



It all starts with semantics

A talk with Wout Hofman, driving force behind the FEDeRATED reference architecture ¹

Wout Hofman is senior scientist Data Ecosystems at TNO, the Dutch institute for applied research. Trained as an electro-technical engineer in the eighties, Wout obtained his PhD in IT in 1994, after 4 years of developing a conceptual model for business transaction management, also known as ‘control tower’. This modelling experience never ceased to occupy his mind, neither as a consultant to a multitude of clients, nor as a researcher. One of the first words Wout had to learn to pronounce properly when joining TNO was ‘interoperability’ –the word that captures the essence of his professional mission ever since. From 2015 onwards, Wout established himself as one of the chief designers of the DTLF federated network of platforms approach, as well as the Dutch Basic Data Infrastructure (BDI). Both constitute the foundation for the reference architecture of FEDeRATED. Besides that, he is engaged in multiple advisory roles, at home and abroad.

Wout works 3 days a week, because he is winding down his career. He has been active in this field –building infrastructures for data sharing in specific markets and communities– since 1985, so close to 40 years. Wout is from the generation of Tim Berners-Lee, the inventor of the World Wide Web.

Just like Berners-Lee invented the Internet for global purposes, Wout – in collaboration with the FEDeRATED IT Architecture Board and the EU DTLF – is inventing a ‘global data infrastructure’ . Formulated differently: creating a *single common* dataspace, ‘open and neutral’ and safeguarding ‘data sovereignty’, for every actor within the larger freight transport and logistics community. This goes beyond the divide between public and private sector. Wout works from the perspective of the communal and global.



FEDeRATED envisages digital integration for seamless multimodal freight transport. Wout: “Interoperability of transport modes comes with supply chain visibility, which creates the opportunity to optimize the use of assets and infrastructure.” To which he adds: “Besides, it also comes with all kinds of opportunities for third parties to develop new services. That is what we need to move towards seamless supply chains.” The DTLF network of platforms approach should eventually result in a digital grid that aims to enable any participant in the logistic chain to act as a decentralized node sharing data at source in a secure way with other participants.

