



Global coverage of the business intelligence ecosystem

Finnair Unlocks the Business Value of Data with Data Catalog

By Ron Powell

This article is based on a podcast Ron Powell conducted with Minna Karha, Head of Data for Finnair, where she is responsible for enabling her business units to maximize the full potential of their data assets. Prior to Finnair, Minna held positions in BI, data management, analytics and data warehousing. Ron is an independent analyst and industry expert for the BeyeNetwork and executive producer of The World Transformed Fast Forward Series. His focus is on business intelligence, analytics, big data and data warehousing.

Minna, Finnair is the largest airline in Finland and has been consistently recognized as one of the safest airlines in the world. Let's begin by having you tell us more about Finnair?

Minna Karha: Finnair was established in 1923, so we will soon celebrate our 100th year of operations. We also are the national carrier of Finland and one of the world's longest operating airlines. For the tenth consecutive year, Finnair has been selected by Skytrax at the World Airline Awards event as Northern Europe's Best Airline, which makes us very proud.

Currently we carry more than 13 million passengers annually and have 6,300 employees. We have 19 destinations in Asia, 18 in the Americas and more than 100 in Europe. We are also a member of the One World Alliance.

What is your role at Finnair?

Minna Karha: My role is to help our units maximize the full potential of our data assets. The starting point, of course, is to understand that data in a modern business world is currency. We need to understand and manage the value of our data assets. This will enable us to unlock the business value of our data, thereby allowing us to continuously increase customer satisfaction. This means, of course, providing tools and best practices and also building the mind-set on new roles within the organization.

I realize that your goal is to become a data-savvy, data-driven organization. To do that, everyone must have access to relevant data. Also, the data literacy rate within your organization must be high. What initiatives have you put in place for people and processes to support your vision?

Minna Karha: Two years ago we started by establishing the role of Head of Data. That was when I joined Finnair. One of my first things was to get to know the organization and understand which roles and skills were lacking. I needed to determine what kind of team I should build to support the already existing organization and be a steward of the roles already in our company. So we ended up building a data solutions development team. We hired data engineers, data architects and data asset product owners. All of these roles were completely new to the company. Also, close cooperation with our analytics and data scientist teams is extremely critical. My data solutions development team is focusing on being a task force that can help different teams of different maturity levels in different situations use data and analytics. We try to help each team when they need help. Sometimes they need more help accessing the data so the data engineering team comes in to help. Sometimes they need more help in data utilization or understanding the inventory or just help to use already existing tools and dashboards, for example. Not only have we added these roles, but we have also started to build a tool. We are currently building a modern data platform that enables collaborative data and analytics work, supporting both machine learning and real-time analytics.

The other big tool or platform that we have been establishing to make our data assets visible is a data catalog. Making our data assets visible through the data catalog allows us to know what data we have and realize its full potential.

Gartner, a leading analyst organization, has identified data literacy as a critical component for a data-driven organization. Can you talk about the importance of data literacy for your team?

Minna Karha: I agree that data literacy is very important. For me it means that everybody who uses data understands the origin of the data and is able to estimate how much they can rely on the specific data. Then they can understand what decisions they should make based on that data and the information that they have at hand.

How do you plan to leverage your data assets?

Minna Karha: Obviously we are talking about an organizational transformation so it's very important to introduce tools, platforms and technologies. The value, however, actually comes from the skills and mind-sets. It is especially the mind-sets and company culture that require a lot of time and a lot of work. For that, it is very important that there is a commitment from the key stakeholders in the organization.

We have defined Finnair's data and artificial intelligence (AI) strategies. We have commitment and ownership for both strategies from our executive board and board of directors. The support starts from the top, and we also work closely on a daily basis with key stakeholders across the organization to build a network that then can support the change management and the transformation, which is very critical.

That's fantastic. Without top-down support, it's very hard to implement an enterprise-wide vision, wouldn't you agree?

Minna Karha: Yes.

It seems like everybody's talking about using AI extensively today, but I've really only seen it being used sparingly. I think the hype is still way beyond the actual implementation capabilities of many companies today.

Minna Karha: I agree that many companies are still at the starting point. In our case, we identified a lot of potential. Therefore, we have been doing machine learning for a few years, mainly focusing on the commercial unit – customer experience, how to predict what customers want, what the propensities are for certain actions, and how we can recommend the most interesting options for them. For example, during the past year, we have tested and implemented machine learning with flight prediction models.

In general, I am confident we are on the right path moving forward, but, of course, the core enabler for all of this is the quality and sufficiency of data. That is something where we have work to be done to make sure that our processes, partnerships and integrations create constant quality data that will fuel our machine-learning models.

You mentioned tools and platforms. What does your data technology stack look like?

