ANNUAL REPORT 2022

—

DNB Data Science Hub

DataScience
Hub

DeNederlandscheBank
EUROSYSTEEM

Contents

Introduction

It is with great pleasure that I present to you the second annual report of DNB's Data Science Hub. In this report we showcase the steps we have taken to become an even more data driven central bank and supervisor. I'm very proud to see how much we have been able to accomplish together with other departments, with the IT division, with Chief Innovation Office Supervision (CIOT) and with our team.

Our mission is to become the first place you turn to if you want to get help in starting a data science project at De Nederlandsche Bank. To accomplish this mission we initiate a wide range of activities. For one, we do projects together with other departments and you will see a selection presented here. We also have a limited number of projects where we try to deliver public goods -- think for instance of the DNBDatafetcher package that makes accessing external data so much easier. In addition we have organized several events to bring to together the community of data scientist. Examples are the Open Source Lunches, the Open Source Workshops, Open Source Expert Groups and the Data Party for people entering the bank. To foster interaction with academics we organized a data science conference. We also contribute to the policies governing data and code use. For instance, we are pushing to make code shareable as widely as possible while adequately tackling confidentiality and security issues. Finally we are actively seeking cooperation with other central banks and supervisors.

How well are we doing? Well, according to the people we work with, not bad. For all our activities we measure the baseline and how happy the clients are once the project has been finished. The average grade for our projects is 8.9. Our workshops receive a 9.5 on average. These are numbers that make us blush and we hope to be able to live up to the expectations in the time to come. Looking ahead, I see a year with many opportunities for us. Needless to say there are hurdles as well but nobody ever said that it would be easy. In the end, I'm confident we will be able to push the Data Science Manifest, in particular Rule Number #10: celebrate the completion of the projects!

Iman van Lelyveld

Our team in 2022

Bernard vd Boom

Zooey Bossert
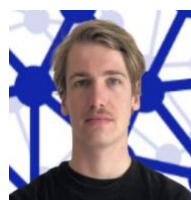
Daniel Christiano

Patty Duijm

Tim Haarman

Ronald Heijmans

Jasper Hoek

Robert Hofman

Justus Inhoffen

Natalie Kessler

Abel Koch

Nana Lange

Iman van Lelyveld

Michiel Nijhuis

Gert Rietveld

Raphael Schmidt

Kai Schellekens

Aniek Sweijen

Hidde van der Zalm

DataScience Hub

Our year in numbers

13
data science projects finalized

8
collaborations with different divisions

8.9
overall client satisfaction

4
Open Source Lunches organized

9.5
average rating on our Open Source Workshops

150
unique participants that attended our activities

DataScience Hub

Quotes from our clients

"The collaboration was excellent. Work progressed without having to closely monitor it, and the DSH itself took initiative to improve the solution"

-Sjoerd van der Zwaag

"The DSH colleagues were professional, helpful and very pleasant to work with. The research was well executed despite some setbacks and provides real added value to our department."

-Jorrit van den Broek

"Good collaboration with the Data Science Hub. Very knowledgeable staff, and pragmatic and efficient communication and division of workload"

-Sjors Altemühl

"DSH was very quick with their responses and very helpful during the project. We experienced many changes in the MAPS project, diverse obstacles and had to deal with multiple external factors. However DSH was very flexible and quickly adapted."