¹ Interview by Minne Buwalda

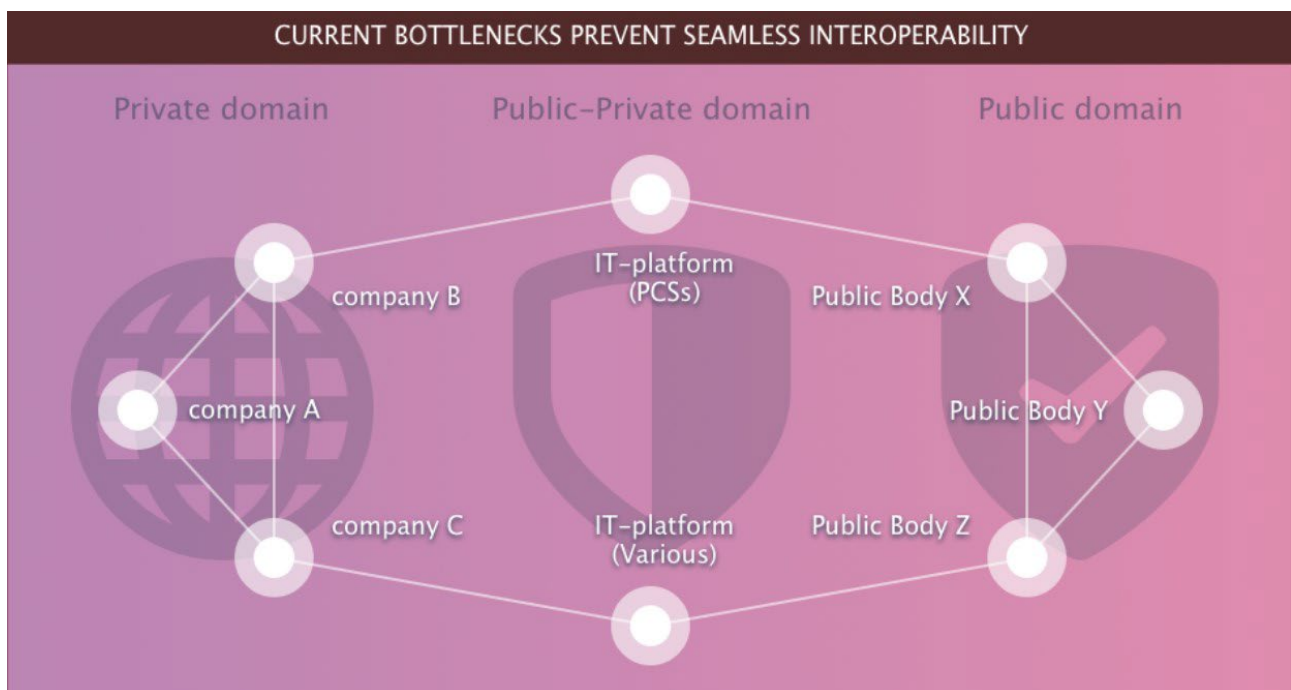




Working at a meta-level

The complexity of Wout's job has to do with the fact that such a single, common dataspace has to be developed at a meta-level. Working at this abstract level makes it difficult to communicate, but Wout gives it a try: "In the case of FEDeRATED we are developing a blueprint for a digital infrastructure that is configurable by way of semantic models." Wout strongly advocates interoperability through semantic web technology, and not in the traditional way, for instance based on message standards only.

Pointing at the present situation and the digital future he is aiming at, Wout explains: "A dataspace is a system of agreements between companies and organizations on how they want to share their data. These agreements may concern standards, for instance; or in the case of FEDeRATED, semantics." He continues: "Currently, a lot of organizations create their own dataspace and link other companies and organizations to it. Think about supermarket chains that build a dataspace in order to exchange data with their suppliers. Or think about ports, that all have different agreements, different PCSs, different dataspace."



He continues: "But the present fragmentation of dataspace –all these silo's– is not the ideal solution, and I see it as nothing but an intermediate stage, because ultimately it is about dataspace interoperability." And: "Our main focus is to support individual organizations in partaking in a single, common dataspace for supply and logistics."

Digital twins and events

Getting a grasp on Wout's grand design cannot be done without knowing some of the basic concepts he and the FEDeRATED team put forward. To start with, there are crucial concepts like 'digital twin' and 'event'.





A 'digital twin' is a digital representation of the physical infrastructure and all possible transport objects concerned. Wout starts with the basics: "It is essential to have a digital twin that can be used by all for data sharing." Then he starts expressing himself in more technical terms: "A digital twin is a taxonomy, a hierarchical structure of objects from the real world of transport. Part of this taxonomy is equipment: a trailer, a shipping container, a railway carriage... You can also include terminals, cranes, straddle carriers, and then further expand the taxonomy, even into the field of production equipment in industry." Wout, enthusiastic: "That is a very powerful idea."



But of course, in the end it is not about such real-world objects per se; it is about the movements of or between such objects; it is about the *activities* –or 'events'– that take place on that physical infrastructure. Therefore, apart from the digital twin of the physical infrastructure, an inventory of all possible business activities needs to be created. In Wout's IT-vocabulary such activities are categorized as 'subtype of event'. Wout: "By categorizing events, we categorize what companies do." And: "This way we specify the relationship of data to be shared between two organizations."



It all starts with semantics

So according to Wout, making digital technology accessible for all stakeholders within the world of European freight transport and logistics starts with the realization of semantic interoperability. Stakeholders have to be able to understand one another and make themselves heard. Preferably this should not lead to administrative burdens and costs.

In his design of a global dataspace, Wout therefore emphasizes semantics. He says: "Semantics contain everything for data sharing." And: "The FEDeRATED reference Architecture differs from other corporate and government initiatives in this area because of our focus on semantics, which is the core of our system." To which he adds: "Another focus is on 'data at source', which is not that complicated, even though it does mean a change for many organizations."

So, the FEDeRATED network of platforms approach is preferably based on semantic web technology, but that does not mean Wout designs a new, overarching multimodal *semantic standard* –which could be likened to some kind of Esperanto for European logistics. Instead, he designs a multimodal *framework for the alignment of existing standards by way of semantics*. In short, Wout





does not design a new language; he designs “a mechanism by which relevant stakeholders can design and configure their data sharing solutions”.

