

ANNUAL REPORT 2023

DNB Data Science Hub

DeNederlandscheBank

EUROSYSTEEM



DataScience
Hub

Foreword

I am very happy to present the latest Data Science Hub Annual Report to you. 2023 has flown by, and looking back fills me with pride. We have completed a wide range of projects across DNB with very happy clients: nine out of ten stars is a score that humbles us and which we hope to live up to in the coming year. The projects are also becoming more mature as the organization becomes more data driven.

With our other activities, we try to enthuse the wider data science community. The Open Source Lunches are still a popular platform to showcase interesting projects attracting 121 unique visitors. To improve the knowledge base, we ran several workshops with in total 132 participants. Externally, we have organized a workshop with the Netherlands Bureau for Economic Policy Analysis (CPB) for data scientists in government. That was a fantastic afternoon that should lead to a repeat event.

Further afield, we had bilateral discussions with several central banks and supervisors. We are all looking at the same issues, so the aim is to identify a topic that is of shared interest and then see if we can develop something together. An example of this is a Digital Twin for physical climate risk where we work together with the Banque de France and the Hong Kong Monetary Authority as part of a BIS Innovation Network workstream.

Data science is a group effort, so none of this would have been possible without collaborating with other related departments such as our colleagues in IT, the Data Office, the supervision's Chief Innovation Office and all the data scientists across the bank. And, of course, the DSH team including interns and visiting researchers.

Looking ahead, there are several major hurdles that we want to tackle in the coming year. We have developed a coherent view on what structural improvements we need to realize to be able to deliver more value. For example, we need to make scalable compute available in order to run complex models and allow us to implement cutting-edge solutions. But, much simpler, we also need to streamline information and training for starting colleagues. They often still find it daunting to navigate the bank's systems.

More than enough excitement to look forward to. Hope to meet up at some point and see if we can travel the data science road together.

Iman van Lelyveld



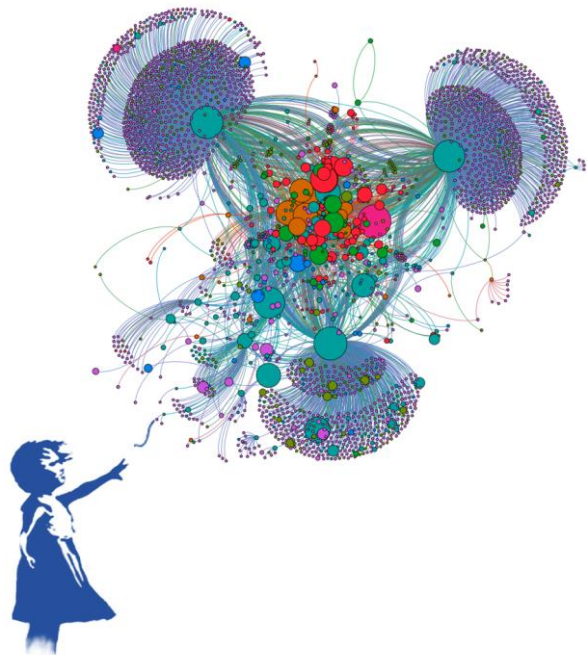
DataScience Hub

What do we do?

The DSH aims to offer advice, guidance and execution of data science projects for the entire DNB organization. It is the hub in a hub-and-spoke model. Specifically, the DSH:

- builds data science solutions with all departments, proactively searches for data opportunities and proposes new 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 2023.



