ANNUAL REPORT 2024

DNB Data Science Hub

Foreword

I am yet again very happy to present our Annual Report to you. Looking through the pages of the 2024 edition, I'm happy to see that we are maturing as a department and, more importantly, as an organization. Thanks to great work of our IT and data colleagues at DNB and the ECB, our tooling and access is improving. This allows us to focus more fully on the core data work. Moreover, our projects are becoming bigger and more ambitious.

We have continued with providing advice, guidance, and project execution of data science projects across the organization. We finalized 7 projects with 6 different divisions with clients still awarding us very high marks. To enthuse the wider data science community we still organize the Open Source Lunches which now also feature external presenters. To further improve the knowledge base, we ran several workshops with the GIT for version control workshop as a steady favourite. New topics were a two-day intensive hands-on course on *Machine Learning – Tools and applications for policy* and *Using PyTorch for neural networks*. All together our activities attracted 277 unique participants.

Externally, we have organized the second *Data Science Event for the Public Sector* joint with several government agencies, gracefully hosted by the Netherlands Bureau for Economic Policy Analysis (CPB). We had over 80 participants and the 2025 edition is already in the works at the Frederiksplein.

2024 was also a year where we stepped up our external outreach. Many other authorities have similar mandates and are hence interested in the same risks and the same topics. Our code is open source and hence could, with relatively little effort, be implemented elsewhere. Open source code development thrives if there is an active ecosystem supporting this. To foster such an ecosystem we first need to identify pockets of support for our vision. With projects such as the Digital Twin for Physical Climate Risk and Interconnectedness we hope to kickstart support for such an ecosystem.

Hope to meet up at some point and see if we can travel further on the data science road.

Iman van Lelyveld

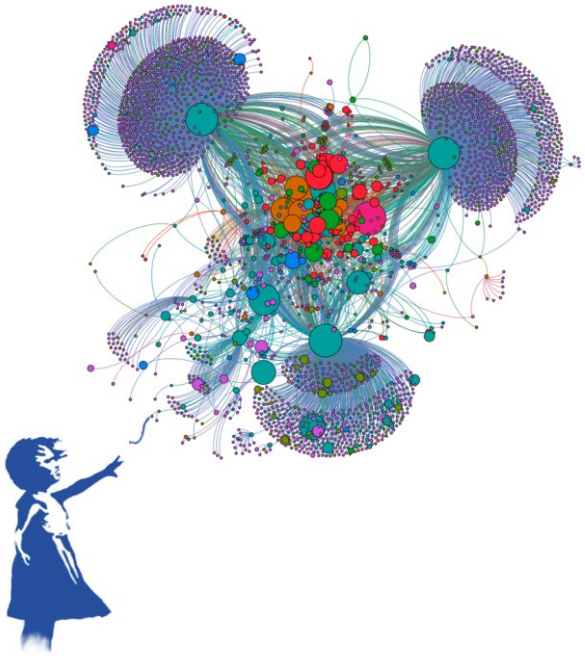
DataScience Hub

What do we do?

The DSH serves as the central hub for data science at DNB, providing advice, guidance, and project execution across the organization. As the core of a hub-and-spoke model, the DSH:

- supports all departments by collaborating to identify data opportunities and develop data science solutions
- promotes the interests of data users in the development of data infrastructure
- supports and connects the data science community by providing training and organising events

This Annual Report provides an overview of all our activities in 2024.



Are you interested in our work and activities? Feel free to reach out to us: data_science@dnb.nl

Our team in 2024


Alessandro Pollastri


Abel Koch


Christian Franssen


Diana Struijk


Hannah Froklage


Iman van Lelyveld


Kristy Jansen


Martijn Buitink


Michiel Nijhuis


Milan Karsten


Natalie Kessler


Robert Hofman


Tim Haarman


Zooey Bossert

2024 in quotes

"Our collaboration was efficient, effective and fun. In the end, this collaboration not only proved to be fruitful for FM, but also for the Central Bank domain."

FM, DRIP

"The Data Science Hub has been very cooperative since the beginning, and always available to discuss the progress on the project."

EUBA, FICO

"We had a good experience working together with DSH. All deliverables agreed upfront were met. In case of questions from our side (which were only few), response was quick and adequate. Great collab!"

EBO, PRISM

"DSH was very cooperative and established the Cosmos Graph DB really fast. They are technically knowledgeable and know what they are talking about."

