Linked data increases the catch qual

Linked data is about adding a semantic value to data gathered from different parts of the organization by using the Internet to connect related data that wasn't previously linked. "This is a way to actually get the catch ou were searchin for when fishin in the Data Lake", says [redacted] at Research & Development IT.

"Data that is connected, and not hidden in silos, enables us to understand relationships, creates meaning and helps us realize new business opportunities", says [redacted] "Data stored in knowledge graphs helps us understand information relations between data through machine reading. This is a way for companies to be less dependent on analysing and interpreting information in the future, but still make great sense of it". One of these methods is linked data, also called semantic data, since it creates meaning and value. The core is to publish structured data so that it can be interlinked and become more useful through semantic queries. Linked data is built upon standard web technologies, like RDF, but rather than using them to serve human readers, it extends them to share information in a way that can be automatically read by computers and make the basis for informed decisions based on knowledge and facts.

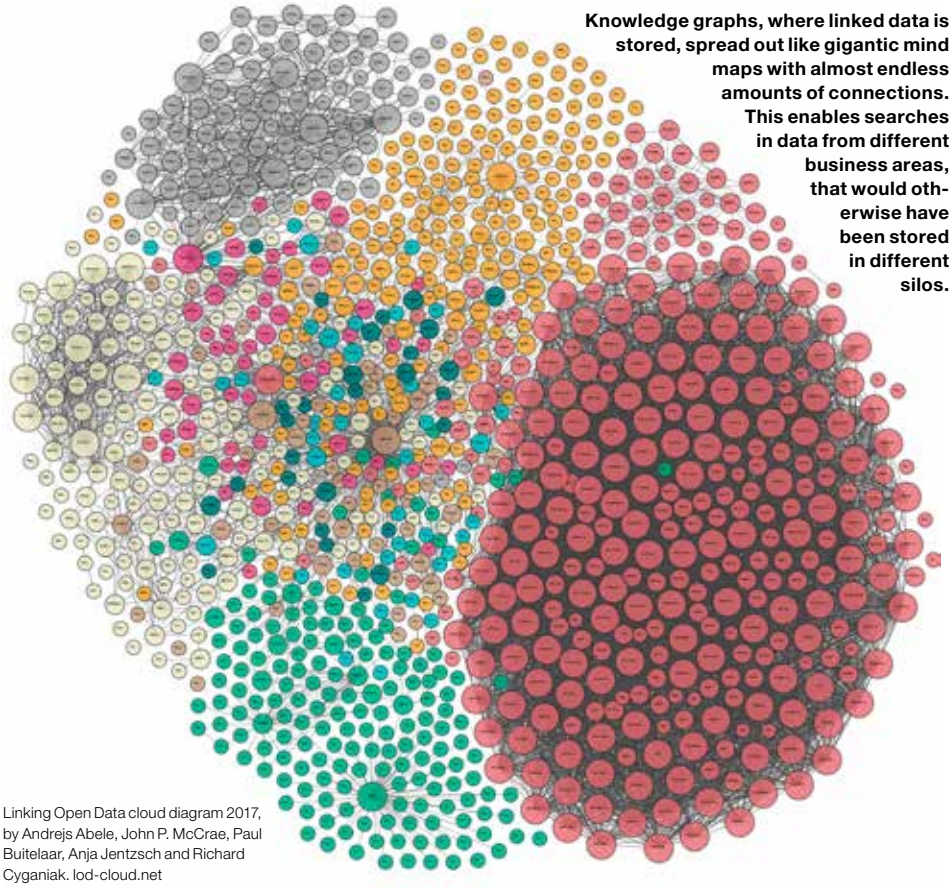
As part of developing the sustainable transport solutions, which are Scania's future, there is a need to keep the complexity level low and to increase the development.The Embedded systems area with electrification and connected vehicles have grown enormously during the last decade and thus also the complexity, especially thanks to Scania's modular development approach which covers more than 6500 product characteristics and thousands of billions possible variants. That is different combinations of vehicles with different features of cabin types, wheel configuration, power train, gearbox etc. These combinations are built on more than 1 000 000 different articles. Then consider that there in every device are multiple Electrical/Electronic (E/E) systems to provide various user functions for the trucks in many versions and configurations. For example in embedded system development, tasks might be to automatically adjust the temperature in the cabin depending on the outdoor temperature or adjusting the brightness of the instrument panel depending on the daylight.