Minna Karha: We are building our data platform on Amazon Web Services, and we are taking the approach of really building it the way software development is done. Because data comes in so many different formats, we want to make sure that what we build is scalable for different formats and different use cases. We have a basic structure. We have the data lake where all the data flows into, and then on top of that we also have built a modeled data warehouse for our users who prefer using the modeled data for use cases. We have AWS S3 as the data lake. We use tools like AWS Glue for data ingestion, and then we have Snowflake as the data warehouse. And then for visualization and analytics, we also use Power BI. Some teams are still using Cognos analytics, but we are moving more toward Power BI as the end users feel it is very intuitive and easy to use.

We already have machine learning implemented, but we don't have a company-wide, common solution there. We are looking into building a data science lab platform that is common for all the teams who are currently coding the models with R, Python and so forth. This is something we still need to unify.

When you talk about quality data and managing all of the data, earlier you had mentioned a data catalog. Where does the data catalog fit into your vision?

Minna Karha: First of all, the data catalog for us is the place for our data asset inventory. We want it to be the place where anybody who owns data or has access to new data makes sure it is also documented in our data catalog. We have helped that work by creating templates so that it's easy to add the new data sources to the catalog for others to use. We also frequently monitor the use of the data catalog to identify the usefulness of the content. Also, data quality is part of the catalog. We have templates for what we call data SLAs, meaning that we want every data source documented. We also want to document whether the data quality meets the service level agreed upon with the vendor or whoever is responsible for the data creation. That will then give us an idea of how many SLAs are in place and how many data sources meet the expected data quality levels. Then we can, of course, monitor that to determine if the incoming data matches the agreement specifications.

Can you talk about the selection process or the criteria you used in selecting a data catalog?

Minna Karha: Our starting point involved an interesting use case. A new data analyst joined the organization. It took him several weeks to find relevant data for his first analysis, and it turned out that he really needed to know who to ask. In our company, knowing what data was needed and how to get access to it was kind of tribal knowledge. We used that as a starting point, and then created the business case from an efficiency point of view – determining how much time was wasted trying to find the needed information. We tested a couple of tools, and we really wanted to do a very quick proof of concept that would test in practice how they work. We tested a couple of tools, and then we selected the best one for our purposes. The driving forces were that we wanted to have a very intuitive tool. We didn't want a tool that required extensive training and hands-on support before the users are able to use it.

We also see the data catalog as a tool for the whole organization so it needs to be very easy to search. We also wanted it to be fairly simple and easy to add content because if that's not easy to achieve, then nobody will keep the content up to date. The best feedback was from an analyst that was part of the implementation project. He said that it's never easy to document, but with the data catalog tool it is the least painful it can be. It is really nice to do the documentation there, which I think is good feedback.

The documentation side has always been a big issue with technical people as well as business people. What data catalog did you select?

Minna Karha: We selected Alation.

With Alation, are there things that you're doing now that you weren't able to do in the past?

Minna Karha: There are a lot of things that we are able to do now. So, first of all, we have one place where we can actually document and make our data asset visible. We also are able to collect in one place our business glossary and the SLA agreements for data. And we now have one place where we are documenting our AI models for the many different teams that are building models. We have found, for example, that a couple of teams had been building models for almost the same purpose. Now we are able to make that work more transparent between the teams.

That's always great to get rid of redundancy within an organization. We always talk about reuse in the data space, but it's so hard to get people to utilize things that have already been built.

Minna Karha: That's true. That's also why we have established KPIs to regularly monitor and publish the amount of usage. Because we are still in the implementation stage, there is a lot of content to be added. We want to be able to monitor the catalog content and show how it's progressing. Then if it's not progressing as we had hoped, we will have ways to take actions. When you have relevant content and when it's useful, then you have a lot of users. At the end of the day, I believe that the amount of daily use is really the key.

From a user perspective, obviously you're seeing growth in the number of users. Are you seeing greater collaboration between the technical organization and the business?

Minna Karha: Yes, we're finding that people who see the data catalog for the first time are excited that, for example, they can find information that they had been unable to find in the past.

That's fantastic. Looking forward, what operational areas will benefit from this collaboration and data utilization?

Minna Karha: I definitely see that when we have the full inventory available, there will for sure be data sets that some or part of the organization did not know existed. They will be able to use that data for their benefit. I also see that currently there are similar data sets used by different teams. Having more alignment brings business value and better data governance because we know the purposes for which the data and the numbers are being used.

That's great. Minna, it's been a pleasure talking with you, and thank you for sharing your vision and the importance of data literacy, which is allowing your business units to maximize the full potential of Finnair's data assets.

Author Bio: Ron Powell

Ron is an independent analyst, consultant and editorial expert with extensive knowledge and experience in business intelligence, big data, analytics and data warehousing. Currently president of Powell Interactive Media, which specializes in consulting and podcast services, he is also Executive Producer of The World Transformed Fast Forward series. In 2004, Ron founded the BeyeNETWORK, which was acquired by Tech Target in 2010. Prior to the founding of the BeyeNETWORK, Ron was cofounder, publisher and editorial director of DM Review (now Information Management). He maintains an expert channel and blog on the BeyeNETWORK and may be contacted by email at rpowell@powellinteractivemedia.com.