications, API formulas etc which allows for smooth data sharing based on FAIR principles
- New spectrum of roles in data sharing ecosystems
  - Data Owners and Data Users co-operation based on agreements of data quality and data usage. Technical intermediaries. Governance Intermediaries, maybe the same as the Technical Intermediary. Overall authorities ensuring applicability among different data sharing ecosystems.

Towards a Water Data Space #1
Which types of data flows?

Water Data are characterized by:

- They are key elements of the broader spectrum of environmental data, which in turn are pivotal for a Digitalized Green Transition (Transformation)
- Water Data relates to three domaines: water in aquatic ecosystems (with most data produced by public monitoring), water in pipes (utilities, industries), water in cities (often related to cloudbursts).
- Water (thus also Water Data) are often related to nexuses: Water & Energy, Water & Biodiversity, Water & Urban Planning, Water & Food
- Water Data can be found in "closed circuits" eg co-operation among utilities, but they are often connected publicly accessible data from one or several of the three domains

➢ Water flows across several domains and administrative structures, thus data sharing may involve a large number of data sharing ecosystems and a huge complexity of stakeholders

Towards a Water Data Space #2

Typical data imaginaries – view on data identified through the project Water Data Space

Altruistic data space

- Medium focus on data
- Data should be free and accessible to all
- Actively engage in providing data
- We are willing to compromise on our own needs for the sake of the common good

Utilities

- Emerging focus on data
- Data is increasingly viewed as business assett
- We are willing to provide data for the right fee
- We focus on our own needs but are willing to engage in focused cooperation with others

Commercial companies

- Strong focus on data as digitalization and technologies based on data in essentiel parts of R&D and service offerings
- Raw data is a key corporate asset – we refine our products and services based on raw data
- We are in generally not willing to share our own data unless we see commercial upside
- Many has a business model build on own data formats, so motivation for standardized data must be clear
- High interest in participating in partnership innovation with water utilities and research

Towards a Water Data Space #3
Regulatory framework

Water is a thoroughly regulated ressource with adjoined monitoring and reporting requirements at different levels:

- EU Directives, reporting from Member States to EEA (Data sharing is once per year)
- National regulation, national monitoring of water, in many cases by a national intermediary organization for data compilation, and in most cases coordinated with regional/municipal regulation, monitoring and data compilation (Continous data flows)
- Water use and monitoring in closed circuits (Commercially in companies eg Food Processing; Publicly or semi-publicly in utilities), with data subject to various degrees of sensitivity (Continous data flows)
- Water use and monitoring at household-level, subject to GDPR regulation
- Water in "projects" with a limited number of participants, time and scope of the topic
  e.g. Research and Innovation Projects, Infrastructure projects and related EIAs (Data flows among partners, data flows to external stakeholders)

➢ As water flows may involve several domains, water data should be enabled to flow similarly – this will in turn require a multifunctional intermediary ensuring technical as well as governance interoperability

Towards a Water Data Space #4
Defining arch types and scenarios for governance design

Outlining principles for Water Data Sharing within one European Water Data Space:

- All relevant data should be enabled to reach EEA from the Member States
- All relevant data should be enabled to reach National Environmental Data Organizations
- Water Data from "projects" (defined as activities with a limited number of participants, time and scope of the topic e.g. Research and Innovation Projects, Infrastructure projects and related EIAs) should be managed internally in the project in a smooth and efficient manner, while at the same type being enabled for all relevant data to transferred to the National Environmental Data Organizations
- Water Data from "closed circuits" and "households" enabled to be transferred to the National Environmental Data Organizations, while taking into account the specific "sensitivity" of the organizations involved

These principles for water data sharing constitute the starting point for implementation of one, interdependent European Water Data Space.

