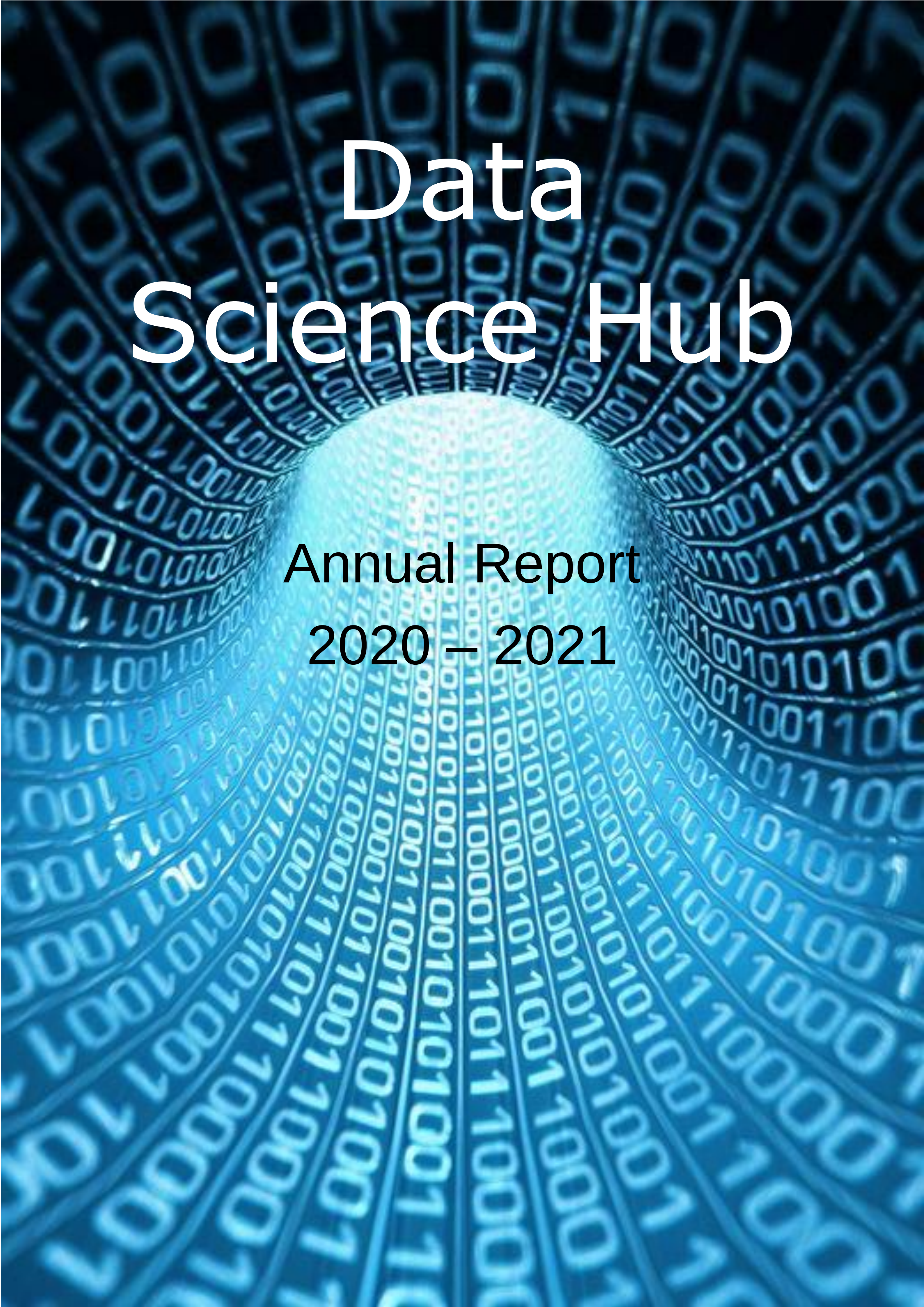


Data Science Hub

Annual Report 2020 – 2021

Contents

Want to stay updated throughout the year? Subscribe to our monthly newsletter by sending an email to data_science@dnb.nl

Our team

Iman van Lelyveld

Hub Head Iman: "*In 2021 the Data Science Hub (DSH) has come up to speed, completing 29 projects with departments all across DNB. We have, for example, been instrumental in bringing innovative applications that improve the reporting process of regulated firms to the cloud and have worked on efficiency improving machine learning models for banknote checks. We are fostering a thriving community of data scientists through lunch meetings, workshops and hackathons.*

None of this would have been possible without the outstanding DSH team, interns and visiting PhDs I present to you here. Also, a great thanks to our clients, IT, data owners and all others involved. It is a quite a journey we are on. Let's go forward, enjoy the journey and travel towards a data driven future."