According to Wout, interoperable semantics can be applied to all processes in business collaboration. “You start with the semantic model and relate it to the collaborative business transactions. On that basis you develop the technical specifications for the Service Registry, the IAA (Identification, Authentication, and Authorization), and the Index, considering various stakeholders and their roles.”

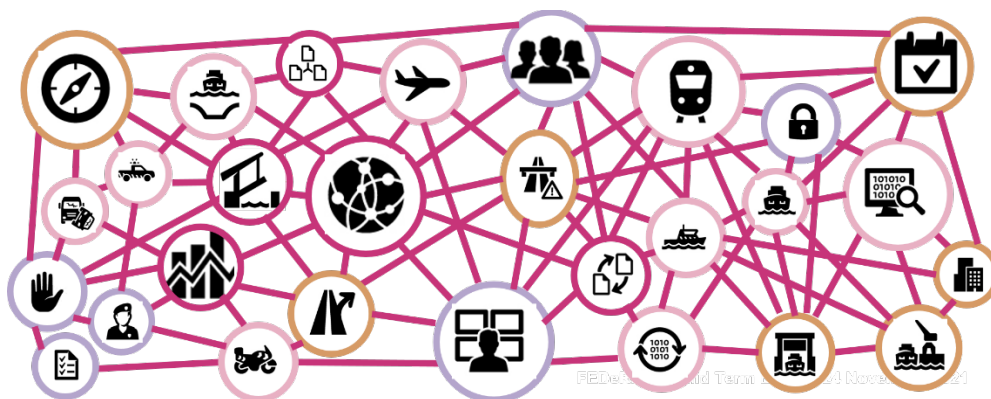
And: “The technical application called ‘Semantic Adapter’ assists in streamlining the data between these technical specifications and existing IT systems of organizations. It is through semantics –the language– that logistics processes –search and find, supply chain visibility– can be optimized.”

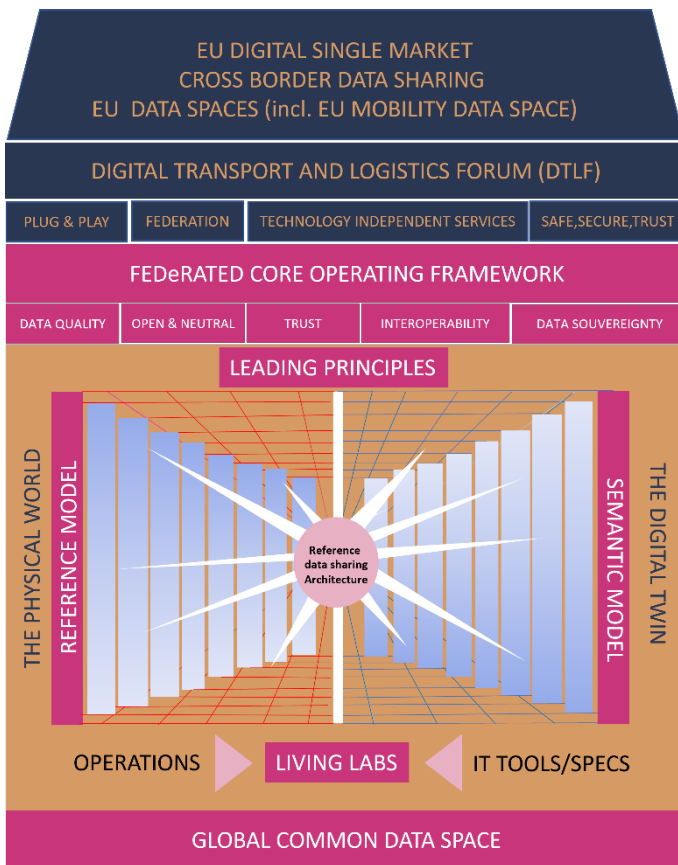
So, it all starts with the ‘semantic model’, which is a semantic representation for data sharing. Wout: “The semantic model gives us the concepts and associations organizations require to share data.” To which he adds: “It is also the semantic model that creates machine-readability and allows data sharing between different systems.” And: “In case of FEDeRATED, it models all types of interactions between two organizations, e.g. orders, bookings, visibility events, logistics activities, business documents such as eCMR, eAWB etc.”

