ANNUAL REPORT 2025

DNB Data Science Hub

Foreword

Our fifth Annual Report lies before you. Yet again, a great year with several projects going into the next phase, some traction with our outreach and very promising developments on the infrastructure and policy front. If you want to read about what we learned in the past few years, we wrote this up in an article in the Harvard Data Science Review but for this year, let's get into it. PRISM, a cloud platform for collecting data and running forecasting models has been onboarded by IT. This is the capstone of a journey with our colleagues in the forecasting department that started with the DataFetcher package which allowed us to move away from manually downloading and collating data. It is an ideal example for the build-by-business (3B) policy that DNB has adopted this year of how co-creation and sequential development can lead to top-of-the-line applications in production.

Our joint Digital Twin project has also moved to the next stage as Project Danu at the BIS Innovation Hub. This project aims to develop a modular digital twin allowing authorities to adapt to locally available data and assess physical climate risks like floods and wildfires.

The necessary conditions for data science are looking up as well. The 3B policy provides a framework for how businesses departments can start experimenting and, over time, migrate towards production level analyses and applications. Our IT colleagues have repositioned themselves with a customer-facing team. Databricks has been rolled out, drastically cutting development time.

Looking outside DNB, we have been, again, in contact with many organisations. In total we talked with more than 30 organisations. A highlight was the third edition of the Data Science Event, joint with CBP and CBS, with over 130 participants from Dutch public authorities. Internationally we are finally seeing some traction on sharing our code with ongoing discussions with several NCBs.

All of this would, of course, not have been possible without our amazing team. We had quite a lot of turnover this year; sad to see people leave, proud to see the cool places they have gone to and amazed by the fantastic new joiners.

Iman van Lelyveld

DataScience Hub

What do we do?

The DSH is the place for advice, guidance and implementation of data science projects

Tailor-made solutions

From strategy to execution for your specific data challenges

Proven products

Proven solutions directly applicable to your data with fast turnaround time

Demand-driven

Community & Development

Increasing data science knowledge through training and events

Representing interests

Representing the interests of data scientists regarding data infrastructure and 3B policy

Supply-driven

This Annual Report provides an overview of our activities in 2025. We'd like to think along with your data, please reach out to us: data_science@dnb.nl

Our team in 2025


Abel Koch


Aimée Donsu


Alif Verkampen


Ans Vlooswijk


Christian Franssen


Hannah Froklage


Iman van Lelyveld


Kristy Jansen


Leonardo Becarelli


Martijn Buitink


Michiel Nijhuis


Milan Karsten


Natalie Kessler


Sofia Yuan


Thao Le


Tim Haarman


Zooey Bossert

2025 in quotes

"With FICO the reporting analyst spend 70% less time on the identification of outliers"

-TNBE, FICO3 project

"The DSH did not only bring expertise but also leadership to the project, taking the initiative the advance the work both within DNB and internationally."

-SFO, SFODT2 project

"Excellent, thanks to the creative and cooperative help of DSH we have streamlined a process and won valuable time"

-BEVO, NUMI project

"The collaboration with the DSH was a pleasant experience, the DSH is very knowledgeable and helpful."

-TGBB, EMCAP project

"The project brought the expertise of the DSH regarding the Digital Twin together with the physical risk data of the sustainable finance statistics teams."

-STAT, STATDT project

"The collaboration with the Data Science Hub was excellent. They were engaged, knowledgeable, and always willing to think along with us, providing valuable support across all aspects of the project"

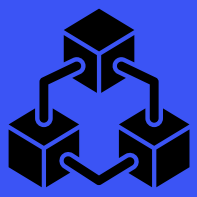
-FS, LSTT project

2025 in numbers

5
Open Source Lunches organized

11
data science projects finalized

9
collaborations with different divisions



9.6
overall client satisfaction

220+
unique participants that attended our activities

Goals set for 2025

Expert

Goals	Measures
At least 20% of the projects have an impact on DNB's working methods by changing work processes	1 50% of the finalized projects has commits (coding) of the business
A minimum of 80% of our activities are focused on "big picture" initiatives, encompassing high-level and innovative data science projects	2 100% of our major projects comply with the Big Picture and 3B policy

Process

Goals	Measures
All our project comply to the 3B policy.	3 All projects have completed the checklist
At least 20% of our projects are reusable internally	4 All projects are included in the 3B register according to the decision tree
Our infrastructure is aligned with the ideas described in the Big Picture to have impactful data science within DNB	5 40% of onboarded projects have the potential to be of (in)direct relevance for other departments
	6 At least one project with a scalable compute