"Enabled" is here understood as the conditions for technical interoperability and governance providing the framework for all involved parties

Scenarios for a Danish Water Data Space

Through the work with involvement of 200+ water professionals, workshops, interviews with SMEs and stakeholders around a Danish Water Data Space, we have identified 3 arch types of governance models

Please find the arch types described in the tabel on the following page

The governance model and the design and implementation of a governance structure will be subject to further project work between the partners in the Water Data Space project

Data Spaces Key Elements / Archtypes	Archtype A Independent Water Data Space Technical co-operation	Archtype B Public, Altruistic Water Data Space	Archtype C Independent Water Data Space Commercial activities
Value Creation	Utilities-initiated, focus on increasing efficiency Research and Innovation projects Environmental Impact Assessment	Environment Quality Management or ressources management eg groundwater. Can be national, international eg EEA or GSEU, or River Commission-initiated within or between River Commission, improving river water quality	Data Marketplace on commercial terms
Organisation	Co-operation anchored in a Project Organization, with agreements about purpose, scope, time horizon, obligations reg data sharing and data use Characterized by being limited in time, scope and participants. Can eventually also be ongoing eg a co-operation between utilities Purpose is a technical result	Administrated by an Authority, or an organization established with this specific purpose Characterized by being ongoing, with the purpose anchored in public, administrative obligations	Commercial Company Characterized by being ongoing, purpose is commercial
Ownership	Co-owned, can be a public-private partnership with all data owners represented, eventually also data users. Management responsibilities anchored within one of the participants or in a separate body	Publicly-owned Water Data Space, financed from Government / Authorities Budget Altruistic; data can be found and accessed for free Data Providers are obliged to deliver data, or can be based on voluntary agreements	Partners investing in the company
Constituency	Agreements on decision-making etc laid down in Terms of Reference describing participation, obligations for participants, data ownership etc	Agreements on decision-making, financing of operational costs including	Commercial terms, decisions on investments, owenership, membership fees, procurement and selling of data etc
Financing/business model	Financing is member-based, covering all costs. At the outset, the co-operation will not have income from sales of data, but members may benefit from using the results in improving their own performance, thus also economic situation.	Public financing of operational costs. Public / project partner financing of innovation costs.	Partners provide basic financing and initial investments. Eventually, the company will be financed solely from income from sales of data.
Ethics, data sovereignty	Follow offical principles on dataethics; GDPR legislation		
IT security	Consider safety and security measures; critical infrastructure		
Others	SLAs, Data Sharing agreements etc		

Towards a Water Data Space
#5Technical coherence

- A universe / community of heterogenous data spaces, committed to work along same lines / approaches of data models and data infrastructure
- Using the same Embedded Tools – open source – same standards, same classification lists, same APIs – allowing for smooth data sharing
- All data sets according to the same content: location (geographical position)-observation (method)-measurement
- Subscribing to the same governance framework, elaborated by the community and endorsed by …. ?