Creating tools for the future

Before moving on to the FEDeRATED applications, Wout reiterates the importance of the FEDeRATED Architecture: “We have to start with removing bottlenecks like the creation of specific data spaces in the context of e.g. port communities and supply chains. Currently, many go too fast to generate technical solutions without taking a holistic approach, because they focus on the silo of their specific business.” To which he adds: “That is why interoperability still tends to fail.”

So, within FEDeRATED everything starts with applying interoperable semantics to all processes by way of the semantic model. This is a basis for the creation of innovative applications for a common, federated future.





Wout defines FEDeRATED applications as “the functional requirements of FEDeRATED translated into tools.” But what functionalities is he talking about? These are functionalities like discoverability, security, and access control.

And what tools is he talking about? Wout: “These are tools like Service Registry for discoverability; IAA for access control; and Index for data quality validation.” He continues: “These tools are articulated by technical specifications, and as soon as these technical specifications are implemented, the tools can be tested by users. Wout: “That is how we make them more precise.”

Illustration: The steps towards Global (common) Data Space, based on the EU DTLF -FEDeRATED network of platforms approach

Eventually, we need opensource tools to realize the objective of the EU and FEDeRATED Vision towards an open and neutral network. Wout: “Once the technical specifications are sufficiently detailed, complete, understandable, and implemented, they can be made available as tools for the wider EU freight transport and logistics community, including SMEs.” To which he adds: “They should eventually be made available as freeware/open source, and that process may be started via tenders.”

Musing, he says: “The technology is not that difficult. The concept and definition for applying the technology are difficult.” Then, wrapping it up: “But eventually we have to hide all that difficult stuff. With a car you don't want to know either how everything functions under the hood.”

So, when FEDeRATED Living Labs implement the technical specifications, they can act as a node in the FEDeRATED grid, and be early adapters of the opensource tools for Europe’s future transport and logistics system? Wout: “This is correct. We must provide guidance to Living Labs and early adopters to implement the architecture themselves or with the help of others. Eventually, they can act as a node in the overall grid. Management of the grid is completely distributed, but must function according to common agreements, like those formulated by legislation.”

Service Registry

When asked what the tool called Service Registry entails, Wout starts explaining: “The Service Registry contains the semantic model -plus the concepts that describe the relationship between





possible transactions, i.e. the process model.” He continues: “The Service Registry allows an organization to specify how they want to participate in the grid, i.e. what services they want, what data they need, and what data they can share.” He continues: “The Service Registry enables organizations to categorize what they do, what events they participate in, what business services they offer.”

And: “Companies and their services not only need to be findable; they also need to be able to agree on rates; make bookings; send tenders; then coordinate schedules. That’s all in the Service Registry.” He concludes: “It is about business services for business activities; it is about the consistency of all those interactions.” Formulated differently, the Service Registry of an organization supports its discoverability and its capability to do business electronically.

Asking Wout how companies arrange that at present, he says: “Companies arrange that amongst themselves, bilaterally or in communities, and then they start packaging those agreements in data carriers, e.g. an EDI or GS1 standard, or a proprietary open API, which is then implemented. But to get such a system fully working between two parties, you need an average of six to nine months, and that is undoable for a lot of SME’s. Besides, it does not provide the agility and resilience required in supply and logistics.” Wout, concluding: “Eventually, a FEDeRATED Service Registry enables instantaneous data sharing, or what is called ‘plug-and-play’.”

The latest Factsheet of FEDeRATED Living Lab #20, from September 2022, says ‘the Service Registry is under development’. Asking Wout about the current stage of development, he says: “We already developed a tool called ‘Semantic Treehouse’. With that tool, we can say: We put the semantic model in it, we define a transport order in the multimodal semantic model, and then we press a button, which generates an API we can implement. Companies can apply this tool plug-and-play”.

Index

Concerning the tool called Index, Wout says: “An Index of an organization contains all the events, with links to the data that have been received or issued. So, every stakeholder’s Index contains the events in which that stakeholder partook, plus links that were shared.”