TV, MGDB

"Terrific. They provide great support. We were well included in the developments. Very nice."

BEVO, SDDM

"The collaboration went smooth, the DSH offered a good solution to the automation we wanted to perform, helped to set it up to make it work well for our application and was easy to reach for support or questions."

FM, FMFICO

2024 in numbers

6
Open Source Lunches organized

7
data science projects finalized

6
collaborations with different divisions

9.3
overall client satisfaction

8
Open Source Workshops organized

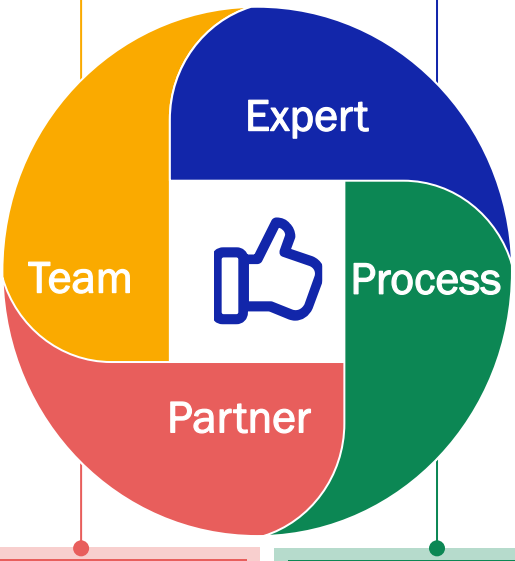
218
unique participants that attended our activities

DataScience Hub

Goals set for 2024

Goals	Measures
We build on our current knowledge	9 50% of the relevant topics are studied and applied
We want to positively influence relevant internal parties	10 The average grade for the input we provided to the DSAP is at least an 8

Goals	Measures
Broaden the knowledge of data work at DNB by educating clients in projects	1 All data science projects take into account the DSH manifest
	2 50% of the finalized projects has commits (coding) of the business
Providing reausable solutions to clients as stated in the manifest	3 At least 20% projects lead to a follow-up project



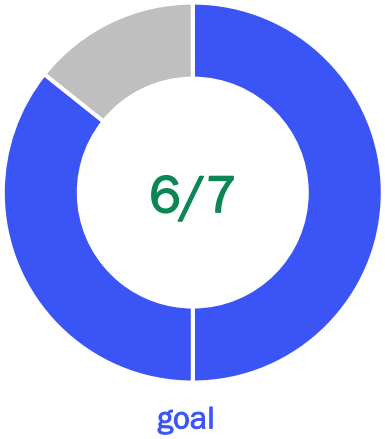
Goals	Measures
We communicate our proceedings so that our colleagues know where to find us for a data science project	6 At least 5 different divisions with a big project
	7 The average grade for our work on projects is at least an 8
We stimulate a data science community	8 150+ unique participants attended activities organized by the DSH

Goals	Measures
We aim for wide usability and the relevance of external stakeholders within our whole working process	4 At least 40% of our finalized projects are (in)directly relevant for another business unit
	5 At least 50% of our (finalized) projects are shared externally

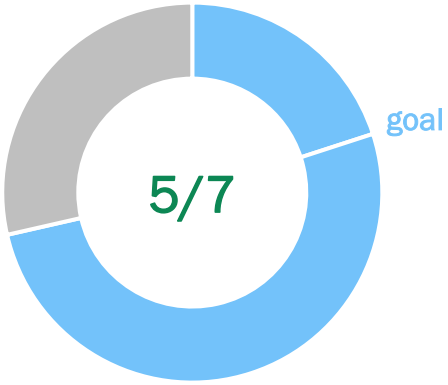
Goals set for 2024

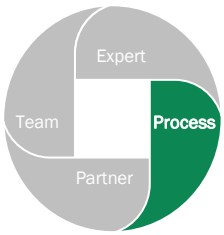
1. All data science projects take into account the DSH manifest