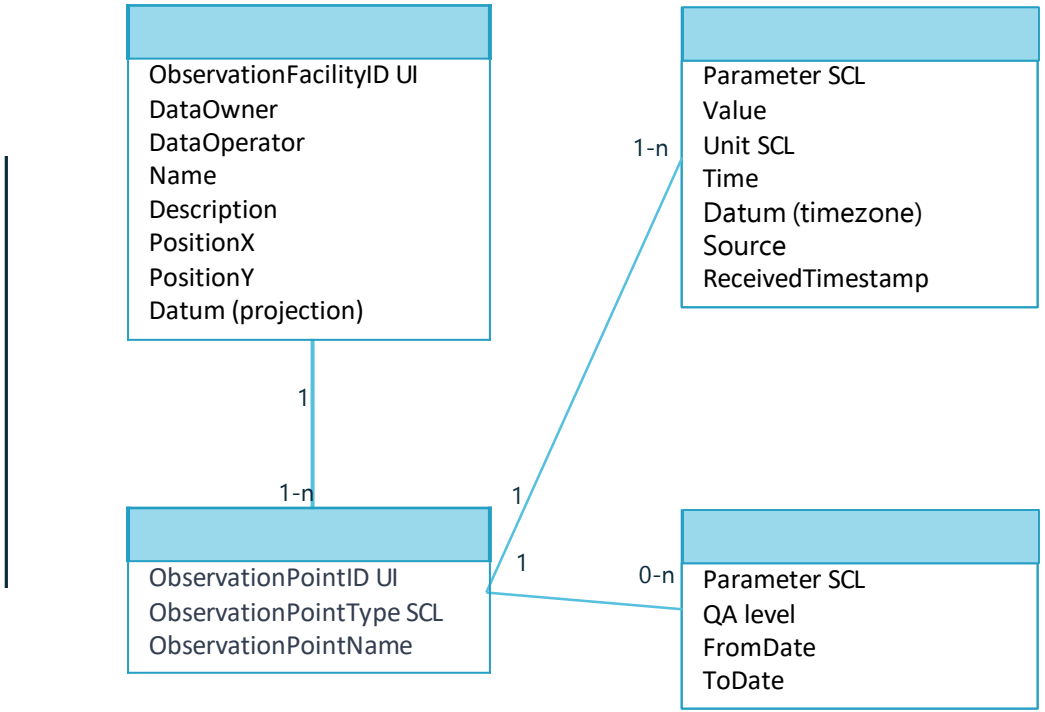
Technical elements for interoperability

- MVP: Must include the full data handling process from data generating components at sample points and related connecting to metadata, via data-in APIs, data storage, data-out APIs and exhibiting of accessible data, search function.

- Processing of data and any type of value-adding, calculations etc. takes place outside of the data sharing ecosystem, but in respect to data sharing agreements

- Result: at the basis of a joint datamodel, a data sharing ecosystem is developed.

A common data model



Takeaways

- QA entity is voluntary
- UI=Unique Identifier
- SCL Stancodelist

Examples of Data Sharing #1 – case Groundwater, Denmark

- National Groundwater Monitoring Program – Measurements of groundwater levels and quality are transferred to the Danish Environmental Portal and from here presented and made available for the public – continously and ad hoc uploads
- Measurements of groundwater levels and quality are carried out in connection with e.g.
  - Construction works and infrastructure projects including EIAs include groundwater analysis, pdf's and/or data can be transferred to the Danish Environmental Portal – ad hoc
  - Research and innovation projects include groundwater analysis, pdf's and/or data can be transferred to the Danish Environmental Portal – ad hoc
  - Private wells extract groundwater as a ressource for use in Farming or Food-processing industries (Breweries, Dairies etc), pdf's and/or data can be transferred to the Danish Environmental Portal – ad hoc
- Data on Groundwater Levels are extracted from the Danish Environmental Portal and used for e.g.:
  - Analysis of impact on aquatic ecosystems
  - Commercial Service Provider carry out for Municipalities on Impact Assessments of applications for establishing new wells
  - Analysis of groundwater levels close to terrain in connection with Real Estate Investments
- Aggregated data on Groundwater Resources in Denmark are transferred to EEA

Examples of Data Sharing #2 – Research and Innovation Projects

- Research and Innovation projects (could also be EIA projects)
- Limited in time, participants and scope
- Data collection and processing take place during the project's lifespan, data may not be ready for sharing until the end of the project
- Some data may be generated by the project itself (eg. Field studies), thus some raw data may exist
- Data usage from other sources including publicly accessible data as well as from various research and innovation organizations is very likely
- Both raw data and processed data originating from the project could potentially be shared
- At the outset, Research and Innovation Projects funded by public sources are obliged to be shared