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

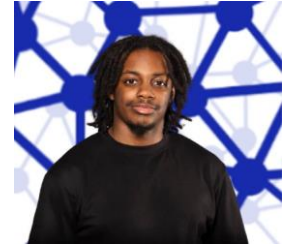
Our team in 2023



Abel Koch



Aniek Sweijen



Bryan van Axel Dongen



Iman van Lelyveld



Karo Oude Groote
Beverborg



Kristy Jansen



Marinthe Straathof



Merel Sliepenbeek



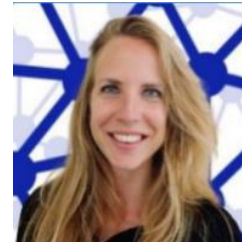
Michiel Nijhuis



Milan Karsten



Natalie Kessler



Patty Duijm



Paun Poonpakdee



Robert Hofman



Tim Haarman



Zoey Bossert

2023 in quotes

“I am pleased with the collaboration with DSH, as they really helped us throughout the entire project with their knowledge, insights and contacts within DNB. I think it saved us quite some time to have someone from DSH involved.”
- Resolution, ECCP project

“Great cooperation, good suggestions. We have been well involved from the start. Blockages were well discussed. Devised beautiful and working solutions for new challenges. Useful discussions had. We have learned a lot from the DSH people.”
- EBO, NIPE project

“It has been a great pleasure working with the DSH. They know what they are doing and are able to figure new stuff out very quickly. Especially the ‘table-issue’ we had took some stamina as the solution required a lot of trial and error.”
-BVM, CCAP II project

“DSH helped us very timely and efficiently with retrieving data on climate risks from GIS maps. The collaboration was very pleasant, and the result was exactly what we aimed for.”
- EBO, BWGIS project

“I really enjoyed the collaboration. I always learn a lot when doing a project with you. I like the informal and flexible way of working, and the quick responses. For this project I was able to do the bulk of the work myself but having someone to spar with whenever running into difficulties, or to discuss the bigger picture of the project which is really helpful.”
- Resolution, DGF project

“The collaboration was very good. There were clear agreements and I think we made good progress during this project.”
- CIOT, TOP project



2023 in numbers



13

data science projects
finalized

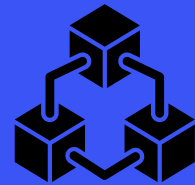
5

Open Source
Lunches organized



10

collaborations
with different
divisions



9.1

overall client
satisfaction

9.4

average rating on our
Open Source
Workshops



240+

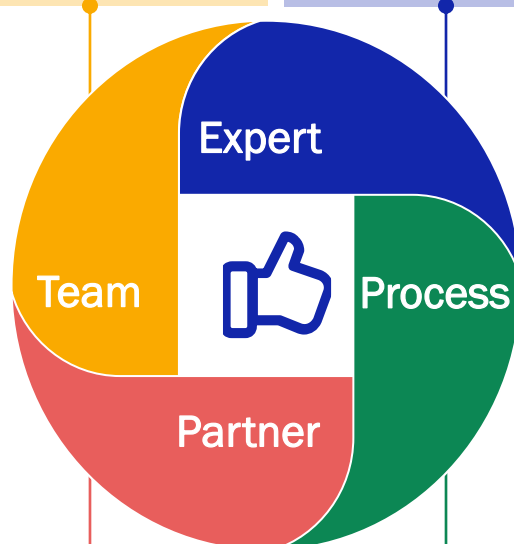
unique
participants that
attended our
activities



Goals set for 2023

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

Goals	Measures
Broaden the knowledge of data work at DNB by educating clients in projects	<p>1 All data science projects take into account the DSH manifest</p>
Providing reusable solutions to clients as stated in the manifest	<p>2 25% of the finalized projects has commits (coding) of the business</p>
	<p>3 At least 20% projects lead to a follow-up project</p>



Goals	Measures
We communicate our proceedings so that our colleagues know where to find us for a data science project	<p>6 At least 8 different divisions with a project</p>
We stimulate a data science community	<p>7 The average grade for our work on projects is at least an 8</p>
	<p>8 100+ unique participants attended activities organized by the DSH</p>