2. 50% of the finalized projects has commits (coding) of the business

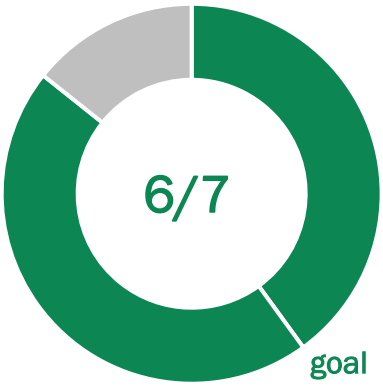


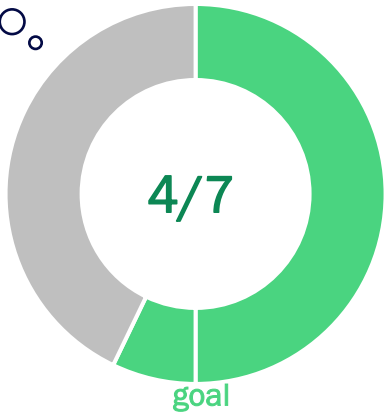
3. At least 20% projects lead to a follow-up project





We had a lot of external contact besides projects, see the external outreach

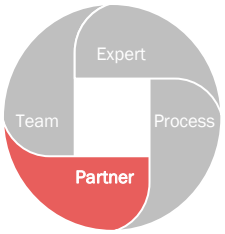




4. At least 40% of our finalized projects are (in)directly relevant for another business unit

5. At least 50% of our finalized projects are shared externally

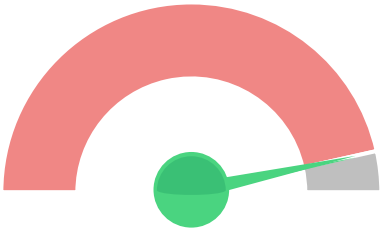
Goals set for 2024



We finalized projects with 6 different divisions

EUBA
FS
TV
EBO
BEVO
FM

The average grade for our projects is a 9.3



Over 224 unique participants attended activities organized by the DSH

6 At least 5 different divisions with a project

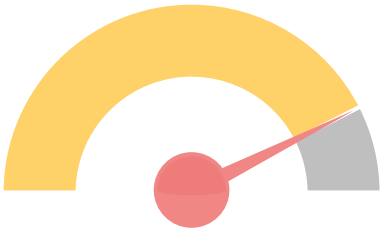
7 The average grade for our work on projects is at least an 8

8 150+ unique participants attended activities organized by the DSH

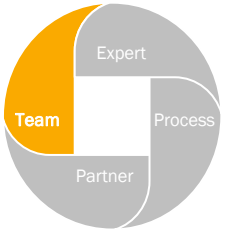
9 75% of the relevant topics are studied and applied

10 The average grade for the input we provided to the DSAP is at least an 8

Our colleagues have actively followed relevant trainings



The grade for our input is an 8.5



DataScience Hub

Finalized projects

PRISM

Platform for Retrieval and Ingestion of Statistics for Models (PRISM) is a brand new, multi-model platform that can retrieve and store input data for models, but is not limited to one specific model, it can handle multiple models with ease. It is the successor to the bespoke platform that DSH built in collaboration with ECMO for the NIPE model.

PRISM modernizes the current process by automating data collection allowing modelers to request a data refresh with the touch of a button.

MGDB

Metadata Graph Database (MGDB) using public metadata of insurers, has been successfully implemented to demonstrate the possibility of creating a Cosmos DB graph database for storing and querying data from DSW.

SDDM

Druktevoorspelling Sensor Data (SSDM) was a project to utilize sensor data of the office building to monitor and predict the occupancy levels of the restaurant area.

FSTR2

The goal of this follow-up project was to extract and clean the Securities Financing Transactions Reporting (SFTR) data contained in the ECB DEVO database and make that data available in a secure way in the DNB PowerBI environment.

FICO

The aim of FICO (FINREP, COREP) project was to build an outlier detection Python package that can be used to monitoring credit risk indicators by supervisors. This resulted in a more effective way to review series that are normally ignored, as supervisors are now directed to the most interesting observations.

The package can be used on the Agora data (FINREP, COREP), or the KRIs (key risk indicators). Moreover, the package is also usable on other data formats, as it is applied to FM market data as well (see the FMFICO project).

FMFICO

FMFICO is a follow-up project where the FICO outlier detection package, which was previously developed by DSH, was applied to monitor market fluctuations and provide intelligence.

DRIP

DRIP investigated whether the NIPE architecture could be used as a blueprint for automating the benchmark process at FM. During the project Build-By-Business (3B) was introduced.

DRIP was the use case to template the 3B-production process, leading to the automation of the benchmark process in Synapse, managed by IT.

Ongoing projects

ODM

Since 2020, pension funds and insurers have provided extensive data to DNB, making manual quality checks unfeasible. ODM is a self-learning outlier detection algorithm to autonomously ensure data quality for FTK and SII data.