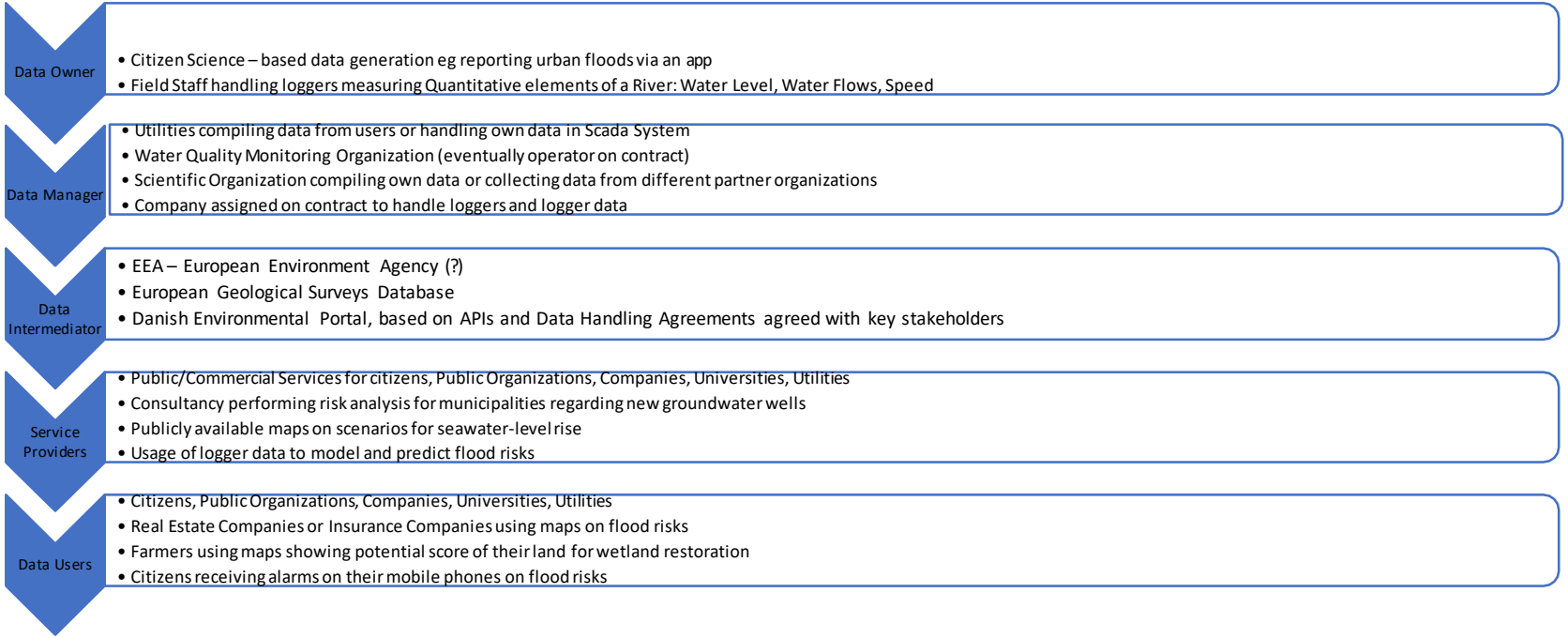
Examples of Data Sharing#3
Almost being a Data Space…
Common public environmental data in Denmark

- Danish Environmental Portal (DMP) (as a service provider) offers costumized exhibition of a number of data-out services e.g.:
  - Climate Change Adaptation (KAMP) https://kamp.klimatilpasning.dk/
  - Exhibit of raw data (Miljødata) from monitoring of environmental quality of lakes, rivers and marine waters Miljødata (miljoeportal.dk)
  - Full overview of the data-out services at Danmarks Miljøportal IT Systemer (miljoeportal.dk)
- Further, DMP (as intermediary: technical management, various co-operation agreements for data-in and data-out etc.) organizes data-in specifically for the following:
  - Surface Water
  - Wastewater
  - Groundwater
  - Soil
  - Nature
  - Environmental Impact Assessments
  - Rats
  - Circular Economy (in prep)
- Further information about DMP: Danmarks Miljøportal Om Danmarks Miljøportal (miljoeportal.dk)

Examples of Data Sharing #4

Roles in a Water Data Space *
Generic types + an example of a "local" Water Data Space in DK

- Water Consumption Users – Water Meters
- Water Quality Monitoring; Measuring Quantitative elements of a River: Water Level, Water Flows, Speed



* Based on Data Spaces Definition developed by Alexandra Institute (DK)

Section 3

- Governance: what is encompassed and why is this the difficult part?

- Which are the challenges to be in focus? Big difference in data required between adressing short-term problem solving (2025); medium-term optimization (2030) ; and long-term transformative changes adressing planetary boundaries (2050); the longer time perspective, the higher complexity, cross-sector solutions as well as uncertainties