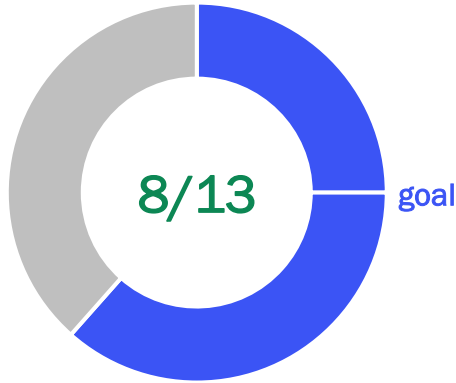
Goals	Measures
We aim for wide usability and the relevance of external stakeholders within our whole working process	<p>4 At least 20% of our finalized projects are (in)directly relevant for another business unit</p>
	<p>5 At least 50% of our (finalized) projects are shared externally</p>

Goals set for 2023

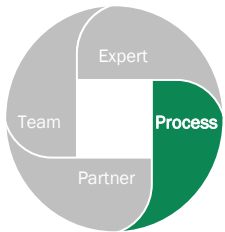
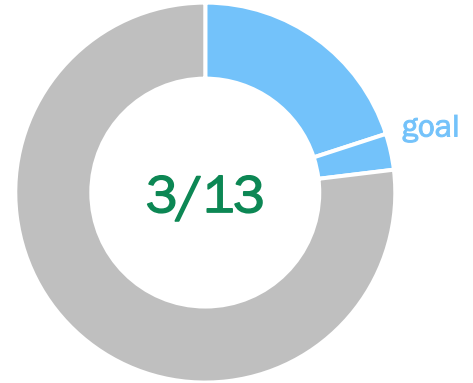
1 All data science projects take into account the DSH manifest



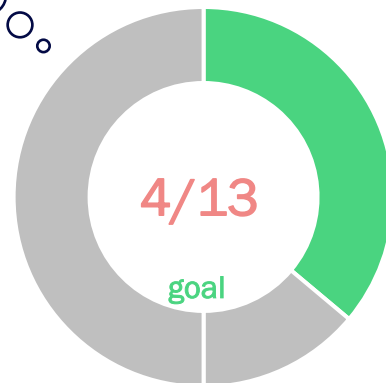
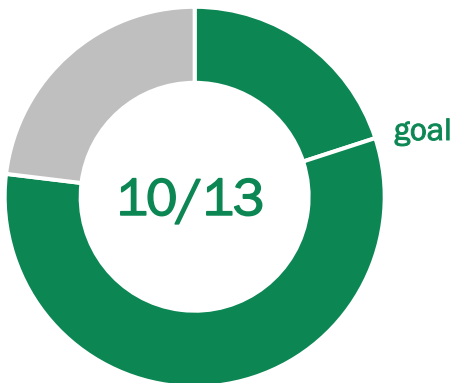
2 25% of the finalized projects has commits (coding) of the business



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



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

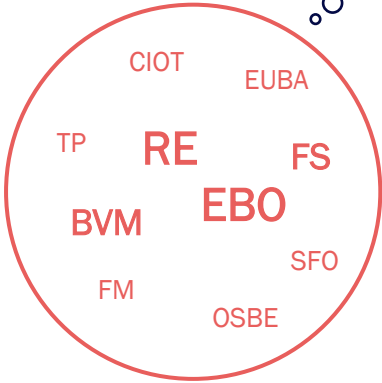
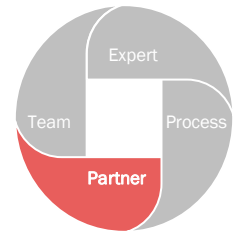


4 At least 20% of our finalized projects are (in)directly relevant for another business unit

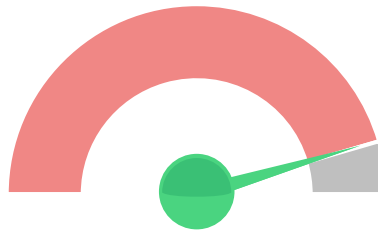
5 At least 50% of our finalized projects are shared externally

