

# Experiences, essentials and perspectives for Data Science in the heart of central banks and supervisors

A case study of the Dutch Central Bank

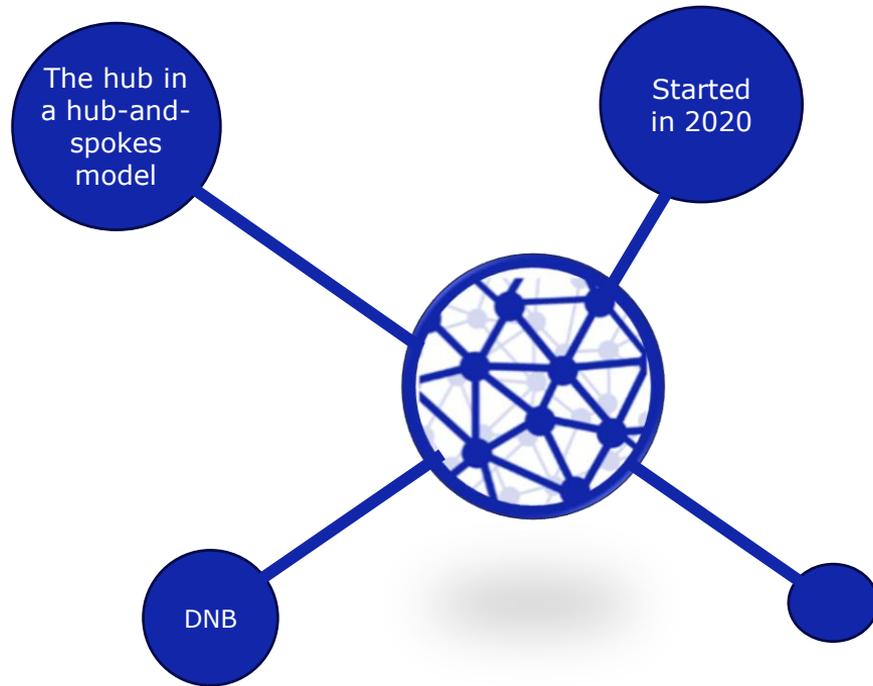
*Patty Duijm and Iman van Lelyveld*

DeNederlandscheBank



EUROSYSTEM

# Data Science Hub



## This presentation

7

Experiences

5

Essentials

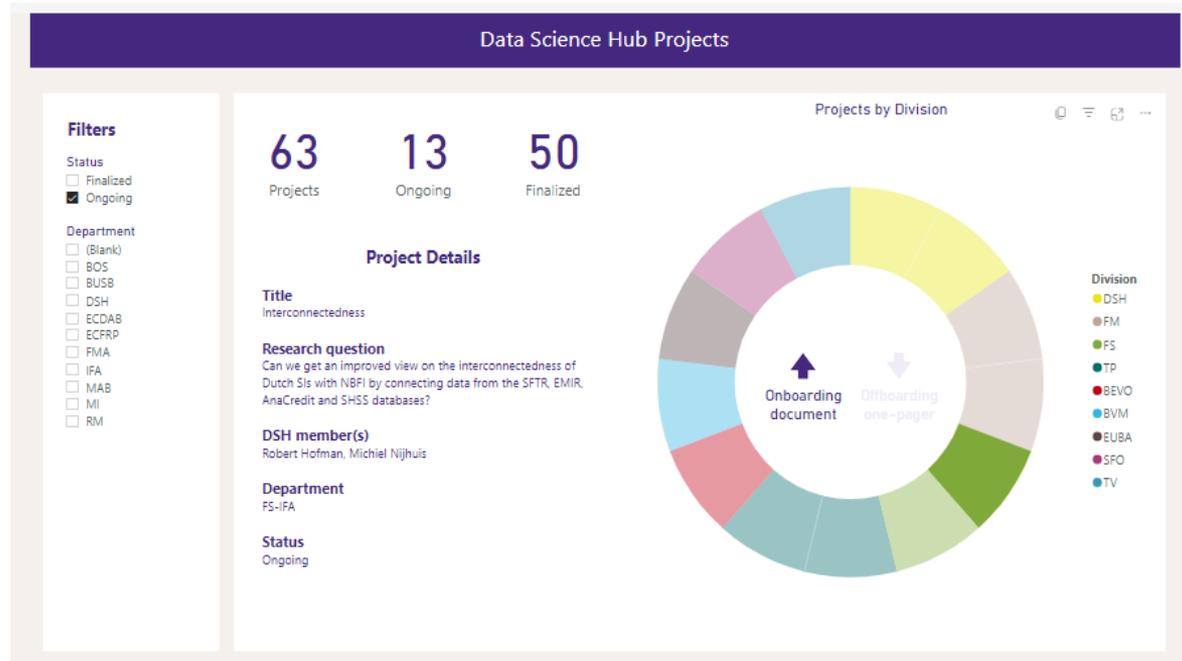


Way of Working

# Seven experiences

Since the start of the DSH in 2020 we have worked on 63 data science projects.

Our experiences are all illustrated with examples from selected projects.



