

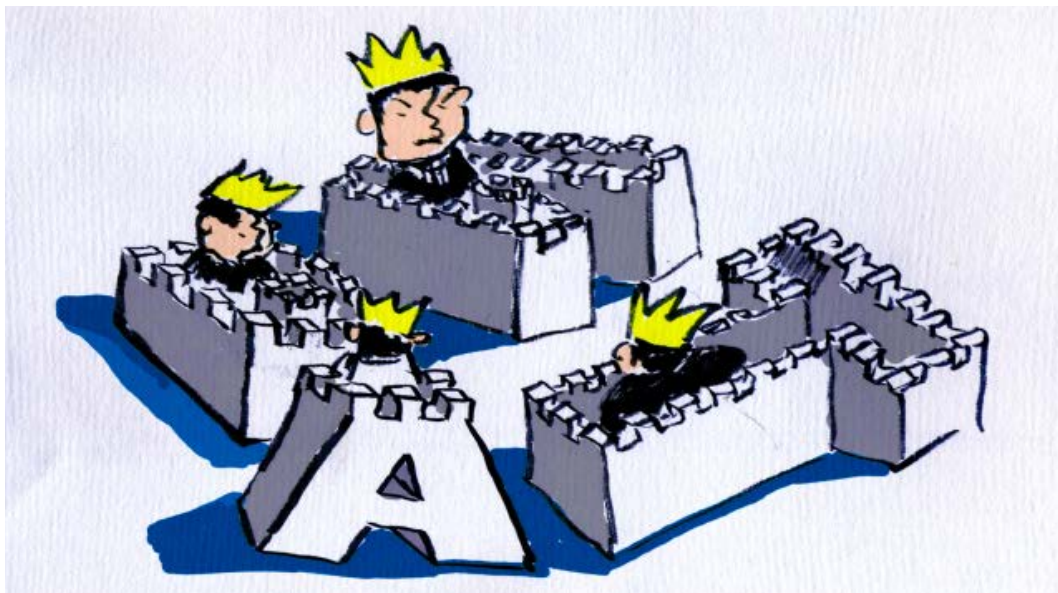
# 'Big data' analysis and modern supreme audit institutions: tearing down the walls of data kingdoms

By Janar Holm, Auditor General, National Audit Office of Estonia

Estonia has adopted e-solutions that have made the country one of the world's most highly developed digital societies. This digital journey has had a considerable impact on its businesses, but also on its citizens. How has this shaped and impacted the work of the National Audit Office of Estonia and its own institutional development? Janar Holm has been Estonia's Auditor General since early 2018. Prior to this, he held several functions in the Estonian government, including being (deputy) Secretary-General of several ministries. In his contribution he explores the concept of big data, what it means for audit offices, including providing examples of how the National Audit Office of Estonia uses data analytics to add value through its audits and to stimulate a better data exchange environment.

## Looking behind the phenomenon called 'big data:' two questions to be asked

I am frequently asked about digital innovation in Estonia and how we, as a supreme audit institution (SAI), are using this success to our advantage – for instance, to what ends are we utilising data from our auditees to generate influential suggestions and what are the examples of big data analysis in our office? When asked such questions, I usually reply with another question – what do you mean when you talk about big data? Because, as a matter of fact, the definitions are somewhat obscure. This time, as I have been given an opportunity to address this issue in a prominent public auditing journal, I would like to raise an additional question – what is the role of a modern public auditor in data analysis, be it big data or not so big data?



Source: Artur Kuus © 2019

My kingdom, my data!

## What do we mean when talking about 'big data?'

The flow of data that constantly surrounds us is beyond common understanding. For instance, during the past month in small Estonia, the number of queries made by various information systems to access interconnected databases via the exchange platform X-road (see **Box 1**) was around 110 million, and this was only information on accessing certain data, not the data itself! Some would call this data, based only on its volume, big data.

To make a distinction between whether we are dealing with big data analyses or simply analysing large amounts of data, I feel it is really important to understand the concept behind this phenomenon called 'big data,' a term which is spreading rapidly in mainstream audit institution language. Our field is known for punctuality in terminology and we expect this from our auditees, so we must practise what we preach.

When looking at prominent definitions today, we must go back to 2001 when Gartner stated that 'big data is data that contains greater variety, arriving in increasing volumes and with ever-higher velocity.' Simply put, big data is large in volume (nowadays reaching petabytes), complex in format (taking data from several sources such as unstructured text, maps, sound, video, etc.) and prompt in pace (so that regular data processing power and software are inadequate). During the last decade, two other aspects have been adopted: value (data has its intrinsic value) and veracity (there are always questions about the reliability of data).

#### **Box 1 - X-Road initiative**

X-Road is a centrally managed distributed Data Exchange Layer (DXL) between information systems. Organizations can exchange information over the Internet using X-Road to ensure confidentiality, integrity and interoperability between data exchange parties. The X-Road is an open source data exchange layer solution that enables organizations to exchange information over the Internet, providing a standardized and secure way to produce and use services.