Bernard vd Boom

Gert Rietveld

Zooey Bossert

Gregory Aerts

Justus Inhoffen

Nirmal Muppiri

Dieter Wang

Michiel Nijhuis

Nana Lange

Kristy Jansen

Vidushi Anand

Ronald Heijmans

Kai Schellekens

Samet Kütük

Patty Duijm

Tim Haarman

Natalie Kessler

Artjom Janssen

Our year in numbers*

29

data science projects finalized

17

colllaborations with 17 divisions

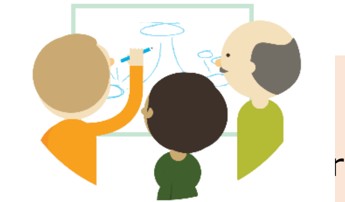
8.5

overall client satisfaction

9

open source lunches organized

6

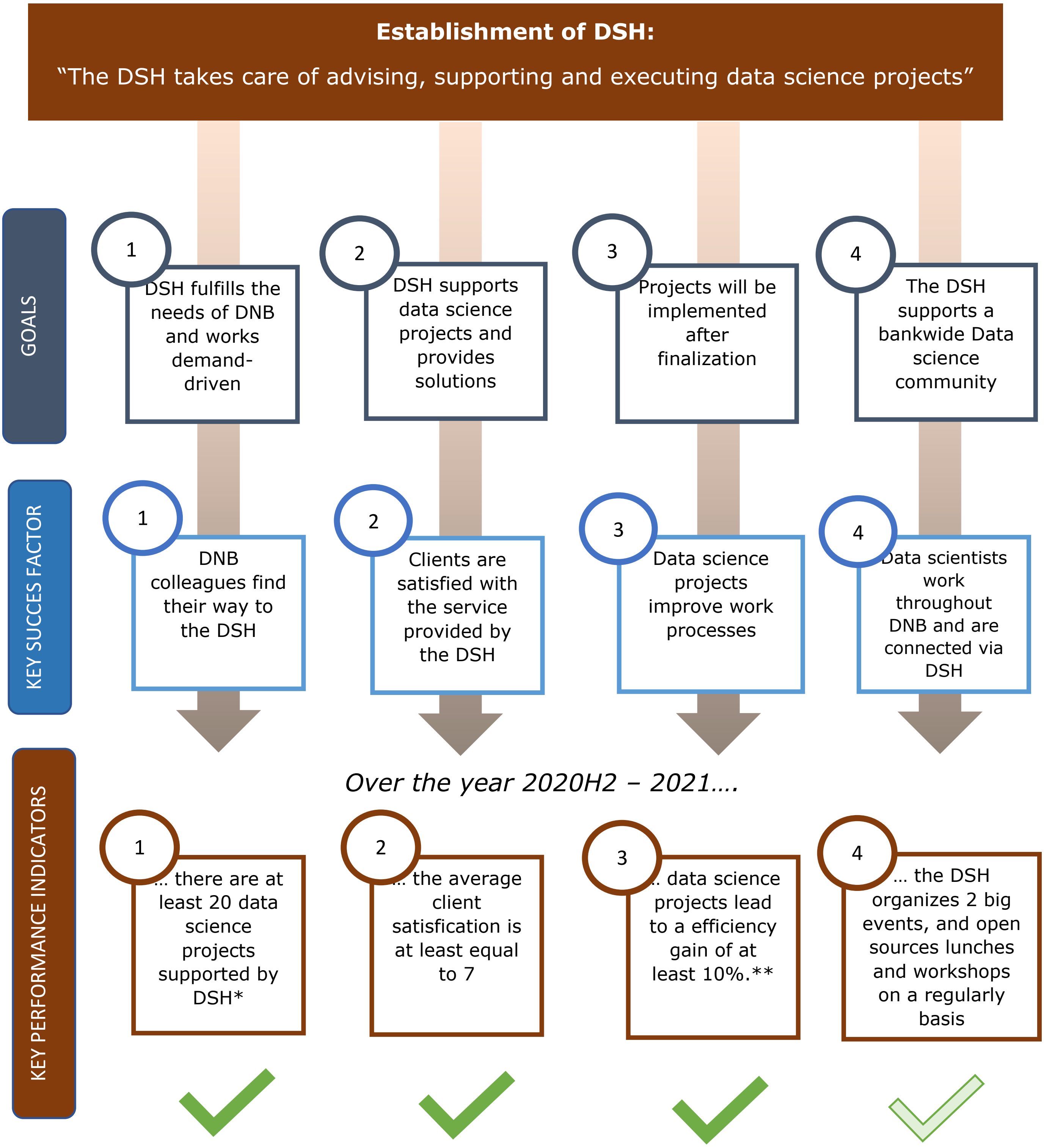


8.6

overall sc rce workshops

*The DSH was established 1 July 2020. The numbers in this report are therefore based on the period 2020H2 - 2021

Goals set at start



Establishment of DSH:

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

	1	2	3	4
GOALS	DSH fulfills the needs of DNB and works demand-driven	DSH supports data science projects and provides solutions	Projects will be implemented after finalization	The DSH supports a bankwide Data science community
KEY SUCCES FACTOR	DNB colleagues find their way to the DSH	Clients are satisfied with the service provided by the DSH	Data science projects improve work processes	Data scientists work throughout DNB and are connected via DSH
KEY PERFORMANCE INDICATORS	… there are at least 20 data science projects supported by DSH*	… the average client satisfication is at least equal to 7	… data science projects lead to a efficiency gain of at least 10%.**	… the DSH organizes 2 big events, and open sources lunches and workshops on a regularly basis
	✓	✓	✓	✓

Over the year 2020H2 – 2021….

*A finalized project is defined as a project i) that is originated from a request by the business (another division or department within DNB) for support by the DSH; ii) that has passed all the phases of the DSH onboarding procedure, including a completed intake form; and iii) where the business with the support of the DSH has worked on pre-set goals.

** Not all projects strive for efficiency. Risk-identification or new insights could also be a goal of a DSH projects. Projects without the efficiency goal are out of scope for this KPI.

KPI #1: Projects

KPI

1

.. there are at least 20 data science projects supported by DSH

✓

From the start, the Data Science Hub supported **29 data science** project throughout the bank.

>>> See the overview at the end of the report for more information on each project.