-Marleen Schlooz

"The Data Science Hub makes room in their calendars to finish the project on time"
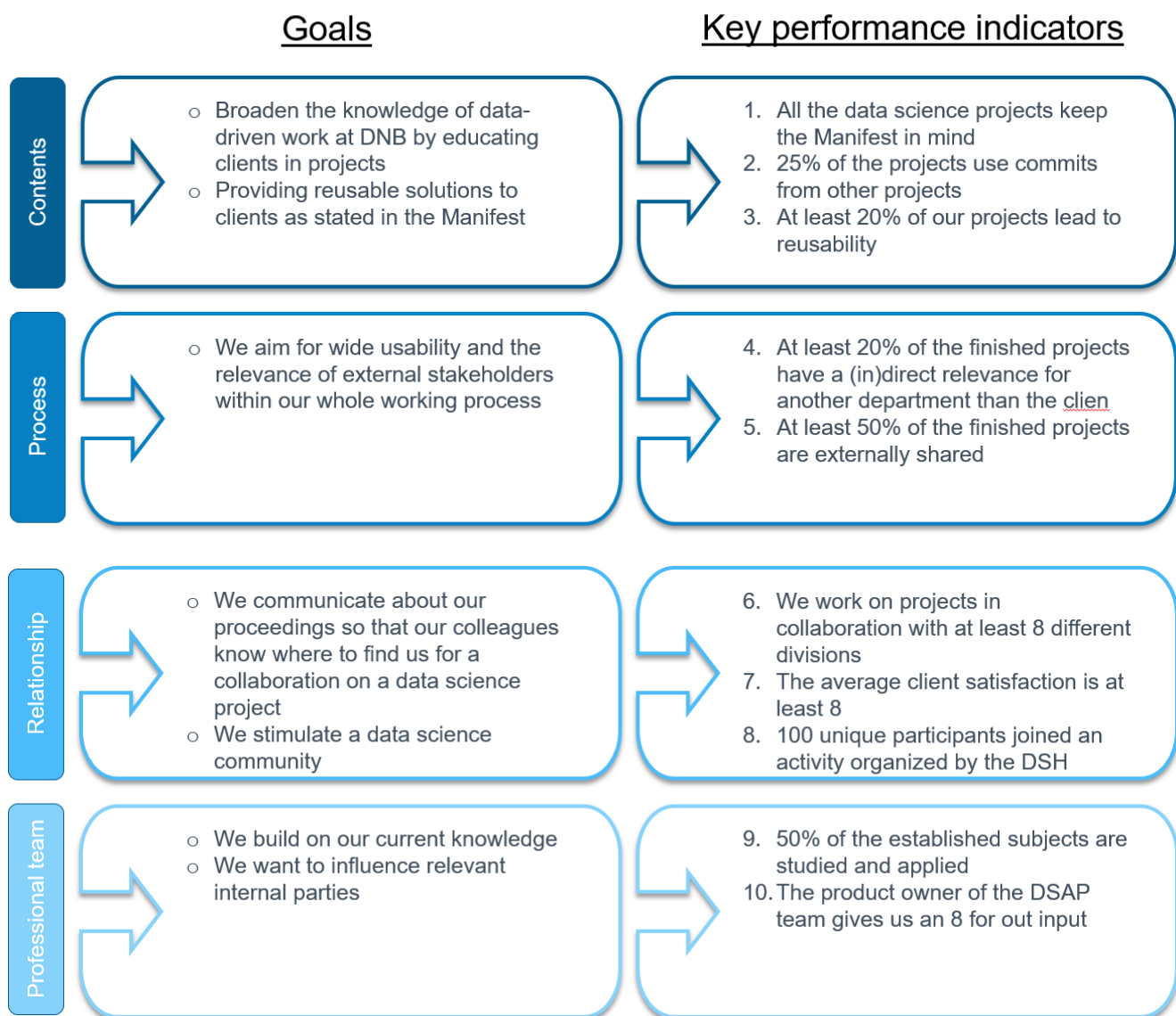
-Chris Gossink

"I had a very nice experience with the Data Science Hub and I hope we will have the opportunity to work together again in the future"

-Sari Butink

DataScience Hub

Goals set at start of 2022

The DSH takes care of advising, supporting and executing data science projects

Goals

Contents

- o Broaden the knowledge of data-driven work at DNB by educating clients in projects
- o Providing reusable solutions to clients as stated in the Manifest

Process

- o We aim for wide usability and the relevance of external stakeholders within our whole working process

Relationship

- o We communicate about our proceedings so that our colleagues know where to find us for a collaboration on a data science project
- o We stimulate a data science community

Professional team

- o We build on our current knowledge
- o We want to influence relevant internal parties

Key performance indicators

1. All the data science projects keep the Manifest in mind
2. 25% of the projects use commits from other projects
3. At least 20% of our projects lead to reusability

4. At least 20% of the finished projects have a (in)direct relevance for another department than the clien
5. At least 50% of the finished projects are externally shared