DFTC

DFTC aims to use the PRISM infrastructure to build and deploy a cloud-based ML model for DFROG that periodically produces GDP forecasts. It involves creating the necessary cloud infrastructure to train, deploy, and monitor the model.

FSTR3

The SFTR Stress Testing with Virtual Lab project aims to develop a real-time stress testing tool for the repo market using SFTR data. This tool will leverage an ETL pipeline from the FSTR2 project and historical data for predictive analysis.

SYN

The lack of representative test data can lead to increased incidents and dissatisfied users. To address this, SYN focusses on developing a package that generates synthetic data, maintaining the characteristics and patterns of the raw data.

INTER

The project aims to enhance understanding of the interconnectedness between Dutch Significant Institutions (SIs) and Non-Bank Financial Institutions (NBFIs) by integrating data from SFTR, EMIR, AnaCredit, and SHSS databases. This initiative addresses the lack of clarity in bank-NBFI exposures, highlighted by cases like Archegos.

The goal is to deliver a comprehensive note on exposures, gap analysis, data source overviews, and capital impact assessments, improving risk visibility and management.

FICO3

FICO3 will use the FICO outlier detection package for FINREP and COREP key datapoints to provide accurate outliers from banks. The new tool will be user-friendly, offering visualizations via a Power BI dashboard and/or outlier scores.

Ongoing projects

SFODT2

SFODT2 or Digital Twin project builds on the existing digital twin and is suitable for experimental use in microprudential and macroprudential supervision. This involves improving the tool's deployment to a cloud provider, enhancing the UI performance, connecting supervisory data, and facilitating real-time data connections.

The tool will help model the effects of climate-related events on financial assets using the latest available data.

NUMI

BEVO_DNS maintains a numismatic collection of over 400k items, each photographed and stored on a network disk. This project aims to automate the process of exporting requested images, which is currently labor-intensive.

FETCH2

FETCH2 is a refactoring of our DNB DataFetcher package (project FETCH).

LSTT

Modernizing the Liquidity Stress Testing Tool (LSTT) by developing an efficient workflow for executing stress tests on banks' liquidity positions. This involves creating a cloud-based Python engine to automate standard and custom scenarios, improving reproducibility and auditability.

LSTT allows colleagues to input custom scenarios and access results independently, serving as a template for other stress tests and ensuring the use of DNB Confidential data with appropriate security.

FICO4

FICO4 will use the FICO outlier detection package to improve the flagging of relevant price changes for FM, replacing arbitrary thresholds.

NOW

NOW aims to nowcast CO2 emission indicators to provide timelier data for users, addressing the time lag in current data sources. Challenges include limited historical data, many missing values, and the impact of the COVID-19 pandemic. The solution will involve creating a Python package with forecasting methods, performance evaluation tools, model explanation features, and validation modules.

Internal outreach

The internal outreach features several key events within DNB that DSH either organized or participated in. Our goal is to increase awareness of data science and data-driven work within the organisation.

Open Source workshops

At the DSH we organize workshops to help increase the data science knowledge at DNB. In collaboration with the DNB academy, we organize three standard courses: Version control with GIT, Clean & Responsible Coding and Explainable AI.

This year we have added one more workshop: Building neural networks with PyTorch. The PyTorch course proved to be a welcome addition with 13 participants signing up within a few days of the announcement! Apart from the PyTorch course, a two-day intensive hands-on course on *Machine Learning – Tools and applications for policy* was given. In total, we have organized 8 workshops in 2024. We are happy that we managed to enable colleagues to increase the level of their coding and data science skills as indicated by the participants via surveys.

The need to have more data skills within DNB is still far from satisfied as the waiting list for our courses is growing again. All in all, the DSH was able to help DNB take another good step towards a more data driven institution!

SupTech Horizon

On October 30th, the DNB SupTech Horizons Event 2024 brought together the Supervision colleagues for inspiring keynotes by Steven Maijoor and Elizabeth McCaul, engaging panel discussions, and personal insights into supervisory tools. The event featured tool demos and lively conversations, as well as workshops on Generative AI, Explainable AI, and Prompting. The Data Science Hub was invited to host the Explainable AI workshop, which sparked great discussions and interest.

A big thank you to our colleagues from Digitale Strategie Toezicht for their efforts and sharing the latest developments in SupTech at DNB.