Partners

Goals	Measures
A minimum of two-thirds of our projects are initiated based on demand.	7 250 unique participants in a DSH activity
	8 Five divisions with a major project
We stimulate a data science community	9 One project involves collaboration with a newly engaged data-intensive department (such as BCM)
We communicate our proceedings so that our colleagues know where to find us for a data science project	10 The average evaluation score is at least 8

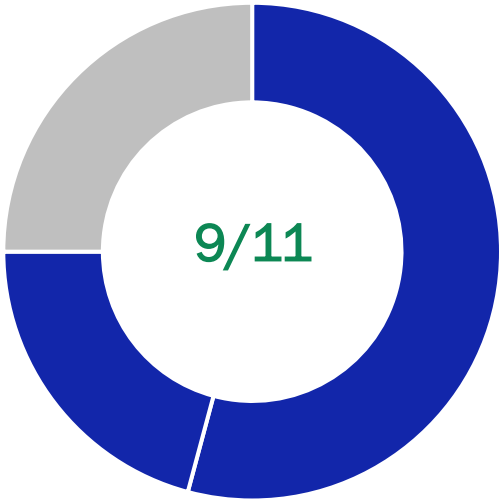
Team

Goals	Measures
We build on our current knowledge	11 80% of the relevant topics are studied and applied
We want to positively influence relevant internal parties and well-functioning DASE service offering	12 The average grade for the input we provided to the DSAP is at least an 8

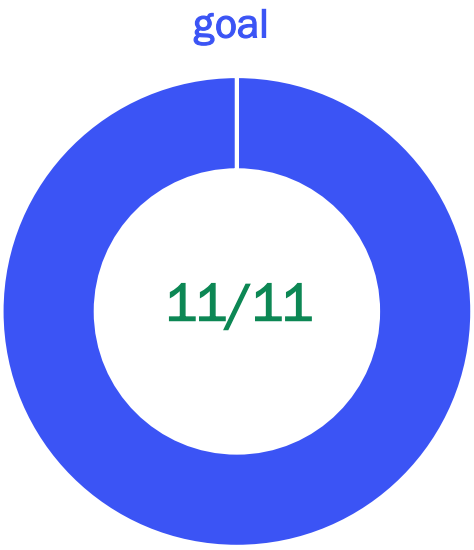
DataScience Hub

Results

1 50% of the finalized projects has commits (coding) of the business

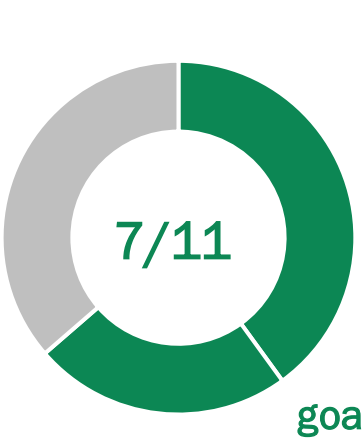


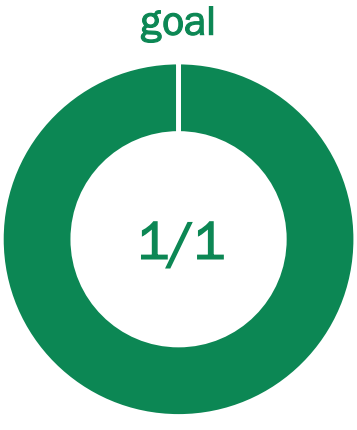
2 100% of our major projects comply with the Big Picture and 3B policy



All projects have completed the checklist

All of our projects are in the 3B project register





3 All projects have completed the checklist

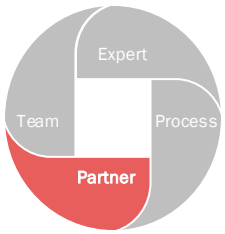
4 All projects are included in the 3B register according to the decision tree

5 40% of onboarded projects have the potential to be of (in)direct relevance for other departments

6 At least one project with a scalable compute

DataScience Hub

Results

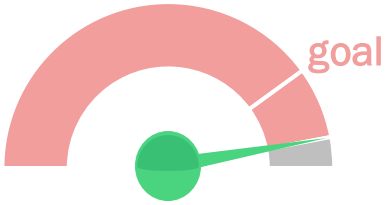


Over 225 unique participants attended activities organized by the DSH

SF
STAT
TV
SFO

We have worked with BCM-MI/BCM-CDS

The average grade for our projects is a 9.4



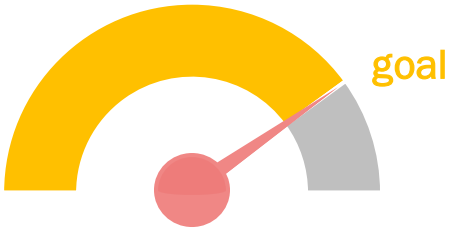
8 250+ unique participants attended activities organized by the DSH