Wout goes into a bit more detail: “In case of the FEDeRATED-grid we’re talking about distributed databases that need to be connected, but we don’t want to copy the data of one party to the other. So how do we proceed? We tell the other party: these are the links you have with my data, and these are the events the data are concerned with.” After checking if he is still understood, Wout continues: “So, everyone has their own data and a set of links to other databases. If we make those links available, then we also have to store which links we share, in order to check if it’s allowed and done the right way. So, how do we do that? For that we develop and use the Index.”

The Index has to secure data quality, so the correctness and completeness of data. This is preferably by each data holder. Besides, the ordered sequence of events, which is called ‘event logic’, should





be validated by way of the Index. Wout: “Once the data are converted from the input format to FEDeRATED semantic data, we have to ask the question: 'Does the data exchange fit the pattern we agreed on for sharing data?'.’” He concludes: “Through the Index we can set up all kinds of validation, and check the quality of the data.”



Identity, Authentication, Authorization (IAA)

Another technical specification for FEDeRATED deals with Identity, Authentication and Authorization, so access control. Yet, building a FEDeRATED tool for this is not an issue for the time being, mainly because first of all European and global regulation and supporting services or tools in this field are needed.

Wout: “Concerning IAA there are issuers, holders, and verifiers of credentials. We can work with established identity brokers –so issuers– but this is temporary, because we do not want central functionalities in the FEDeRATED set-up. Eventually there can be many issuers, but legislation in this field will have to indicate which issuers are trustworthy.” To which he adds: “Such a system grows much faster than by way of a central broker. In the US, they are already moving in that direction, and at present TNO and Dutch Customs are in discussions with them on this topic.”

Semantic Adapter

Apart from the semantic model, there is also talk about a Semantic Adapter recently. Asking Wout if he can describe this Semantic Adapter in a simple fashion, he says: “It is software for data transformations between different formats via the semantic model. The semantic model is at the core of the adapter. The adapter knows how the data input matches the semantic model, and then creates output based on that semantic model. And eventually the adapter translates the data back to its original format.”





And: “The Sematic Adapter enables everyone to use their own formats or standards and be able to share data with others that have different formats or standards. By way of the Semantic Adapter data is shared in the semantic format between the FEDeRATED stakeholders.”

Change management

Realizing a distributed data-based future in logistics requires all stakeholders to be adaptive to change. There are three types of stakeholders that benefit from data sharing in freight transport and logistics: businesses, platform and software providers, and governments. In moving towards a federated data-sharing approach, Wout assumes a proactive role of the government. This is because the public sector is the only stakeholder that can realize an open and neutral network. Such is not a task that businesses want to pursue –each company requires a competitive advantage over others. There is not one company that is able to develop such an open and neutral solution acceptable to others. Besides, companies generally just want to have their own data systems in order and compete in their market, and software & platform providers want to market their own software & platform. Yet, according to Wout, “companies also require a common, federated network for agility and resilience”.

The EC and quite some EU Member States governments opt for an open and neutral network for several reasons. One of those reasons is that they eventually want to be able to steer towards sustainability and circularity in supply and logistics. Wout: “They want to be able to see, for example, whether a law in these areas actually ensures a higher load factor and fewer kilometres. In 2019, the average load factor in the EU was 45%. This might be partly due to the European Cabotage Act. Cabotage is transportation of goods by road in an EU country on behalf of a carrier from another country. This is only allowed to a limited extent now.”

The need for commitment

Within the FEDeRATED context Wout aims at building a solution for the common good of all European stakeholders in freight transport and logistics, but most of these stakeholders experience his contribution as somewhat detached from their daily reality as a market player or regulator. Most of them already have a lot of trouble dealing with their short-term digital problems, and because Wout’s solution is so long term, they do not usually prioritize it. It may be a projection of the ideal situation, but for them that projection is too far in the future to worry about.