AI Panel

In April, DNB, together with the AFM, has presented its vision on the supervision of AI. To bring this to the attention of the sector, a symposium 'The impact of AI on the financial sector and supervision' took place last summer. In the panel discussion on AI and fairness, Michiel Nijhuis shared the developments taking place within DNB in the field of AI and machine learning. He illustrated how DNB deals with the issues that the use of AI and machine learning entails. The discussion provided insight into the various considerations that need to be made when applying AI technologies from both a technical and a social point of view. These types of symposia are important for promoting a responsible integration of AI at supervised institutions.

Open Source Lunches

2024 was a great year for the Open Source Lunches! Not only where fantastic DNB projects presented, but we also had some external guest present their work. Brinn Hekkelman from CPB presented his work on fair algorithms and Lauren Waardenburg presented her research *Knowledge Brokerage in the Age of Learning Algorithms*. Furthermore, there were very interesting presentations on DNB projects with topics ranging from data engineering to data science.

We want to thank everyone who presented at the Open Source Lunches and all participants. We hope you enjoyed it as much as we did! If you have any projects or topics you would like to present the coming year, feel free to contact us!

External outreach

The external outreach features several key events outside DNB that DSH either organized or participated in. Our aim is to promote knowledge sharing and collaboration – for example by exchanging code.

ACE visit to DSH

In June, the DSH hosted the Bank of Finland's Analytics Center of Excellence (ACE). This all-day session provided a platform to discuss how data science is applied within DNB and ACE, alongside the infrastructure that supports the effective use of available data.

The event featured presentations from the DSAT team, Databalie, DSH, and ACE. These sessions offered valuable insights into diverse approaches and strategies, making it a great opportunity to exchange experiences, challenges, and pitfalls.

The collaboration between DSH and ACE underscored the importance of knowledge sharing and collaboration between central banks within the Eurosystem to advance data science capabilities.

BISIH AI Symposium & SupTech TechSprint

The BISIH London Centre organized a three-day event combining a symposium focused on AI with a techsprint on various innovative topics in SupTech.

The first day covered developments in AI and other innovations in SupTech, RegTech, and MonPolTech. The techsprint took place over the following two days as a new way to generate potential projects and catalyze further cross-jurisdictional activity. During the techsprint, multi-disciplinary teams used design thinking to develop technology-based ideas or PoCs to address common supervisory challenges, with our project on Digital Twins for Climate Risk as one of the topics.

The event provided valuable insights for our Digital Twin project. It was an excellent opportunity to discuss key advancements and opportunities for collaboration.

EDKP Hackathon

In early June, the BISIH Singapore Centre and the Monetary Authority of Singapore organized the EDKP Hackathon. The hackathon was an intense and productive three and a half days, where many experts from different central banks and authorities brought a wealth of knowledge and perspectives that truly enriched our time together.

Six different use cases on GenAI and climate risks were developed. Together with our colleagues from Banque de France, who also collaborated with us on our project on Digital Twins for Climate Risk, we worked on a new use case that introduced a new physical climate risk to the MVP framework and used only satellite data, also for real estate. In addition to our team, another team from the Bundesbank worked on the topic, applying the MVP framework to their jurisdiction.

It was an inspiring week that resulted in valuable insights and offered the chance to expand our networks with useful connections.

External outreach

The external outreach features several key events outside DNB that DSH either organized or participated in. Our aim is to promote knowledge sharing and collaboration – for example by exchanging code.

EIOPA Data & SupTech Forum

In October, the DSH got to travel to Frankfurt to participate in the Data & SupTech Forum hosted by EIOPA. This event brought together experts from across Europe to share the latest developments in SupTech.

We proudly presented our project on Digital Twins for Climate Risk, showcasing how innovative technologies can address complex supervisory challenges. Among the many insightful discussions, notable highlights included the implementation of Generative AI at Banque de France and a Data Quality SupTech project presented by one of our DNB colleagues.

It was an inspiring two days, stressing the importance of collaboration in advancing SupTech solutions across the industry.

Latvian Fintech Forum

In November, Iman van Lelyveld delivered a keynote speech at the Latvian Fintech Forum 2024.

The speech discussed how AI might affect the financial sector putting the spotlight on:

- how to use AI in detecting malignant transactions
- an example of how individual actors' AI-driven decisions can lead to macro problems
- a more philosophical point; are we able to really fully use AI if we have trouble understanding it?