LIST OF PROJECTS

1	ELA	Emergency Liquidity Assistance
2	BIO	Biodiversity
3	HFI	Google trends
4	STC	Contagion stress test
5	HRM	HR Mail traffic
6	SCC	FMI supply chain concentration
7	C19L	Covid19-Look through
8	MMSR	MMSR dashboard
9	KYC	Know Your Customer
10	CJB	ConjunctuurBoek
11	LDR	LDR interconnectedness visualisation
12	XBRL	XBRL in Neo4j
13	SBO	Small Banks Outliers
14	RCFP	Relative Carbon Footprints
15	T2O	TARGET2 outlier detection
16	DL2C	Dataloop to the cloud
17	KYC2	Know Your Customer 2
18	AZADV	Azure Advisor API
19	FUB	False Unfit Biljetten
20	REIR	Integrity Risks in Real Estate
21	CRE	Commercial Real Estate
22	GAs	Crypto arbitrage using genetic algorithms
23	TMTP	Text mining pensioenfondsen
24	CSA	Challenge of Self Assesments
25	SIMM	Standard Initial Margin Model
26	DFROG	Automatiseren DFROG
27	CCAP 1	Credit claim acceptance project
28	RIAD	Joining data
29	WEBSTATAPI	DNB Website API

KPI #2: Client satisfaction

KPI

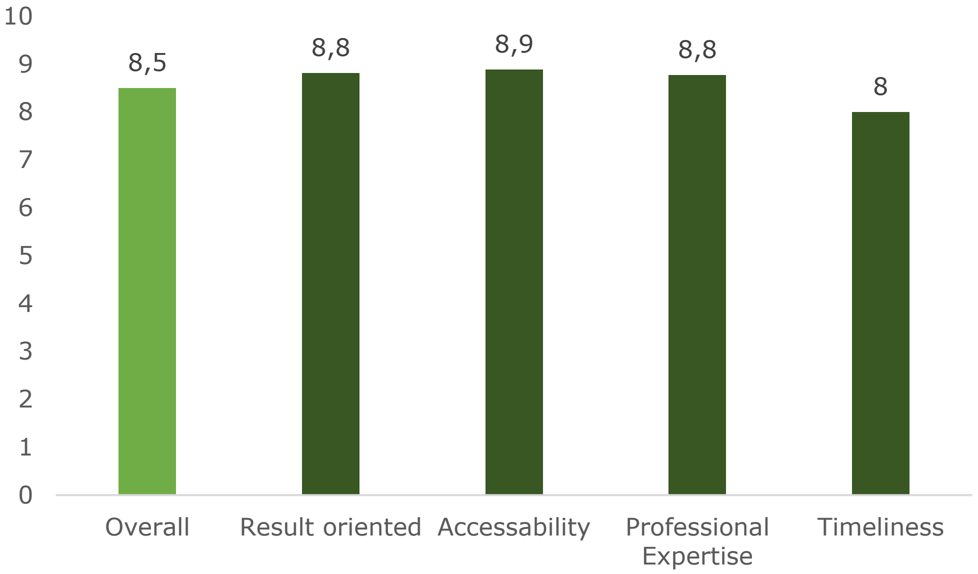
2

… the average client satisfication is at least equal to 7

✓

The overall client satisfaction is a **8.5**

>>> See the next page for quotes by a sample of customers.



At the end of a project, a survey is send to the client to ask for feedback on various aspects. The results in this report are based on 26 surveys (for the 29 projects)

Quotes from our clients

“This was a pleasant collaboration. The DSH provided the technical expertise that we lack, while we could provide financial market insights. I think this project was a good example of combining the 'technical' expertise from data scientists with the more 'practical' expertise from policy makers. If we get the dashboard up and running as discussed and use it for daily monitoring purposes, this will clearly be a useful monitoring tool that helps us in monitoring money markets more efficiently than we do currently.”

Tom Hudepohl

“Collaboration went smooth and the member of the DSH was very experienced. We were able to quickly translate our research questions into algorithms that we tested and adjusted.”

Richard Heuver

“Willing to understand us, and trying to support us in wanting to get more insight in collected data.In the end the offered tools and possible output was not good enough because input was not structured enough. We learned from this and will improve our next data collection.”

Rob Wassink

“It was a pleasure! Hope to be able to work more often with you!”

Joris van Toor

“It was very useful and informative; it is a professionally operating team with whom it was easy to have a fruitful cooperation.”

Richard Derksen

“It was great to have someone who can fully focus on the data side. I experienced an open dialogue on how certain ideas/issues could be entertained/solved. In addition, it was very helpful that we were able to make the translation from the data/analysis to key messages that were ultimately shared with senior management.”

Jeroen Huiting

KPI #3: Efficiency gains

KPI

3

… data science projects lead to a efficiency gain of at least 10%.

✓

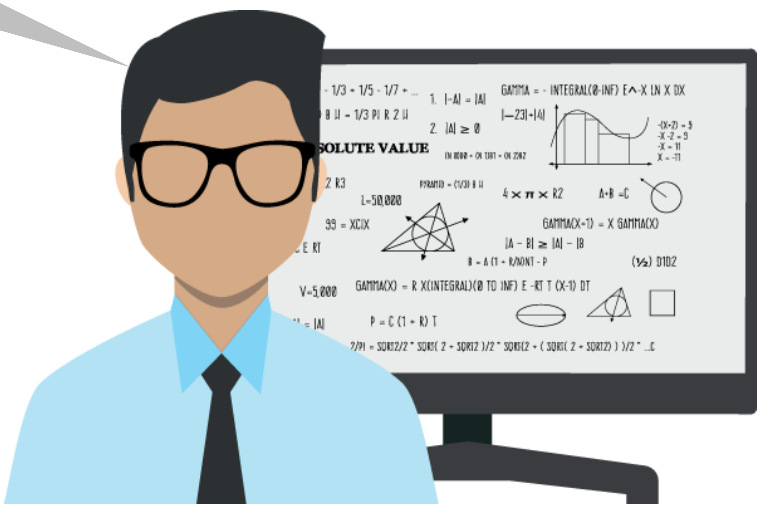
The average efficiency gain equals **50%**

This number is based on 7 projects that had efficiency gains as the main goal of the project. The numbers range from 15% to 80%.