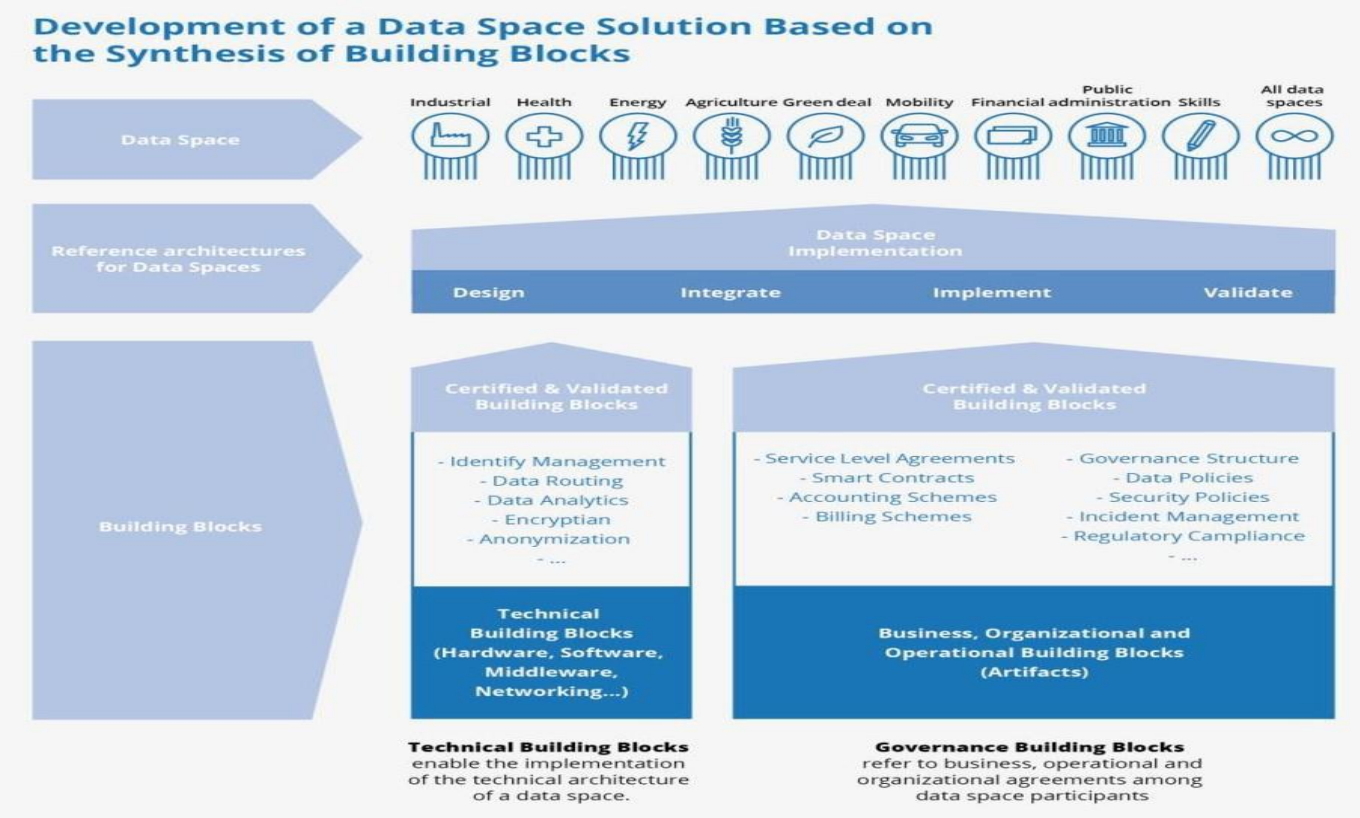
- Who owns the data?
- How to ensure harmonization of data, if the starting point for data sharing is a plentitude of data systems?
- How to ensure leadership attention?
- Who pays for design and operations e.g. support functions?
- Who can access the data and on which conditions?

How to develop and manage a Data Space?
What will be the organisational aspect of Governance?

https://dssc.eu/

- Different visualizations gives different understandings of attention to be paid to technical vs governance cf next slide

How to develop and manage a Data Space #2



Elements within governance * **

ORGANIZATIONAL
– Defining roles, mandatas, responsibilities

OPERATIONAL
– Defining technical framework

BUSINESS
– Defining commercial aspects

➢ Will be presented in detail in the next slides

* Based on a.o. IDSA, Alexandra-Institute and Nordic IoT Center, see *Further Readings* at the last slide