Finland's Impact of AI on Economy, Finance, and Supervision

In November, the Bank of Finland and the Finnish Financial Supervisory Authority jointly organized a two-day seminar on the impact of AI on Economy, Finance, and Supervision. Tim Haarman attended to present details on our Interconnectedness, Digital Twin, and Know Your Customer projects, and later joined an engaging panel discussion on the future impact of AI on our work as supervisors. The ensuing discussions showed significant interest from attendees across Europe, highlighting the benefits of open source development and the exchange of knowledge wherever possible. We hope that this leads to more successful collaborations in the future!

External outreach

The external outreach features several key events outside DNB that DSH either organized or participated in. Our aim is to promote knowledge sharing and collaboration – for example by exchanging code.

SEACEN

In November, Iman van Lelyveld got the opportunity to present virtually to in the SEACEN Course on Retail Payment Systems. SEACEN is a regional learning hub for central banks in the Asia-Pacific region.

The presentation highlighted the findings in a recent publication in the Harvard Data Science Review. In this article the lessons learned from 5 years of operating the DSH are discussed. In particular, our project on bank note quality led to a good discussion.

Data Science Event

In December, the second edition of the Data Science Event took place, which the DSH organized on behalf of De Nederlandsche Bank, in collaboration with CPB Netherlands Bureau for Economic Policy Analysis, the Central Bureau of Statistics, the Ministry of Finance, and the Ministry of Justice and Security.

The event focused on transparency and communication around data science in the public sector. Inspiring keynotes and an engaging panel discussion provided valuable insights and lively conversations. Additionally, various workshops offered in-depth exploration of current data science projects and the opportunity to share experiences.

We are looking forward in hosting the next edition in 2025.

CSL SupTech Week

The Cambridge SupTech Lab (CSL) SupTech Week is the largest annual gathering of global suptech experts and innovators. The DSH contributed by organizing two panels, one on *How to organise data driven work* (youtube) and another on *Tackling climate risk* (youtube). Additionaly, Iman van Lelyveld was a panelist for *Suptech capacity building and cross border collaboration the EU Supervisory Digital Finance Academy* (youtube). All these sessions were very lively and led to good follow-up discussions. Supervisors and central bankers all face the same challenges and thus always benefit from information exchanges.

External outreach

The external outreach features several key events outside DNB that DSH either organized or participated in. Our aim is to promote knowledge sharing and collaboration.

EU SDFA Hackathon

In June we participated in the 6th Advanced EU-SDFA Training Week on AI and ML for SupTech. The EU-SDFA is an EC commission funded initiative to help supervisors become more digitally savvy.

The participants participated into intensive, cross-functional team activities, focusing on problem-solving and prototyping innovative solutions. The course started with advanced theoretical lectures and presentations on use cases from other SupTech projects. Abel Koch discussed responsible coding.

After setting the scene it was time to dive into the hands-on part. It was a dream we entertained for a long time: getting together with data scientists at other NCAs and produce a useful tool. Abel Koch and Robert Hofman were leading groups on The Digital Twin and Interconnectedness, respectively. Others were tackling NLP/LLMs and forecasting challenges.

This week was one of the most valued session of the programme and has led to some follow up meetings. Moreover, we will be involved in organizing the 2025 session as well

Harvard Data Science Review

After a rigorous peer review an article about our lessons learned was published in the Harvard Data Science Review. In the article we discuss that public authorities, such as central banks and supervisory authorities, are not known for their ability to quickly adopt new techniques in a rapidly changing world. However, these authorities play a central role in society, such as safeguarding the financial system. The challenge of keeping the financial system safe is formidable and data science could potentially help. We discuss how to leverage the potential of data science using our experience at one of these organizations: DNB, the Dutch central bank. The dual role of DNB as central bank and prudential supervisor ensures that the lessons learned are of interest to all stakeholders in the public and financial sector. Furthermore, by adopting a strategy that prioritizes cloud-first and establishing a DSH knowledge gained has wider use.

The goal of our study is two-fold. First, we demonstrate the significant potential of data science in nine lessons, all supported by our own projects. Based on our experience, we highlight the aspects necessary for fruitful data science work. We will argue that AI should become part of daily work processes to reap the full benefits. Second, we share how we work at the DSH with the intent of providing practical guidance and inspiration to other organizations that are thinking about implementing data science in their organization. We thus leave out much to the technical details of the—sometimes quite advanced—solutions we have provided to our clients.

DeNederlandscheBank
EUROSYSTEEM

DataScience
Hub