The [redacted] project, having six development team members, is a part of the program Embedded tool chain, driven by [redacted] IT tools, together with [redacted] at Scania R&D. "We work cross-functionally to increase the catch quality in the Data Lake and to enable a better collaboration with VW Truck & Bus. Another target is to meet the ISO 26262 functional safety standard for electronic and electrical safety-related systems in the automotive industry, taking effect in 2018". The team works agile in close cooperation with business users and the [redacted] project, which has been running for several years, and is a collaboration between researchers at KTH and Scania. [redacted] has contributed with methodology, principles and prototypes for tools. The knowledge has been described in articles and doctoral theses and their advice now has direct impact on the new methods developed at Scania and the [redacted] Project.

"This is a new opportunity in the area of discovering and synthesizing insight from relationships among data."

Jawad Munir, [redacted]

"This is a new opportunity for Scania in the area of discovering and synthesizing insight from relationships among data", says Jawad Munir, [redacted] "Linked Data is not a new concept. It is being widely used on the World Wide Web to integrate data. Google, Facebook as well as healthcare, news industry and the Intelligence community already benefit from Linked data". Sometimes Linked Data



Knowledge graphs, where linked data is stored, spread out like gigantic mind maps with almost endless amounts of connections. This enables searches in data from different business areas, that would otherwise have been stored in different silos.

Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak. lod-cloud.net

ity in the Data Lake

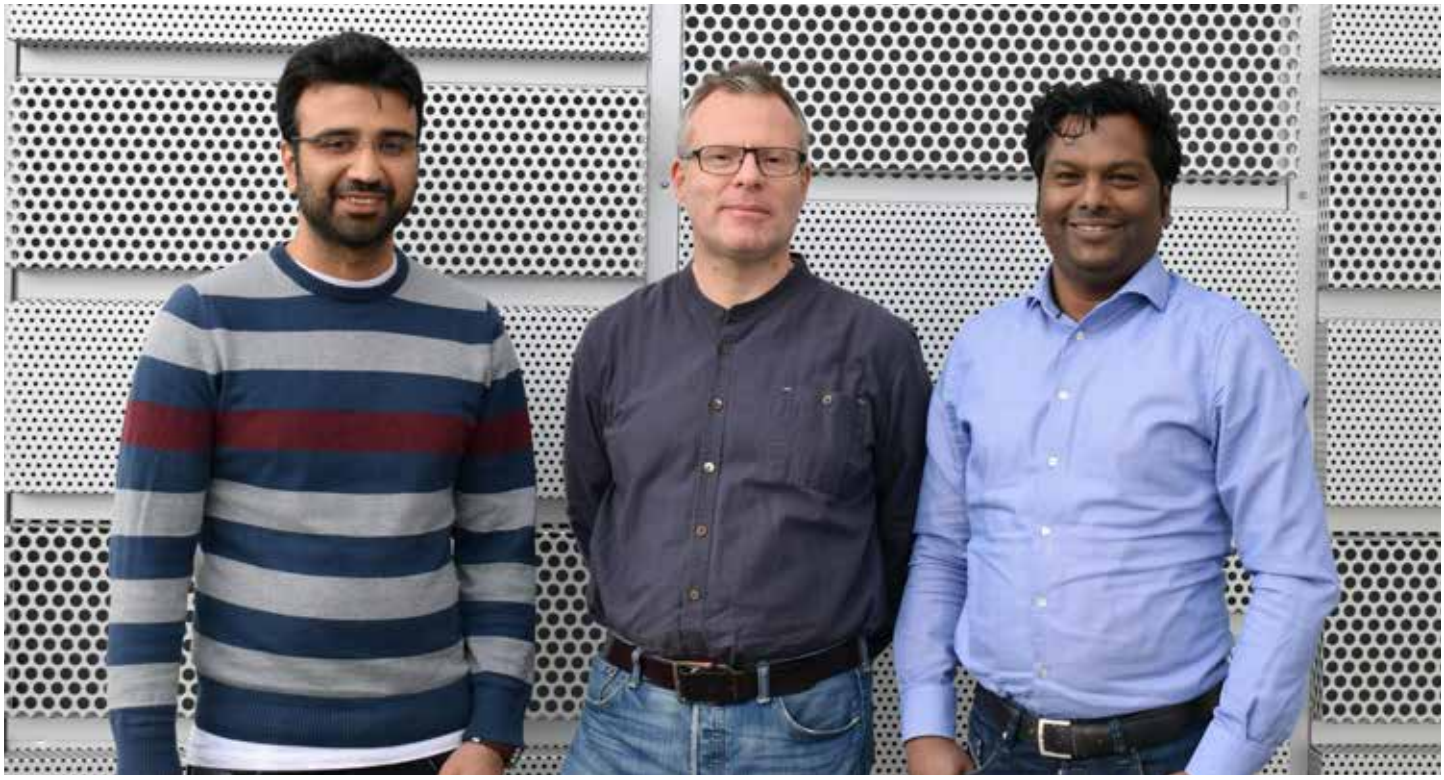


Photo: Kristina Axelson

Jawad Munir, [redacted] and [redacted] representing the [redacted], are working with linking Embedded data together.

and Data Science are considered as two separate schools of Advanced data management, while in reality both of these disciplines are complementary to each other, working on a similar problem i.e. to extract meaningful information from large amounts of data.

Scania has its own Data Lake and multiple data sources, a huge storage repository of all enterprise data of both structured and unstructured data, kept in native format in one place. By tagging and linking the raw data, identifying relationships and adding a meaning to it, we build knowledge graphs. This makes Scania able to make data-driven, fact-based, commercial decisions. By using linked data in the Scania Data Lake, you might build bridges to extracting more powerful and more relevant insights from the Big Data analytics. Building a way to navigate through all the data keeps the Data Lake fresh and clean and prevents data from becoming unusable in terms of creating business value. This is a way of opening up data for consumption throughout the whole company and breaks the silos. With cross-domain traceability, data in one tool can uniquely reference data in another tool, allowing information previously available only in for example Commercial Operations to be consumed and utilized by for example Research & Development or Purchasing.

Linked data, with an RDF database, is based on so called triples. The subject denotes the resource, and the predicate denotes traits or aspects of the resource , and expresses a relationship between the subject and the object.

"We have business scenarios to address" says [redacted] "A Scania vehicle should be able to be stopped anywhere at any time and provide information about identity, what functions it supports, how it was approved for production, where and how it was tested, who approved the test results, who wrote the requirements, how much was test coverage and all other similar queries". Traceability is therefore one of our greatest motivators in this work and will become even more important as we need to support ISO26262 to ensure quality and safety. We will also benefit greatly from an increased traceability as to what kind of new features can coexist with old ones and these can also be used for better precision in our forecasts.

"Today we have limited the work to embedded systems but this method has great potential to reach far beyond these systems", says [redacted]

The purpose with building a tool-chain is to achieve a quick time to market, in an even more competitive market with connected services. At the same time the product quality increases, due to more knowledge based activities, and the complexity is reduced by gaining overview of cross functional information. In addition, a structured environment creates conditions for more creativity and innovation. "Limitations are often a way to try to sort the information, but as soon as you define a limitation you start to create silos. That is why this work is so complex. If we succeed with this I think it will have similar effects of the product information as the modularization has had for the product development", [redacted] concludes. ●