The first X-Road iteration was developed and launched by Estonia's Information System Authority (RIA) in 2001. In February 2018, Finland's and Estonia's data exchange layers were connected to one another. In 2017, Finland and Estonia established the [Nordic Institute for Interoperability Solutions](#) (NIIS) to continue the development of the X-Road core.

X-Road can be monitored online <https://www.x-tee.ee/factsheets/EE/#eng>.

#### **Data analysis in a modern SAI**

We can probably agree on the following aspects of our work as public auditors:

- our audit offices are acquiring data in large quantities, but I believe we are managing only gigabytes of data in one project, not tera- or petabytes;
- we are using data from various sources but mainly rely on (and check the integrity of) structured databases generated by our auditees (or we try to generate them ourselves);
- usually in an SAI, we deal with data that are not volatile. Generally, we also generate suggestions based on a fixed point in time.

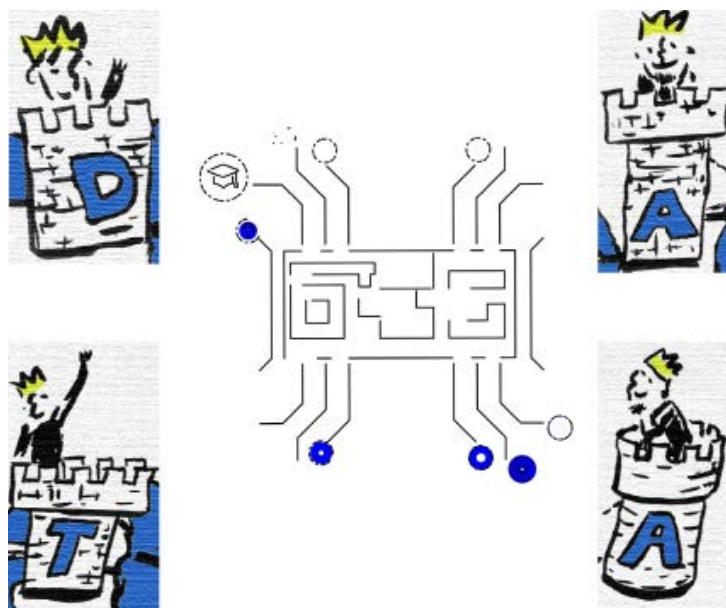
One could argue that we have to apply several methods of data analysis to tackle the challenges of large quantities of data, for example, in some cases, traditional Excel does not suit the job and can become unstable. Then we turn to alternative analytical means such as Python or R programming language for example. Speaking of big data, the data sets we are trying to consider are so voluminous that we cannot manage them without an unusual amount of effort. Nowadays, European SAIs are easily capable of handling quantities of data believed impossible 10 years ago, using the tools suitable for their auditor's expertise. Having different digital tools for different tasks in our toolbox, however, does not imply that we are dealing with big data. Putting up a swing for your child with a professional drill does not mean you are a megaproject constructor.

Regardless of definition, in this new data area, the modern SAI is usually involved in some form of data analytics that requires looking into millions of data fields, comparing various datasets from multiple domains and sometimes even using algorithms to predict potential scenarios. Estonia is no exception – almost all audit planning and implementation requires data mining: mapping appropriate sources, acquiring and investigating data and looking for patterns previously undiscovered. We have combined datasets from various ministries and found ways to enhance their services. For instance, in a relatively straightforward case this year, we gathered data on expenditure in the IT sector in all ministries and their ICT services-providing institutions. Doing this for the first time ever in Estonia, we were able to point out the potential lack of funding for newly developed IT projects and for the sustainability of the sector altogether.

### **Data use, analysis and exchange instead of data collection**

Providing new insights into the data matrix offers great potential for supreme audit institutions but also conceals the threat that an obsession with big data analyses, data mining and gathering becomes a means in itself, not a tool for advancing a state's decision making and improving the wellbeing of people. Instead of generating a massive data warehouse and analytical system in our office, we in Estonia prefer information systems and analytical online tools provided by our stakeholders. Before diving into comprehensive analysis, we investigate what has previously been done in the field. You cannot be effective in pointing out potential problems if you are tied down by reinventing the wheel that is available in the field.

Although a SAI can and should provide value through innovative ways of data management, I believe data analysing should not be duplicated by an SAI and instead we should nudge stakeholders to perform influential analyses themselves. The complementary role is verifying the accuracy of these analyses and promoting the implementation of changes based on these analyses.



X-Road linking the Kingdoms in Estonia!