6. We work on projects in collaboration with at least 8 different divisions
7. The average client satisfaction is at least 8
8. 100 unique participants joined an activity organized by the DSH

9. 50% of the established subjects are studied and applied
10. The product owner of the DSAP team gives us an 8 for out input

Projects

Key Performance Indicators

1. All the data science projects keep the Manifest in mind ✓

2. 25% of the projects use commits from other projects ✓

3. At least 20% of our projects lead to reusability ✓

In 2022, the Data Science Hub supported 13 data science projects throughout the bank

>>> See the section 'Finalized Projects' for more information on each project

1 LIST OF FINALIZED PROJECTS

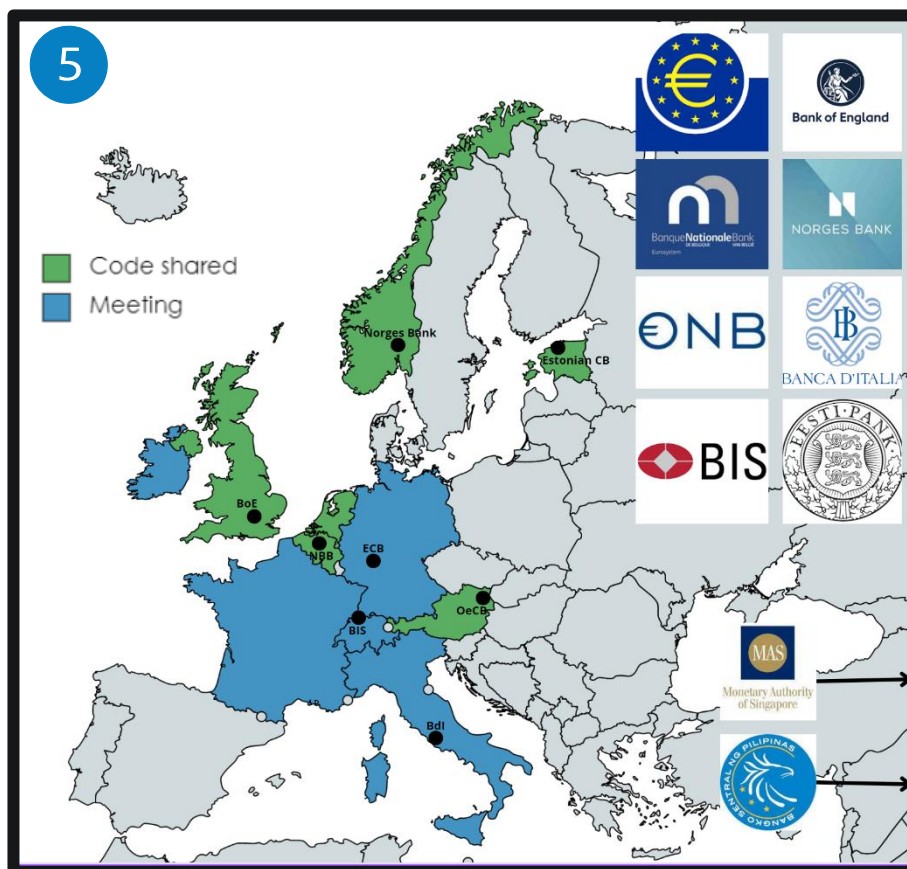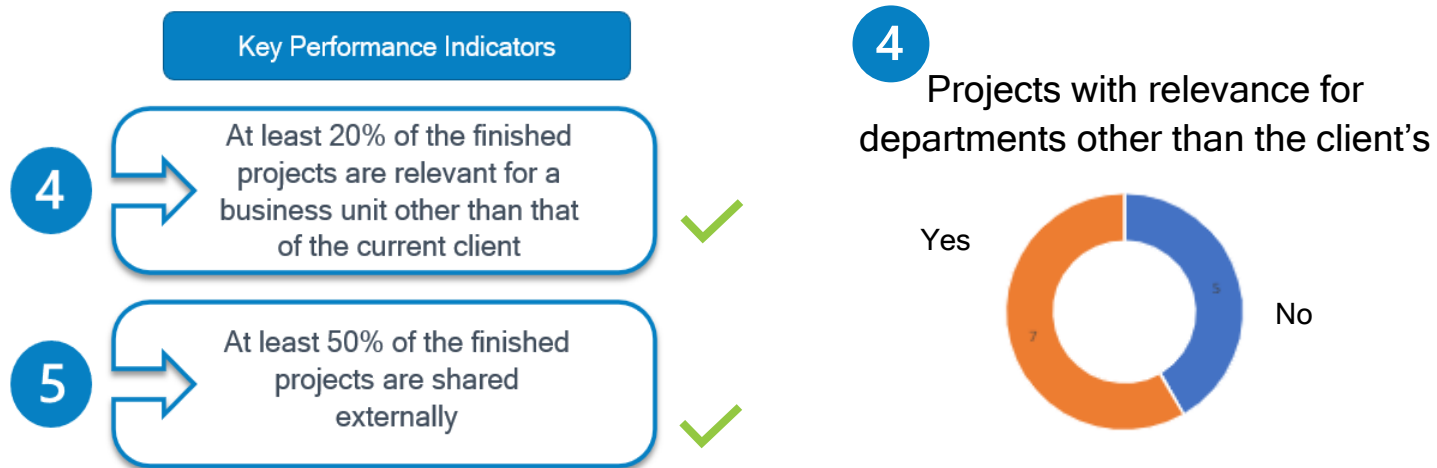|	
ERA	Employee Resignation Analysis
KYC3	Know Your Customer 3
FMDWH	FM-RM Temp Warehouse
IRCRE	Integrity risks in commercial real estate
FETCH	Dnbdatafetcher
VGB	Differences analysis
DELFI	Automate DELFI model
CoBo	Detecting and checking covered bonds document
SFODT	Digital twin for physical climate change
DLML	Dataloop ML
PACTA2	Pacta for banks
ACFS	AnaCredit for FS
M2TC	MMSR to the cloud