9 At least 5 different divisions with a big project

10 One project involves collaboration with a newly engaged data-intensive department

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

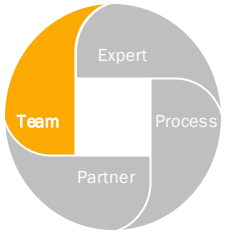
Our colleagues have actively followed relevant trainings



The grade for our input is an 8.0

11 80% of the relevant topics are studied and applied

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



DataScience Hub

Finalized projects

SYN

DNB's SYN project tackles the challenge of realistic test data in development environments. It offers a reusable module to generate synthetic data—artificial yet structurally accurate and privacy-safe—improving testing, accelerating development, and supporting data quality. The first use case focused on the CBS counterparty dataset.

Click on this polygon to see our live projects dashboard

SFODT2

The project advanced toward an MVP with a reusable architecture and a new web app. BIS Innovation Hub now leads the next phase—Project Danu—to build a full end-to-end solution.

NOWC

To improve the timeliness of CO_2 emission indicators, a forecasting model was developed to nowcast data that typically arrives with significant delays. The project delivered an optimized model and a Python package, 'Oraculo,' enabling data preparation, forecasting, and evaluation. This approach supports faster, more accurate sustainability indicators for policy decisions despite challenges like limited historical data and pandemic effects.

NUMI

BEVO-DNS oversees a numismatic collection of 400,000 items with images stored on a network disk. To streamline frequent image requests, the project automates exports given article numbers, the system retrieves corresponding images and generates a zip file, saving significant time and effort.

STADT

We explored using Digital Twin for geospatial sustainability data visualization. The pilot showed strong potential but highlighted the need for a stable environment. With Databricks now available, development can continue a production-ready platform.

Finalized projects

LSTT

FS-MAB modernized its liquidity stress testing process, replacing a manual, time-intensive workflow with an automated framework. The new system reduces inefficiencies, minimizes human error, and enables rapid adaptation to evolving stress testing requirements.

Click on this polygon to see our live projects dashboard

ODM

Since 2020, Dutch pension funds have reported detailed investment data, improved transparency but posing quality challenges due to inexperience and scale—over 1 million lines quarterly. This project applies machine learning-based outlier detection to enhance accuracy and reduce manual effort, with further optimization through reinforcement learning in Dataloop.

FSTR3

Building on the FSTR2 ETL pipeline for SFTR repo data, the FSTR3 project developed a real-time stress testing model in Virtual Lab to monitor systemic risks in the repo market. This graph-based approach simulates liquidity flows and supports DNB's supervisory role, addressing IMF recommendations for enhanced repo market oversight and stress testing for pension funds.

FICORM

FM_RM currently monitors daily price changes in MAPS for equities, fixed income, and other market data using fixed thresholds. This project enhances the process by introducing advanced outlier detection methods that analyze time series and product categories, improving accuracy and reducing false alerts.

FICO3

Within ECTB, a new outlier detection tool based on the FICO package has replaced the previous quarterly process. It delivers more accurate results and eliminates unnecessary follow-ups. The tool also improves usability by providing visualizations and scores through a Power BI dashboard, reducing manual work for analysts.

Internal outreach

Open Source lunches

We organised several Open Source Lunches in 2025; knowledge sharing sessions where one topic or project is the center. Each year we aim for 5 lunches and this year we have reached that goal once more. The presentation topics ranged from how we can build IT solutions according to the 3B policy to the implementation of Synthetic Data for testing within DNB.

We would like to thank everyone whom attended and presented this year. As always we hope to host many more of these sessions in the next year and if you have any projects or topics you would like to present, please don't hesitate to reach out!