Goals set for 2023

We finalized projects with **10 different divisions**



The average grade for our projects is a 9.1



Over **240** unique participants attended activities organized by the DSH



6 At least 8 different divisions with a project

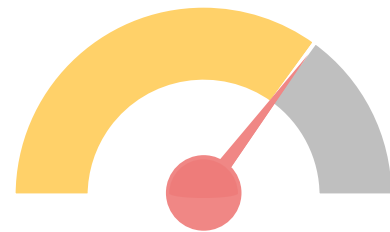
7 The average grade for our work on projects is at least an 8

8 100+ unique participants attended activities organized by the DSH

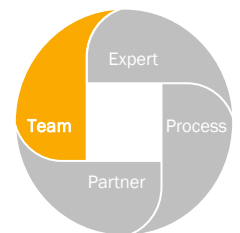
9 50% of the relevant topics are studied and applied

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 a 7.0



Finalized projects

NIPE

NIPE is a model that is used to periodically make inflation projections. To run this model, a large amount of data from external sources are needed, which must be updated every time the model is reevaluated (4-6 times a year). Updating this data was done manually, which is a time consuming and non-tractable process.

This project modernized this process by automating the data collection in the Azure cloud, aiming to provide the possibility for modelers to request a data-refresh with the press of a button. This model currently runs in production and is maintained by the DSH.

ERI

This project set out to map the degree of exposure to climate risks at financial institutions through ESG-risk indicators. Using ESG databases, four environmental risk indicators are calculated and projected in the ESG dashboard, so that supervisors can get insights.

FSTR

A project with FS to analyze what type of information one could obtain from the transactional repo data of SFTR .It is found that there are many intermediaries involved, and strong seasonality patterns appear.

CCAP2

This project investigated the feasibility of identifying and evaluating criteria in credit claim documents. Applying NLP applications using Pytesseract, criteria were efficiently extracted, and the usability of the framework is displayed in a dashboard.

IMAS

Findings from inspections are stored in the IMAS system, these findings consist of descriptions. By analyzing all the descriptions of the findings for a single bank these root causes can be uncovered from the findings, which is now done in a manual fashion.

Using data science techniques this process could be automated and the root cause analysis could be performed periodically, giving the most common topics in the findings. The subjects were extracted with the use of Topic Modelling, giving probability distribution functions over a set of words and the distribution of the topics over the whole set of documents.

ECCP

A project using EuroCCP transaction data to automatically update in a CCP dashboard. Two milestones are reached: a working BVM dashboard, and with an infrastructure set out on Azure.

Finalized projects

TMAR

A project with the goal to gather daily analyst reports, identify topics, and archive them in a database. Due to legal constraints and new Bloomberg functionalities, the project was concluded.

BWGIS

This project aimed to enrich CBS microdata on Dutch households with climate risk indicators. Geographic information for all dwellings in the Netherlands was obtained from the BAG dataset to match on a household level and added to CBS microdata.

DGF Tool

A project with the goal to make an app to pilot how to work with a git branching strategy, ADO pipelines, infrastructure as code, and performing code reviews. The goal succeeded: the app can obtain data from the datalake in a secured manner.

TOP

This project aims to find topics within the meetings minutes of committees of banks for supervision. Using a LDA framework, an unsupervised ML algorithm, on TONE data, it is found that it is indeed possible to find topics within meeting minutes.

Interconnectedness

In the recent past, we have seen that several non-bank financial institutions can expose banks to serious losses. Currently, it is impossible to readily observe the Dutch banking sector's (in)direct exposures to a (number of) defaulting NBFI(s).

In this project, we link big datasets, on loans, derivatives, money market products, and securities holdings, and identify interconnections between traditional banks and NBFIs. We apply a number of (name) matching, cleaning, and aggregation steps, to end up with a network of connected entities and their related exposures.