2 30% of our finalized projects have commits from others

3 23% of our finalized projects are used for our current projects

➤ See our dashboard here

DataScience Hub

Annual Report 2022

Process

Key Performance Indicators

4 → At least 20% of the finished projects are relevant for a business unit other than that of the current client ✓

5 → At least 50% of the finished projects are shared externally ✓

4

Projects with relevance for departments other than the client's

Yes

No

7   5

5

Code shared
Meeting

ECB
Bank of England
Banque Nationale Bank
NORGES BANK
ÖNB
BANCA D'ITALIA
BIS
EESTI PANK
MAS Monetary Authority of Singapore
BANGKO SENTRAL NG PILIPINAS

Norges Bank
Estonian CB
BoE
NBB
BIS
OeCB
BdI

This year we talked to *more than 80 people* outside of the Dutch Central Bank that were interested in Data Science Hub activities

DataScience Hub

Relationship

Key Performance Indicators

6 → We work on projects in collaboration with at least 8 different divisions ✓

7 → The average client satisfaction is at least 8 ✓

8 → 100 unique participants joined an activity organized by the DSH ✓

At the end of a project, a survey is send to the clients to ask for feedback on various aspects.

>>> The results in this report are based on 13 surveys for the finalized projects

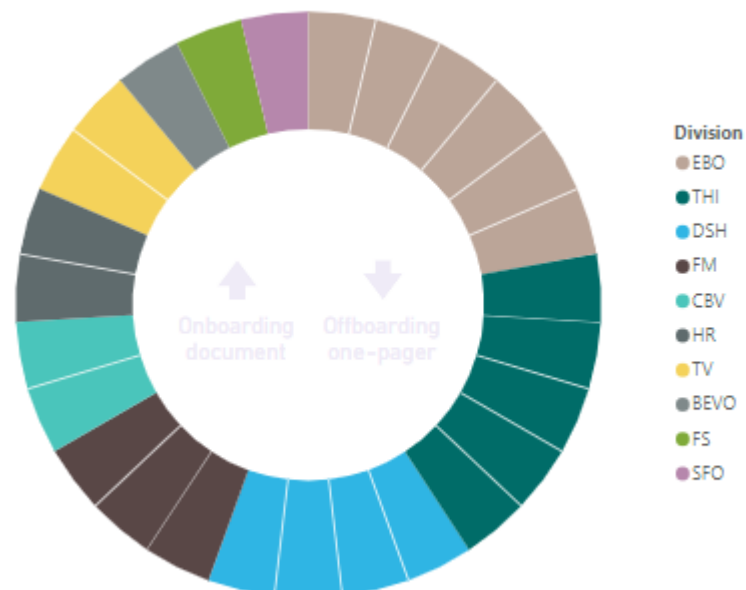**7** How do you judge the collaboration with the Data Science hub in general?