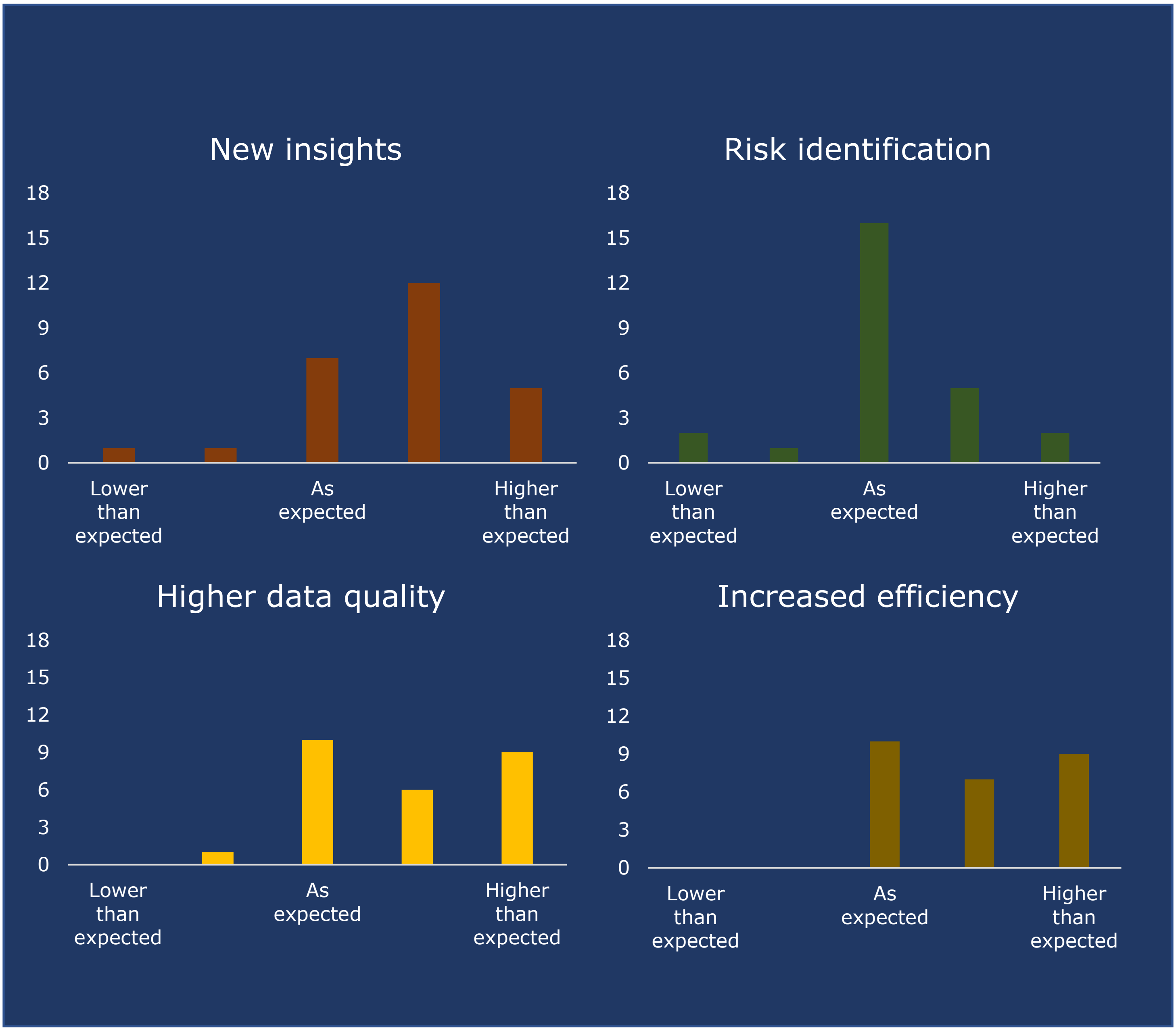
With the onboarding of a project, clients are asked to set a main goal. This can be i) increased efficiency; ii) risk identification; iii) new insights or iv) higher data quality. At the end of the project they are asked:

1. *To provide an estimate of the realized efficiency gains*
2. *To indicate whether the added value of the project based on all of these 4 goals was lower or higher than expected at the start*

see next page



The added values of a project turn out to be often higher than expected at the start of a project, according to our clients. The vertical axis represents the number of projects.



KPI #4: Community activities

KPI

4

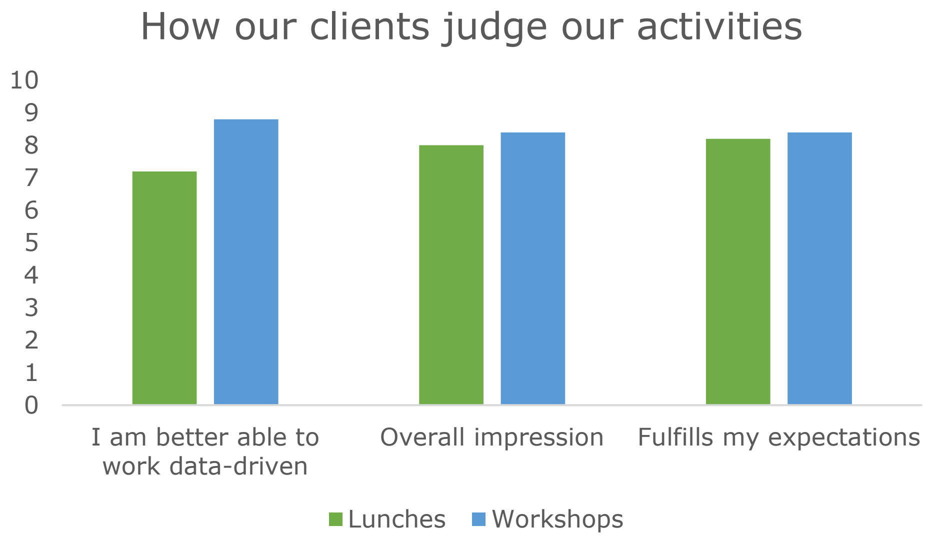
… the DSH organizes 2 big events, and open sources lunches and workshops on a regularly basis