IKB

This project focused on performing automated checks on the validity of the confidential IKB data, with the addition of outlier detection. A functioning connection to the data and a set of default checks ensure that the data is in perfect shape when resolution is needed.

Internal outreach

Open Source workshops

At the DSH we organize workshops to help you increase your data science knowledge. In collaboration with the DNB academy, we provide three different workshops: Version control with GIT, Clean & Responsible Coding and Explainable AI. Also, we gave a workshop on the basics of Data Science, called Data Science 101.

We organized over ten workshops in 2023 with our standard Git Workshop clearly the most popular. On request, more specific training courses were also provided. And we were, of course, also present at the DNB's Digital Festival to provide training. Overall, the workshops were very popular, reaching more than 85 colleagues and scoring, for instance, our Git Workshop with an average of 9.4!

We recently started working with the DNB academy to expand our training courses among DNB colleagues. Next year, we'll continue the workshops via the DNB academy on a structural basis (GIT once per quarter) to meet demand.

Digital@DNB Festival

Wherever the word digital is in the name, that's where we need to be. So, as expected, we were also present at the well-attended DNB Digital Festival held last October.

We were actively involved in the preparation of the event and were present at the digital plaza with our own stand. A great opportunity for everyone to find out what we do, and a good opportunity for us to show what we do. Lastly, we held two workshops: Explainable AI and a Data Party.

Data Party

Since this year, we started organizing Data Parties for new employees. In this workshop, hosted together with the Data Balie of the Data Office, is intended for new employees to give them an insight into all the different teams that work with data within DNB. In an interactive way, colleagues get to know the Data Office, Data Balie and the Data Science Hub better. The Data Party is organized once every 3 months.

But there are also options for colleagues who are already working. For example, a data party was organized during the digital festival, and you can join a future data party. All you have to do is show your interest by sending us a mail (data_science@dnb.nl)

Open Source lunches

In 2023 we were able to organize the Open Source Lunches as they are supposed to be: in the office with a real lunch. There have been a total of 5 Open Source Lunches in 2023 with a large variety in topics presented. From topic modelling to R packages and Azure projects, we have seen it all.

More than 150 of our DNB colleagues joined an Open Source Lunche. It is great to see the interest of other data scientists and data driven colleagues! 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

BISIH EDKP Open Tech-community Workshop

In November, the DSH organized a workshop based on our Digital Twin project for the BIS Innovation Hub EDKP Open Tech community. Michiel Nijhuis started the workshop with a deep dive on the Digital Twin project, a project on the impact of physical climate risks on the financial system. The second half of the workshop we invited NVIDIA to share their experience on working with digital twin technology.

The workshop inspired the community to think about their use cases for the Digital Twin project, to implement during the hackathon mid-May 2024.

Data Science events

During the year, we organized two Data Science events (DSE) in collaboration with other Dutch (semi-) public institutions. The first DSE was held in the beginning of 2023 together with Data Scientists of the CPB to exchange experiences. As this was found to be so useful, another DSE was held later in the year, this time with more organizations. This time, with the addition of people from the Ministry of Finance and CBS, several workshops and a panel discussion were held. There will undoubtedly be more to follow next year.



Way-of-Working paper

The DSH has now matured from a start-up to a regular department. As you can read elsewhere in this report, we have made good progress on many fronts. However, along the way, we also encountered growing pains and had to overcome hurdles. To document the various projects and, more importantly, to draw lessons that might be useful to others, Iman van Lelyveld and Patty Duijm wrote a paper presented at [a conference](#) and which will be included in an e-book published by the European University Institute. One key lesson is that, in addition to the carrot that the DSH offers in terms of free consultancy, there should be a stick where the business is incentivized to change. This is a delicate balance, because we are not aiming for change for changes sake: only when new techniques improve existing processes are they useful. And this is a priori not clear and hence difficult to steer.



DeNederlandscheBank
EUROSYSTEEM



DataScience
Hub