** Some papers use five levels for this: Business, Legal, Operational, Functional, Technical

Governance reg Organizational Elements

- Who initiates a Data Space, and with which purpose? To operate as a closed circuit only, or interoperable with other data spaces? Altruistic or commercial? Open for all or only via fee-based membership?
- Who will constitute the Constituency of a Data Space? Who will have a say on constitutional documents?
- Organizational Design of the data base, especially the intermediary functions (Technical, Governance, Business-wise eg payments)
- Mandate to an Intermediary to handle governance, issued by a Constituency based on an overarching co-operation agreement
- Financing of Intermediaries – establishing and operating the data space comes with costs of an intermediary, how to finance this
- Data sovereignty and Data Responsbility Agreements - Responsibility for the Data Quality lies solely with the Data Owner (defined as the one paying for the measurements, i.e. also if a data manager is contracted), but the Data Owner is only obliged to label the level of quality control, not to perform any steps in this regard

Governance reg Operational Elements

- A number of technical elements needs to be defined and agreed among the participants of a data space in order to ensure interoperability and to define obligations of the technical intermediary eg monitoring and logging of data-to-day exchanges.
- The MVP technical infrastructure is shown at slide #25, in this slide the rules and agreements concerning the technical elements are presented.

- Service Level Agreements: Monitoring of data exchanges, monitoring of software infrastructure, ensuring back-up, setting up a support service for users etc.
- Parameters used in measurements must all be based on same lists eg chemical substances or biodiversity species lists - Standards and Classifications to ensure applicability of data
- Data-In to the Intermediary as well as Data-Out based on agreed APIs
- Ad Hoc Data Transfers to the Intermediary based on Data Agreements with D
- Continuity model
- Vocabulary

Governance reg Business Elements

- When adding commercial elements to a data space, the complexity of governance increases

- Co-operation between utilities may represent a special type of “business” as data from water supply and wastewater treatment represents an economic value

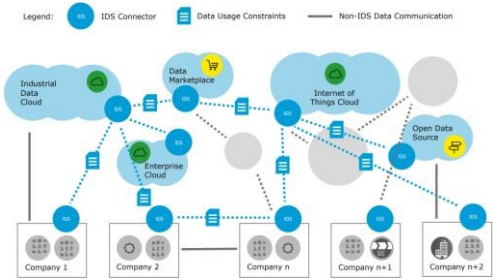
- Business Models – Data for free or for procurement? – Services from Commercial Providers based on free, public data but with payments for the value-adding

- Business Models of Component Producers offering data-based services, how do they fit with data sharing of a data spaces

- Smart Contracts – Machinereadble contracts defining usage policies, legal aspects, SLAs etc. - Who formulates templates? – Who offers mediation between data owners and data users? - Support structures and Legal Advice

- EU Data Act requires all data to be FAIR, what will this imply for researchers?

- Who handles payments, if relevant?



To conclude

- The concept of Data Spaces allows for seamless data sharing, enabling more data for more users
- Moving towards Data Spaces requires substantial efforts reg definitions, roles and agreements – AND needs financing
- As concerns water, altruistic data stemming from public monitoring will in most case be the turning point
- Technically, joint data models including choise of standards and classifications are required, but barriers exists
- Governance constitute the biggest difficulty, involving e.g. decisions on data sovereignty and financing
- Still today, the first fully-fledged data space in the water sector are still to be seen. For this to happen, a challenge with a strong upside potential, requiring multi-stakeholder data-sharing and a solid model for financing has to be identified = a strong answer to WHY?
- 3 arch types of governance based on the value driver has to be taken into consideration when designing and implementing the data space governance

Further reading

- International Data Spaces Association, Position paper, "Design principles for data spaces", April 2021, 1. Gather Knowledge - How to Build Data Spaces? (internationaldataspaces.org)
- International Data Spaces, position papers, Position papers - International Data Spaces
- Data Spaces Support Centre: "Starter Kit for Data Spaces Designers", March 2023, DSSC-Starterkit-Version-1.0.pdf
- Alexandra Instituttet: "Data Spaces: Kortlægning af aktører og initiativer", January 2023, Data spaces - Alexandra Instituttet
- Nordic IoT Centre: "Data Spaces in Denmark", Data spaces in Denmark – Nordic IoT Centre