(data_science@dnb.nl)

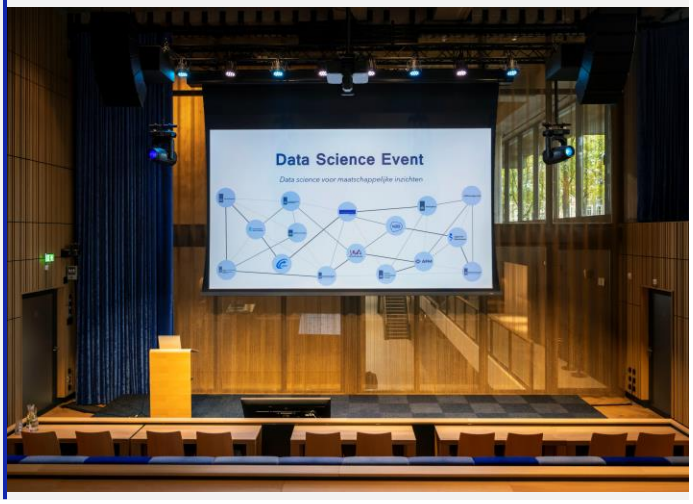
Open Source workshops

The DSH participates in enabling colleagues to keep up with their data skills and explore new ones by organising a variety of workshops. In collaboration with the DNB academy we offer three workshops: Clean & Responsible Coding, Explainable AI, and, the most popular one, Version Control with Git.

This year the workshops continue to be a great success; participation and feedback have been very positive. This enthusiasm and demand suggests that we keep hosting these workshops in the future.

The interest in AI and Data across DNB is growing rapidly, thus, in further collaboration with the AI Programme and Data Programme, we introduced a new training: *Data and AI Impact* - designed for colleagues unfamiliar with data or AI but want to work with it. Additionally, we contributed to the BCB training, where we emphasize how to operate as a data-driven organisation.

Looking ahead, the DSH will continue to deliver and expand data-related workshops to meet the increasing demand and ensure DNB colleagues are Data-focussed.



External outreach

This year, we continued to focus on connecting with colleagues and partners around the world for external collaboration and the international exchange of expertise. Our work on synthetic data received a lot of interest across various institutions, including the Bank of Finland, BIS Innovation Network's SupTech group, the Federal Reserve, and the Bank of England.
We also spoke about our Digital Twin and Interconnectedness projects with a wide range of partners both in the Netherlands and abroad, such as Rijkswaterstaat, the European Central Bank (ECB), the European Systemic Risk Board (ESRB), and the European Banking Authority (EBA).
In addition, we connected with colleagues from the National Bank of Slovakia (NBS) and the National Bank of Ukraine (NBU), allowing us to further broaden our network and share our experiences.
And throughout the year, Iman van Lelyveld regularly represented our team at various "Data Science and AI in Finance" conferences. As a speaker and panel member, he showcased our work and contributed to discussions across the European central-banking community.

Trip to Brussels

In June of 2025, the DSH had the privilege to visit the Data Science teams of the National Bank of Belgium (NBB) and Financial Services and Markets Authority (FSMA) in Brussels. During the two-day visit, we exchanged ongoing projects and discussing the role of data science in central banking and supervisory authority.
Several topics have been discussed with both organisations, at the NBB some included outlier detection and a tool for seasonal adjustment in time series data: JDemetra+.
During our visit with the FSMA, they presented a new approach to processing Key Information Documents, by replacing the traditional regex-method with LLMs.
If this visit has taught us anything, it is the importance of sharing (data science) knowledge with other organization and expand the Data Science Community internationally. It is what drives us to new insights and innovations. Many thanks to our Belgian colleagues and we look forward to more opportunities like this.

Data Science Event

This year marked the third edition of our annual data science event, organized by and for a diverse community of data scientists working within the Dutch public sector. The event aims to facilitate the exchange of experiences and knowledge while strengthening collaboration across institutions. This edition brought together over 130 participants from more than 15 organizations, including AFM, RIVM, and the Belastingdienst.
The program featured two keynote presentations and a series of contributions from participants, delivered through presentations and interactive workshops on a wide range of data science topics. The opening keynote by Lokke Moerel offered a compelling perspective on both the opportunities and challenges surrounding AI. This seamlessly connected to the afternoon keynote by Jesse van Oort (GTP-NL), which focused on developing a Dutch language model.

In addition, 11 institutions hosted (interactive) sessions that provided valuable insights into ongoing projects and innovative approaches. CPB, NVWA, and ILT sharing their views on the responsible use of algorithms in high-risk applications. Belastingdienst presenting an agent-based model to simulate taxpayer compliance behavior. DNB and NAz outlining the steps to implement their own GenAI applications. UVW discussing techniques for synthesizing registration data. AI4Oversight showcasing responsible AI applications in supervision.
Participants greatly appreciated the opportunity to network and learn about initiatives across organizations. The event continues to serve as a platform for fostering collaboration among data scientists in the public domain.

DeNederlandscheBank
EUROSYSTEEM

DataScience
Hub