★ ★ ★ ★ ★ ★ ★ ★ ★ ☆ → 8.93 average score

6 Projects by division

8 How many unique participants attended the event?

Division
- EBO
- THI
- DSH
- FM
- CBV
- HR
- TV
- BEVO
- FS
- SFO

Onboarding document   Offboarding one-pager

DataScience Hub

Finalized Projects

Know Your Customer 3

Follow-up of the first two Know Your Customer projects, again applying outlier detection to select clients for a deep-dive integrity investigation. In previous projects we created an outlier detection model to automatically make a selection of potentially interesting cases. This worked well for private clients, given that their expected behavior is relatively homogeneous. During this follow-up project we adapted and extended this model to a wholesale-client portfolio, to see if we could also apply our approach to a smaller set of more diverse large corporate clients.

Employee Resignation Analysis

In cooperation with Human Resources we investigated which variables correlate with the decision of employees to leave DNB. Variables that were found to be correlated are age, function type, relative salary position, temporary contract, and years of service. The graph, for example, shows that the percentages of employees which are inactive (i.e. they resigned) in age group 2 and 3 is higher than average.
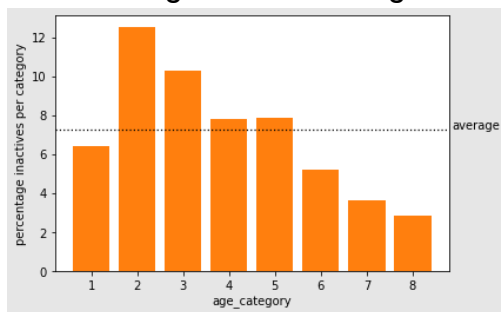
1: <30   **2**: 30-34   **3**: 35-39   **4**: 40-44
5: 45-49 **6**: 50-54 **7**: 55-59   **8**: 60+

Automate DELFI

The econometric model department of EBO (ECMO) planned to automate the data collection of their inflation forecasting model, DELFI. To do this the dnbdatafetcher package, developed by the DSH, was used. The automation of this was done by ECMO. The DSH gave advice on how to best implement the datafetcher in the project, and added a new source to the datafetcher (i.e. OECD) that was needed for the model.

Credit analysis

Several indicators from the AnaCredit data were extracted. With these indicators a better view of the state of the Dutch lending market has been established. The datapoints extracted for the credit monitor can also be used as a first step into working with the AnaCredit data. The indicators are also chosen to show a broad range of the data available in AnaCredit. They show information on the counterparties, the instruments, the protections and the chance of default of the credit issued in the Netherlands.

Differences analysis

DNB receives many parcels with banknotes. Sometimes there is a difference between the value of the banknotes the DNB counts and the value that is indicated on the parcel. In the project, we tried to see if we could identify what the causes of these differences are. As the data is noisy, we build ML models to predict the size of the differences so as to analyze which variables are most important for the prediction model. In this way we were able to draw conclusions if certain differences were extraordinary. By building these ML models, we could also flag multiple data issues, potentially leading to a better data quality and decisions.

Integrity Risks in Commercial Real Estate

The goal of the project was twofold. The first goal was to identify whether the CRE dataset (commercial real estate data) can be used for the identification of integrity risks or subversive activities. The second goal was to find out whether the CRE dataset can be used in relation to sanctions legislation

The result of the project is a report that summarizes the findings. This report can serve as an introduction to some aspects of the CRE dataset and can help colleagues decide whether the CRE dataset is useful for their project/analysis. In addition, certain data quality checks have been performed.