✓

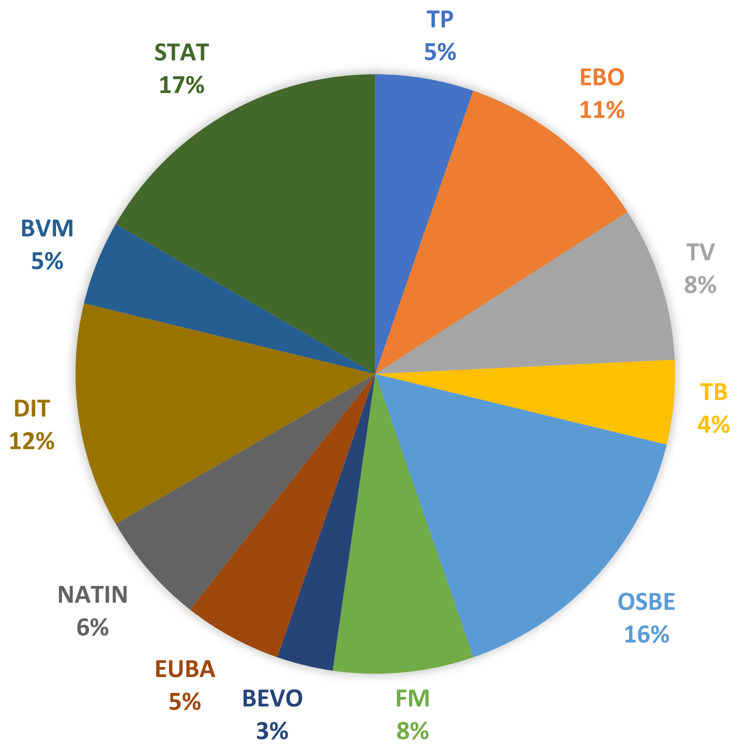
Due to Covid-19, the DSH was not able to organize to big events. These are however scheduled for next year! But.. Despite Covid-19 the DSH was able to organize **9 Open Source Lunches** and **6 Open Source Workshops**! On top op that, the DSH organized a hackathon, presented during training weeks of other divisions and has a slot in the introduction programme for new DNB employees.

Picture taken during the DSH hackathon 2021

```
if date == datetime.date(2021, 6, 25) :
    print('join dsh hackathon 2021')
elif date == datetime.date(2021, 7, 2) :
    print('join dsh hackathon 2021')
else :
    print('eat pizza')
```



Participants of the Open Source Workshops represent a bank-wide community. The figure shows the participation rate by DNB division.

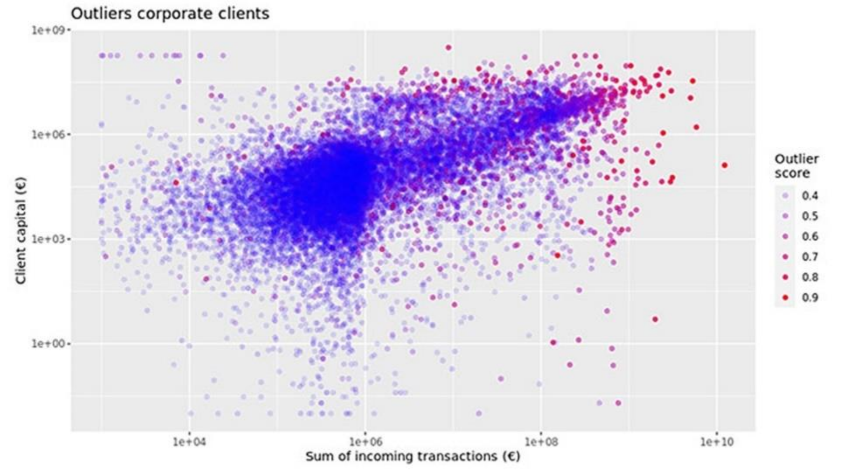


Finalized projects

Know Your Customer

Financial institutions act as gatekeepers. They must prevent criminal money from going through the financial system. To do so, they must apply the know-your-customer principle. But how does a bank monitor millions of transactions a day, and how does DNB supervise those processes? This involves data science. Smart technologies allow us to spot potentially high-risk transactions in a single large database. Our **anomaly detection model** that makes **outliers stand out from the rest helps supervisors** focus their examinations.

The chart on the right plots corporate bank customers, their wealth and the sum of incoming transactions. Each dot in the plot represents a customer. If the algorithm considers customers unusual it marks them red.



In collaboration with Supervision Horizontal Functions and Integrity.

Azure Advisor API

With this project, **a history of Azure Advisor recommendations** was constructed to get an insight in how people make use of the cloud, particularly in terms of security and costs. The development of these recommendations over time can provide an insight into the risks and other insights.

In collaboration with Data- and Information-Technology.

TARGET2 Outlier Detection

The main objective of this project was to **extend the plain autoencoder with memory (GRU and LSTM) models** to increase the performance of the neural network and to provide feedback to improve the data science and analytics platform (DSAP). The data was not rich enough to train an autoencoder with memory and therefore it is not possible to add such a feature to the model.

In collaboration with Payments and Market Infrastructures and Payments Canada.

