How operators are taking on the 'dirty data' challenge

Operators' data is often siloed, inconsistent, incompatible and incomplete. What can they do to ensure it's of a high enough quality to enable analytics, AI and automation? By Kate O'Flaherty

Artificial intelligence (AI), automation and analytics are the foundation of mobile operators' digital transformation. But in order to take advantage of these technologies to better serve customers and improve network functionality, operators need access to usable data.

This is not as easy as it sounds. Despite years of grappling with so-called 'big data', operators' information is often siloed, inconsistent, incompatible and incomplete. In fact, 57% cited 'dirty data' as the single biggest obstacle to leveraging AI in operations in a recent TM Forum report.

As the stakes get higher with 5G, virtualisation and cloudification looming, automation and orchestration are becoming more critical to service delivery. This is resulting in operators adopting a range of approaches and strategies to ensure the data they have is clean.

It has been a long time coming: In the past, operators haven't taken data very seriously, says Ravi Palepu, Global Head of Telco Solutions, Virtusa. Now data is increasing rapidly, with more added all the time. "The biggest challenge is how to manage that data and how to clean what you already have," he says.

It's a complex issue – and operators' interest in exploiting the potential of AI is exposing the challenges in gaining access to the right data across the organisation, says Mark Newman, Chief Analyst at TM Forum.

He admits 'dirty data' is an issue but says an even bigger problem is the way information is stored and categorised across an organisation. In fact, he says, there are technical challenges involved in aggregating, filtering and using disparate data sets. "That is assuming, of course, that people know where to go in their organisations to access the right data and that different departments have the same policy and approach as to who is allowed to access and use it."

Data silos

Indeed, according to Jennifer Kyriakakis, Founder, MATRIXX Software, telecoms operators haven't been able to leverage data in the same way as modern digital companies because their sources are often "a mishmash of

legacy applications with too many databases that aren't synchronised".

These siloed stacks and multiple billing systems create many differing views of the customer, resulting in data that's not leverageable in the moment it's most valuable, says Kyriakakis.

At the same time, a major challenge for mobile operators is the fact that the technology enabling the collection and storage of data has developed faster than techniques for ensuring its quality and reliability, says Kamal Bhadada, President, Tata Consultancy Services Communications, Media and Technology. "Simply put: We're generating data faster than we can manage it."

Operators accept this, but some are already starting to gain value from their own strategies. Kyriakakis says the operators MATRIXX Software works with are looking at improving their digital IT stacks. "They are simplifying, automating and streamlining processes to ensure a single source of truth for customer, service and usage data that can drive analytics and AI. This data can then be used in real-time to drive actions towards the consumer that are relevant in the moment."

She says clean data starts with a new digital stack so "out of date, legacy, batch-based processes" can be phased out. "Through building out a new digital stack, telcos are creating an IT environment best suited to leverage AI, offering accurate, real-time data to all other systems and channels with the ability to trigger actions in real-time based on network or customer behaviour."

Vodafone's strategy

Katia Walsh, Chief Global Data and AI Officer, Vodafone Group, has other ideas. She thinks the notion of "pristine data" is outdated. She admits there are cases where trusted data is key, such as financial results, "But it doesn't mean every data point has to be exact: it's just agreeing on the key points."

The operator is already having some success with this pragmatic approach. As part of a wider strategy, Vodafone is using data to boost customer experience including communications with the customer and offering predictive care. This area has hugely improved over the last few years, Walsh says. "Before that, Vodafone was doing good marketing, but it was not personalised and informed by what specific customers needed."
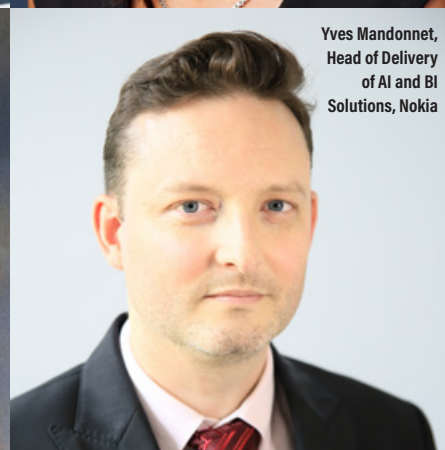
Dr Athina Kanioura, Managing Director and Lead Data Scientist, Accenture Applied

Jennifer Kyriakakis, Founder, MATRIXX Software

Katia Walsh, Chief Global Data and AI Officer, Vodafone Group

Yves Mandonnet, Head of Delivery of AI and BI Solutions, Nokia

Now, around 60% of Vodafone's customer interactions in Europe are supported by AI. "We understand which customers need what, for which point in time, and for what reason – and the time and day of the week that is most relevant to them.

"We want to predict with some certainty when a customer is likely to call and to proactively contact them through the My Vodafone app before an issue happens. That also provides internal efficiencies for Vodafone."

Other parts of the operator's strategy focus on smart internal operations such as network planning and optimising retail footprint and preventing fraud; areas within B2B such as understanding people's mobility around branches or shops, data science and R&D.

Vodafone is not alone in its proactive approach to using data. Laurent Mons, Data Governance and Quality Leader at Belgium's largest operator Proximus says it is investing in data governance as "a vital component" of the wider business strategy.

He explains, "With huge volumes of data coming in from multiple sources, leveraging data governance helps break down data silos across the organisation. More importantly, it empowers employees to go beyond the mere consumption of data and begin trusting and using it to drive business value through analytics and AI."