Dataloop ML

In the Dataloop ML project we developed an outlier detection algorithm for the Solvency-II and FTK line-by-line asset data. The outlier detection algorithm find outliers within these datasets on top of the business rules implemented and can thus improve data quality. The parameters of the outlier detection model are adjusted based on the feedback from the institutions about the datapoints which are flagged by the algorithm.

The original goal of the project was to implement this outlier detection in Dataloop. Within the context of this project that has not been achieved, however a new project has been started to implement the outlier detection algorithm into production.

Datafetcher

The dnbdatafetcher project is a project initiated by the Data Science Hub. Having done several automation projects, it became clear that many colleagues need data from mostly the same external institutions. Instead of writing ad-hoc code for all these projects to collect data, providing general to do this would be much more efficient. The goal of the dnbdatafetcher project was to provide this functionality by means of a Python package.

The package is currently available to everyone withing DNB and supports easy data collection with Python from CBS, ECB, Eurostat, BIS, Refinitiv Datastream, OECD, and the World Bank.

MMSR to the cloud

MMSR to the cloud (M2TC) is a project where we have realized a daily pipeline to fill a database with the complete European Money Market Statistical Reporting (MMSR) dataset. This allows for more efficient and more detailed analysis of the European money market. In an earlier project we already developed dashboarding. However, in that project the data was loaded into the database by hand. Therefore, this project has resulted in the capability to monitor liquidity-stress of institutions, market segments or the market as a whole practically real time. This is an important capability for monetary policy implementation and micro- and macroprudential policy.

Digital twin for physical climate risks

In the digital twin project, we developed a MVP of a digital twin for the calculation of the effects of climate events on the financial system. In a BIS Innovation Hub project we worked together with other central banks to ensure the MVP is sufficiently generic to be able to handle mutiple different physical risks and that it could be applied to different jurisdictions. The MVP has been hosted as an Azure webapp as this enables easy acces.

Covered Bonds

This project aimed to extract to-be-checked values from standardized investor reports that supervisors receive from various financial institutions. We found that with the use of the Python package Tabula, it was relatively easy to extract the target values. To check if institutions are compliant, we also implemented an automated check, based on a set of defined criteria. With this project we showcased how standardized reporting can reduce the number of repetitive tasks and contributes to a more efficient way of working.
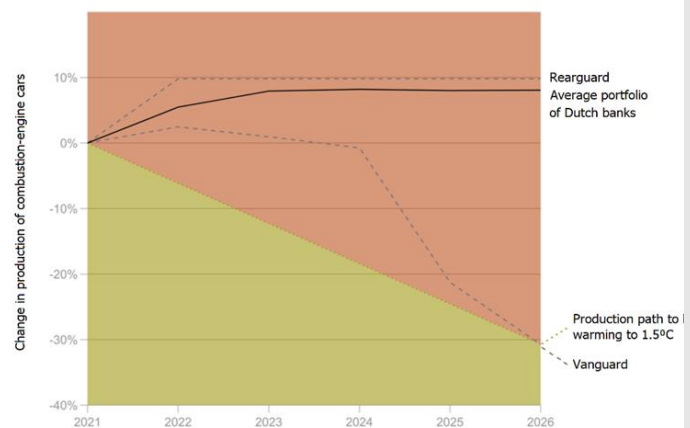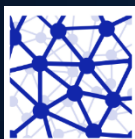
Temporary data warehouse

Transition periods when moving from one to another data platform can be difficult to manage. For one of the divisions, this was also the case when moving to a new data warehouse, but where they still wanted to generate daily pricing computations.

Therefore the DSH built a tool that automatically processes this data, generates the required output, and stores it into the risk management database, until the new system goes live.

PACTA for banks

We conclude that the companies in the most carbon-intensive sectors in the loan portfolios of Dutch banks are not transitioning quickly enough. The analysis reveals that car manufacturers are not switching to electric cars rapidly enough to meet the Paris climate goals. The switch from fossil-fuel-based power generation to renewable sources is also too slow. In addition, oil extraction by companies in the loan portfolios of Dutch banks will increase in the coming years. However, in the natural gas production sector on average the companies are adapting quicker than the transition scenario calls for.