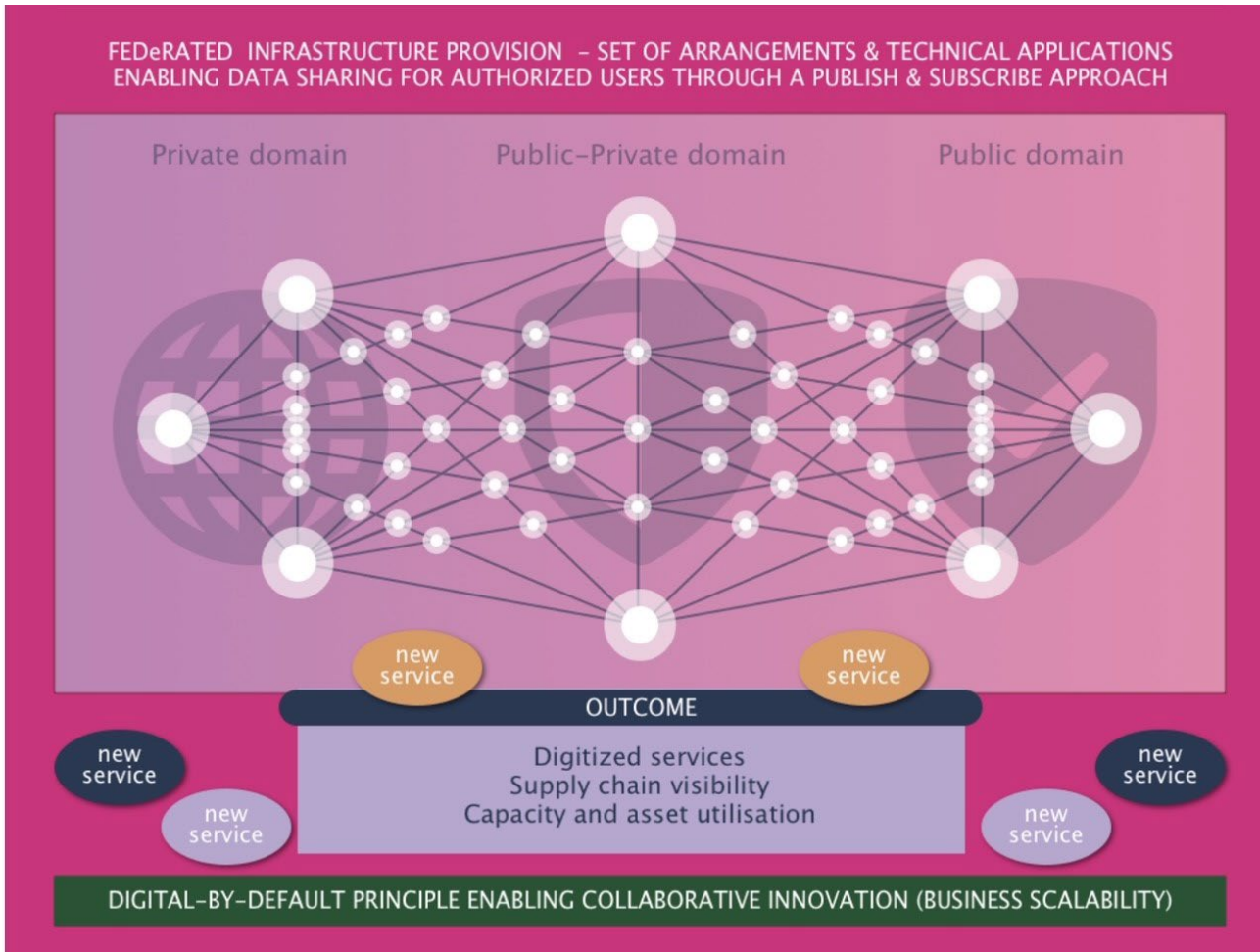
Yet, stakeholder involvement from within the market is needed to make the Global Common Data Space design become a reality. Wout: “Based on the semantic model, we specify and categorize what the input can be. But eventually companies must be able to provide that input.” Early adopters, in the sense of companies and government agencies that work as the first nodes in the envisioned grid, are crucial for the success of FEDeRATED’s endeavor to create a grid for us all. Furthermore, “a migration strategy and supporting implementation guides are also a prerequisite for success”, according to Wout.

It may be clear by now that building a federated network of platforms from a meta-level is not something that is done overnight. Wout: “It is a complex environment we operate in, where various





private and public organizations can have competing approaches in realizing this vision. Implementing such a concept takes at least five to 10 years.” By that time, Wout will be enjoying his pension. So, maybe his job is about starting the movement towards this distributed logistic grid of the future, and let the younger generation realize it, in their own way.



FEDeRATED's Vision towards an infrastructure provision = a global common data space for supply and logistics