Proximus uses data governance software by Collibra to contextualise its data landscape and make information fit for use. This enables the firm to develop and maintain actionable algorithms based on reliable and qualitative data sets, according to Mons.

For example, he says, it helps improve field operations: "If a customer is experiencing problems with our service, we can correlate the incident data with our technical data to accurately predict what kind of technician they will need, and the type of support required to be sent out into the field.

"Leveraging data via these types of improvements would be difficult without governance: it plays a vital role in allowing us to find, understand and trust the data assets in our organisation."

The skills gap is key to resolving the data crisis

Skills are key to navigating the rapidly changing data landscape. Yet as networks transform to include concepts such as virtualisation, many operators are finding they are lacking them on a big scale.

Technology has moved very fast, but there is a lack of skillset to use it effectively, says Athina Kanioura, Managing Director and Lead Data Scientist, Accenture Applied Intelligence: "Historically their IT is legacy: they have data management, data architects and engineers, but no AI architects. The fact they have the technology but not the talent to optimise this and use it in [their] business verticals…slows them down."

Ravi Palepu, Global Head of Telco Solutions, Virtusa, agrees there is a huge skillset gap: "I think the option in telcos is not that great largely because the domain is very complex. So just putting a data scientist in place will not solve the problem."

"We need to close the gap between technical competencies when dealing with huge amounts of data, and people who understand the business. This is one of the biggest difficulties," asserts Yves Mandonnet, Head of Delivery of AI and BI Solutions, Nokia, concurring that skills are an issue, whether at an operator or vendor.

Nokia is both retraining employees and recruiting people from outside, he says.

Vodafone is also working hard on increasing skills. In addition to hiring people, the operator is re-training 100 employees. "We put people in a classroom for two months and they get hands on training on machine learning and agile and then they go into data science jobs," Walsh says.

Data lakes can become a swamp

In the struggle to ensure quality, data lakes can certainly help. Among the benefits, Kyriakakis says, data lakes can be used for long-term planning, segmentation and tariff analysis, and process effectiveness. "They help fix macro-level issues. Real-time analytics more directly impacts individual user's experience on a day-to-day basis and is critical for things like 'next best offer.'"

All of the operators have invested in data lakes, says Athina Kanioura, Managing Director and Lead Data Scientist, Accenture Applied Intelligence. She explains, "This started a long

time ago. Many telcos have cloud infrastructure and the data lake is hosted there, but many have multiple data lakes within the organisation which can create complexity with data integration."

This is because it's easy to get carried away getting more data – and the data lake can become a "data swamp", says Kanioura. Therefore, she says, a data quality assurance process is key.

She explains: "There are various technologies in the data lake space. A lot of the data operators have is structured but increasingly, it's unstructured. Many of our clients have struggled to consolidate, cleanse and collate the information together. You have to use statistical techniques and AI to fill the gaps."

> " **The data lake can become a 'data swamp'**

Bhadada reckons data lakes will be ineffective unless they include metadata management and enterprise data cataloguing through advanced data-wrangling techniques; that is, transforming and mapping data from one 'raw' data form into another format to make it more useful for downstream purposes like analytics.

They could include machine learning-based solutions that continuously learn from data quality patterns and feed back into an engine. This would see them constantly building data quality rules without human intervention and improving accuracy in detecting and correcting issues. He says, "This approach promises to free employees within the organisation to pursue more advanced and qualitative tasks, and leaves the heavy data lifting to the machines."

It's true that technology can help, but it's also critical for mobile operators to understand that this on its own is not enough, says Kanioura. "A lot of their investments in the past have been on the technology side. They need to rethink their business processes [as] they are old school, they are slow."

Data lakes also need governance to make sure they work, says Kevin Hasley, Head

of Product at RootMetrics and Executive Director of Performance Benchmarking at IHS Markit. As part of this, he says: "The organisation needs to agree on inputs and outputs. That can be a huge burden especially when data you accept is dirty and does not have a framework before it."

When you clean that data, it should be categorised in the right way, states Hasley.

The results of doing this in an effective way are clear: Walsh points out that when combined, customer and network data can provide new insights. "Sometimes, when you can combine data types [that have] never been used before, you get great results.

"We can see if customers are having network issues using network data and billing issues using billing data. You are combining data sets that have never been put together in an aggregated or sophisticated way and this allows you to discover things you never knew before."

Vast amounts of information are available, but at the same time, data protection is important; even more so since the EU Update to General Data Protection Regulation (GDPR) came into place in May 2018.

At Vodafone, Walsh points out that data is always collected and analysed with customers' permission. "We had a programme in place even before the GDPR: data is pseudonymised so they cannot be identified."

Clearly, operators are making progress, but the 5G networks rolling out over the coming year will need AI and analytics to properly operate and of course, data will continue to be at the heart of this. So, what strategy should operators be following?

Overall, it is important to have good data to analyse, says Tom Foottit, Senior Director, Product Management, Accedian. Once it is clean, Foottit says, the system must be able to share data with other elements in the orchestration ecosystem, via query and real-time notification using an 'event bus' that allows components to communicate with each other.

"Open APIs and the ability to share the results or analysis using machine learning and AI allow this data to greatly improve automation," he says.

It's a major challenge, but rather than having perfect data, Walsh concludes it's important to focus on the outcome. "That's the beauty of AI: We can put the customer front and centre, and provide a predictive and proactive decision for each person."