Contagion Stress Test

Based on the common asset holdings of Dutch banks, we are interested in how exposed these banks are to other major Eurozone banks. We created a **network graph to visualize exposure and impact** and wrote a note to inform board member discussion in a Corona policy group.

pay

In collaboration with Financial Stability.

Relative Carbon Footprints

We proposed **adjustments to the often used Weighted Average Carbon Index** (WACI). We show that adjusting relative carbon disclosure metrics for inflation and exchange rate fluctuations makes a significant difference to the level and dynamics of these metrics over time.

The results of this project are published in an Occasional Study. As a next step, the study is enhanced and will be submitted to an academic journal (target: Nature Climate Change).

In collaboration with Statistics.

Financial Market Infrastructure Supply Chain Concentration

The aim of this project was to investigate whether financial institutions (banks, pension funds, insurances) are exposed to undetected risks through reliance on similar software and/or technologies. Therefore, a **visualization of how different financial institutions are indirectly linked through different technologies** is created.

In collaboration with Payments and Market Infrastructures.

Google Trends

Given the potential added value of the Google Trends data, we created an **easy accessible environment** (database) of this data at DNB. This gave rise to a better understanding of the data and address search limits.

In collaboration with Econometrics & Models.

Business Cycle Insights

The main objective of this project was to **automate the data collection** for the business cycle insights booklet. This booklet contains a snapshot of the Dutch economy.

In collaboration with Economic Policy.

XBRL in Neo4j

During this proof of concept, we investigated the possibility to **put XBRL data in a graph data base**. Based on the experience with the graph, more use-cases for a graph database might be identified.

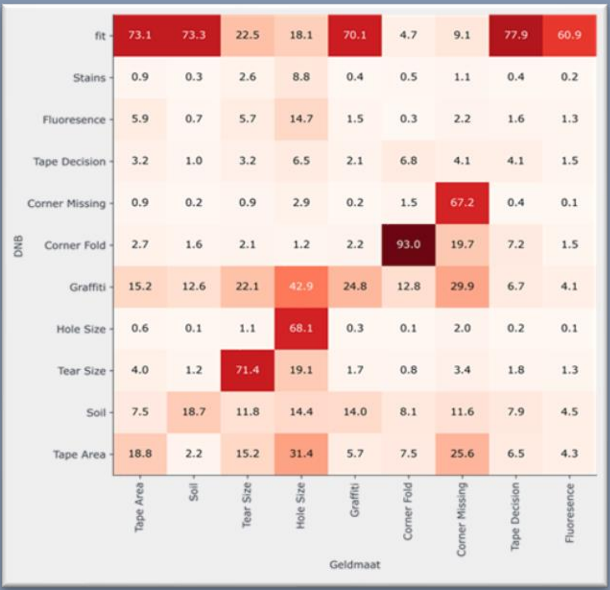
In collaboration with Supervision Insurers and Chief Innovation Office Supervision.

False Unfit Banknotes

DNB receives unfit banknotes from market parties that have been deemed no longer fit for circulation by Geldmaat's counting centers. These unfit banknotes are checked again at DNB because DNB has specific authentication sensors to determine whether a banknote is counterfeit. During the sorting process at DNB, it appears that a large percentage of these unfit banknotes are still evaluated as being fit. Apparently these banknotes are false unfit.

With this project, we determine **the cause of the high percentage of false unfit banknotes** and investigate how this percentage could be decreased. By looking at matched banknotes, it can be seen where the classification differs between DNB and Geldmaat and pinpoint specific rules which do not add up. If everything was going perfectly the diagonal of the matrix on the right-hand side would be filled with dark squares, as the Geldmaat trigger would be the same as the DNB trigger. The number of fit classifications for DNB if Geldmaat detected a problem shows the extend of the false unfit problem. Only the hole and tear size and the corner defects are often both triggered by Geldmaat and DNB for the same banknote.

In collaboration with Cash Operations.



	Tape Area	Soil	Tear Size	Hole Size	Graffiti	Corner Fold	Corner Missing	Tape Decision	Fluoresence
fit	73.1	73.3	22.5	18.1	70.1	4.7	9.1	77.9	60.9
Stains	0.9	0.3	2.6	8.8	0.4	0.5	1.1	0.4	0.2
Fluoresence	5.9	0.7	5.7	14.7	1.5	0.3	2.2	1.6	1.3
Tape Decision	3.2	1.0	3.2	6.5	2.1	6.8	4.1	4.1	1.5
Corner Missing	0.9	0.2	0.9	2.9	0.2	1.5	67.2	0.4	0.1
Corner Fold	2.7	1.6	2.1	1.2	2.2	93.0	19.7	7.2	1.5
Graffiti	15.2	12.6	22.1	42.9	24.8	12.8	29.9	6.7	4.1
Hole Size	0.6	0.1	1.1	68.1	0.3	0.1	2.0	0.2	0.1
Tear Size	4.0	1.2	71.4	19.1	1.7	0.8	3.4	1.8	1.3
Soil	7.5	18.7	11.8	14.4	14.0	8.1	11.6	7.9	4.5
Tape Area	18.8	2.2	15.2	31.4	5.7	7.5	25.6	6.5	4.3

Covid-19 Look Through

For this project we first of all combined and processed the data sources needed for the calculation of the effects of the different measures which have been taken to alleviate the stress on the banks after the Covid-19 pandemic began. Additionally, we provided assistance for this analysis.

In collaboration with On-site Supervision and Banking Expertise

Small Bank Outliers

The main objective of the project was to single out outliers in credit risk data. The thresholds were first decided by the business. After working with the data, the thresholds were decided by calculating the materiality percentage based on corelated attributes.

In collaboration with National Institutions.