Source: Artur Kuus © 2019

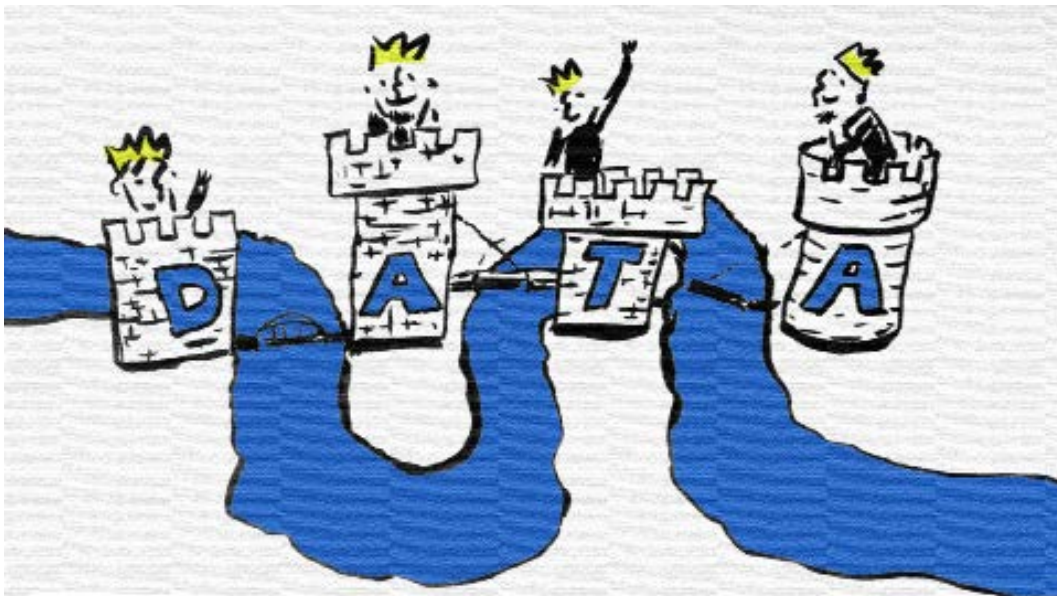
As I pointed out at the XXIII International Congress of Supreme Audit Institutions (INCOSAI) in Moscow in September 2019, we in Estonia see the role of the modern audit institution as more of a promoter of data exchange environment creation and better performance of data analysis by our auditees. An audit institution's role is to dismantle data kingdoms and build bridges between authorities. Freedom to manage data should be a fundamental right for every public authority seeking to serve its citizens better. We are in a unique position as the government has developed X-Road in Estonia, providing a secure exchange layer for all public institutions to utilise. Unfortunately, however, this is still not the case, as data is gathered in silos and even in one field, we found that officials face obstacles when obtaining data from inside their institution.

### **Spreading best practice of data analysis between different public sectors**

In one of our audits, which is currently being finalised, we see various stakeholders in Estonia performing cutting-edge analyses in order to provide better services and even saving lives through better data usage. For instance, the Rescue Board of Estonia is currently mapping all buildings taking data from the buildings registry. They are cooperating with local authorities to obtain data on abandoned buildings and help people who are prone to fire incidents, etc. To do their best in fire prevention, they are using multiple datasets from several institutions, even data from private companies, and neighbours from Nordic countries are visiting to learn from their best practices. So, we do not have to teach the stakeholders how to analyse data, sometimes we should learn from them.

At the same time, we see several authorities lagging behind because they have not taken the time to investigate all the data available, and additionally there are data quality and technical issues that hinder the usage of data. There are problems with blurry responsibilities when developing and executing state services, and the mindset of certain organisations is tangled up in the old way of performing their duties. For instance, in our healthcare system, there has not been any advancement in getting citizens to participate in voluntary cancer screenings. There is not enough data on people who should be in potential focus groups and there is significant potential in using IT to reach out and get in touch with them. When looking for solutions, institutions are pointing at each other and the flaws in information systems, but this has lasted for many years and the solution is nowhere to be seen.

Promoting best practices and pointing out the bottlenecks hindering the usage of data analytics is the main goal of our audit this time. I believe that this is more influential than performing an audit in one single field only to find out that services are not provided in the most efficient and economical manner. Benchmarking data usage for better decision making in multiple fields simultaneously prevents institutions from using their usual argument, 'everybody has the same problems.'



Source: Artur Kuus © 2019

Our kingdom, our data!

### **Developing data analytics capabilities to build bridges**

Tearing down the walls of data kingdoms and being a constructive partner for our auditees, modern audit institutions are facing huge challenges even without adopting big data in our vocabulary. We must develop the mind-set and skills of our auditors so they can ask the right questions, find the answers and the analysis available now, and use the tools necessary for data mining and advanced analysis.

In our new strategy - which we are developing at the moment - the National Audit Office of Estonia will introduce a focus on developing its data analytics capability, and we are looking forward to sharing our expertise and learning from best practices all over the world. This is also a topic we are pushing forward when taking over the IT working group of the European Organisation of Supreme Audit Institutions (EUROSAI) and developing a working programme for the next three years in cooperation with our colleagues across Europe.

### **Ways to move forward for modern audit institutions**

Building on what we have already put in place, and what remains to be done, there are three main propositions we - as public auditors around the world, but also here in Europe - should act upon when it comes to data:

- promoting the creation of a data exchange environment at governmental level and nudging the auditees to perform relevant data analysis;
- identifying best practices and pointing out the bottlenecks hindering the usage of data analytics;
- sharing our expertise and learning from best practices all over the world.