Commercial Real Estate (CRE)

We improved the **current accuracy of the model used for predicting the defaults in the CRE** data set via normalization, transformation and parameter creation.

In collaboration with Chief Innovation Office Supervision.

Human Resources E-mail Traffic

With this project more insight is gained in the impact of Covid-19 and working from home. This is done by **exploring any trends in the number of (internal) mails sent within DNB**.

In collaboration with Human Resources.

Dataloop

We have developed the Dataloop application to improve data quality in supervisory reporting. Dataloop does this by centralising and visualising data from different sources. It also offers several feedback loops, for example between analysts and machine learning tools. In addition, such loops will soon be available between DNB and financial institutions and other supervisory authorities.
Dataloop achieves 20% efficiency gains in compiling statistics. Also, machine learning allows supervisors to detect new patterns, which enables them to focus their efforts.

In collaboration with Statistics.

Emergency Liquidity Assistance

Emergency Liquidity Assistance (ELA) is often initiated at an advanced stage of liquidity problems at a bank. However, central banks benefit from early identification of an increase in liquidity problems. Therefore, for this project we aimed to obtain a **prediction model for the liquidity assistance**.
Unfortunately, the project was hampered by data availability (procedure) and a secondment of one of the collaborators.

In collaboration with Financial Markets.

Money Market Statistical Reporting (MMSR) Dashboard

MMSR is a relatively new dataset that gives DNB the opportunity to better understand the European market for short-term funds. We created a **local database to store these reports**, established an **automated data pipeline** between ECB and DNB, and created an **easily accessible dashboard**.

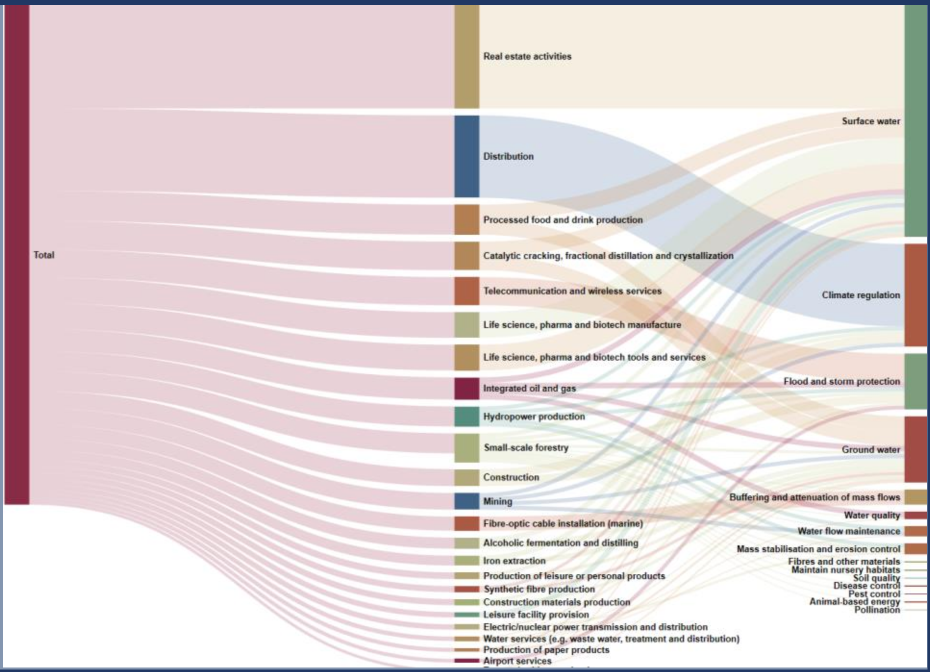
In collaboration with Financial Markets.

Biodiversity

Making our economy more sustainable is high on our agenda. We recently investigated biodiversity loss and its impact on the financial sector. We created an **interactive web application** to represent how much the Dutch financial sector is exposed to **biodiversity risks**. A screenshot of this application is given on the right-hand side and the interactive version can be found here.

We use these innovative visualizations in our Indebted to Nature study to show how different sectors depend on biodiversity.

In collaboration with Supervision Policy.



Genetic Algorithms detecting crypto arbitrage

This was an experiment to apply **genetic algorithms to crypto currency exchange data to detect arbitrage opportunities**. The data consisted of public information on crypto currency exchange rates on several different crypto exchanges. The results show that there are large arbitrage opportunities in crypto currency markets.

In collaboration with Payments and Market Infrastructures and Payments Canada.

Real Estate Integrity Risk

Two (confidential) reports were composed that provide **an overview of several characteristics of the Residential Real Estate (RRE) dataset**. The reports provide new insights for the business in terms of what the RRE dataset is and what information can be extracted from it.

In collaboration with Supervision Horizontal Functions and Integrity.

Challenge of Self-Assessments

For years, self-assessments are used as preparation for a supervisory examination at an institution. In these surveys the pension funds and insurance companies provide valuable information about the state of their institution. However, with hundreds of reporting institutions, there is not enough time to look at each assessment one-by-one. **We therefore developed a risk model to give a clearer and objective insight for the supervisors where potential risk lie**, and presented the results in an interactive dashboard to get a good overview of all the answers.

In collaboration with Supervision Pension Funds and Chief Innovation Office Supervision.

Text Mining Pension Funds

The main objective of this project is to develop a tool to **automatically extract relevant qualitative and quantitative details** from pension funds documentation. However, during this project we found that the data quality and organization made it hard to work with and needed improving first, which CIOT is currently working on.

In collaboration with Supervision Pension Funds.

PACTA Name Matching

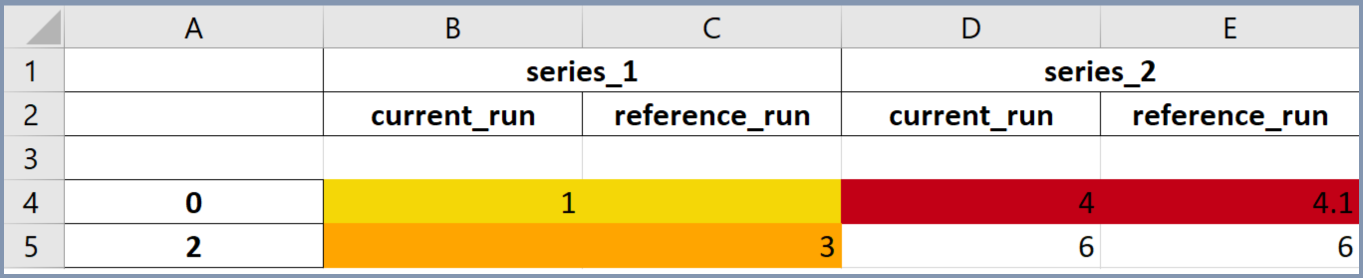
To what extent is the lending of Dutch banks in line with the Paris climate accords? To be able to answer this question data from the loans of banks should be matched to data of climate impact of companies. **We therefore developed a name matching algorithm**. Which links the AnaCredit data to the PACTA tool.

In collaboration with Supervision Policy.

Automate DFROG

Many data sources are needed for the Dutch Forecasting Model for Real Time Output Growth (DFROG) model. Currently, data collection is done by hand. We first identified the data sources where automation is possible, and then wrote software in Python that **automatically collects the necessary data**. This code was written in a generic way, such that in the future throughout the bank people can use this package to retrieve data from sources such as **CBS, ECB, Eurostat or Datastream** without any manual steps. We also built in data quality checks to ensure the data is correct and any changes over time are clearly displayed.

In collaboration with Econometrics & Models.



	A	B	C	D	E
1		series_1		series_2	
2		current_run	reference_run	current_run	reference_run
3					
4	0	1		4	4.1
5	2	3		6	6

Data Fetcher Package

In the “Automate DFROG” project we wrote code to automatically retrieve data from various sources using APIs and code, removing the need for manual data retrieval tasks. We are currently working on putting this code into a neatly documented package that can be used everywhere in the bank to easily automate data collection.

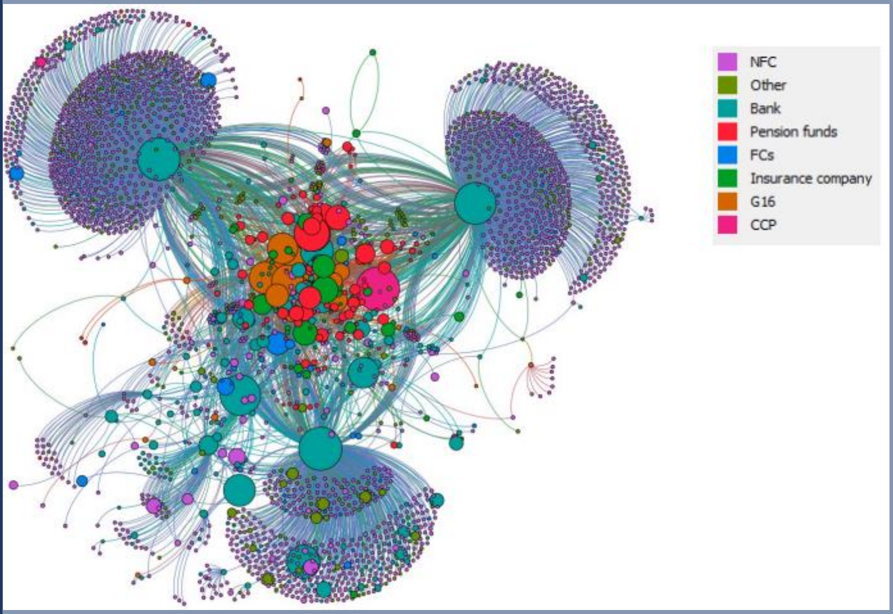
Under development, to be used by various departments.

Credit Claim Acceptance

With this project we support the check on Eurosystem eligibility of credit claims delivered by financial institutions to DNB, by setting up a method that **detects crucial elements in these claims** a collateral expert looks for. The conversion of the scanned pdfs to a readable format works well. It is possible to find elements in these documents that are currently checked by hand.

In collaboration with Payments & Market Infrastructures.

Standard Initial Margin Model (SIMM)



The European Market Infrastructure Regulation (EMIR) data contains information on all derivatives within Europe. We use this data to build **a model that computes the prices and sensitivities of derivatives** and use it to **determine the initial margins** based on the SIMM model. These findings can then be used for further research.
Using these insights, we have written an Analysis in which we provide an overview of the Dutch interest rate swap market. The interconnectedness across institutions in the swap market is represented by the graph on the left-hand side, where the colors of the dots indicate the different sector types and the sizes of the dots reflect the (logarithmic) aggregate size of the derivative positions.

In collaboration with On-site Supervision and Banking Expertise

DNB Website Statistics API

The main objective of this project is to **facilitate the process of downloading data sets from the DNB website**. A wrapper (API) around the website is developed as well as a visual web app so that users can select and download multiple data sets at once.

AnaCredit for Financial Stability

The AnaCredit data contains detailed information on loans in the Eurozone. We explore the **usability** of this data set and **incorporate the data into the monitoring and stress-testing frameworks of DNB**.

In collaboration with Financial Stability and Statistics.

Digital Twin for Physical Climate Risks

The main goal of this project is to **investigate whether digital twin technology can assist DNB** in identifying and predicting physical climate risks related to real estate

In collaboration with Sustainable Finance Office.