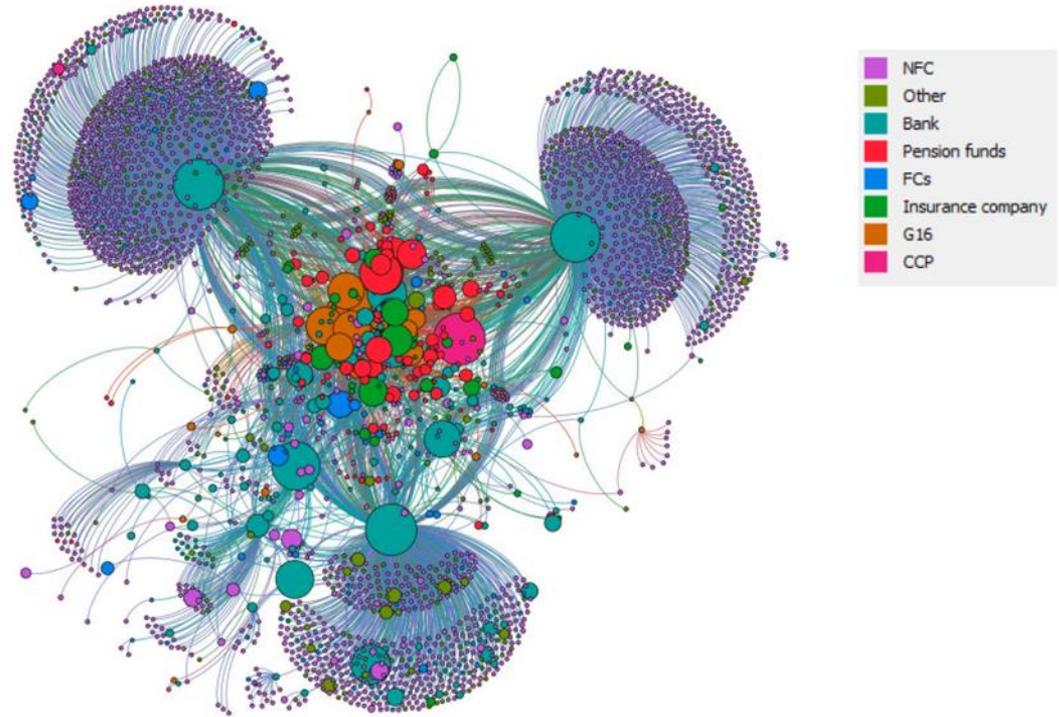
# Granular data

"The **combination of - sometimes novel - granular data** offers new insights useful for both supervisors and policymakers."

Examples are SHS, Anacredit, EMIR, MMSR.

Three challenges:

- Counterparty names as free-format text
- No single unique identifier
- One needs to consider different sources of granular data



## Project: Network analysis of interest rate derivatives

The colours of the dots indicate the different sectors and the sizes reflect the (logarithmic) aggregate size of the derivative positions.

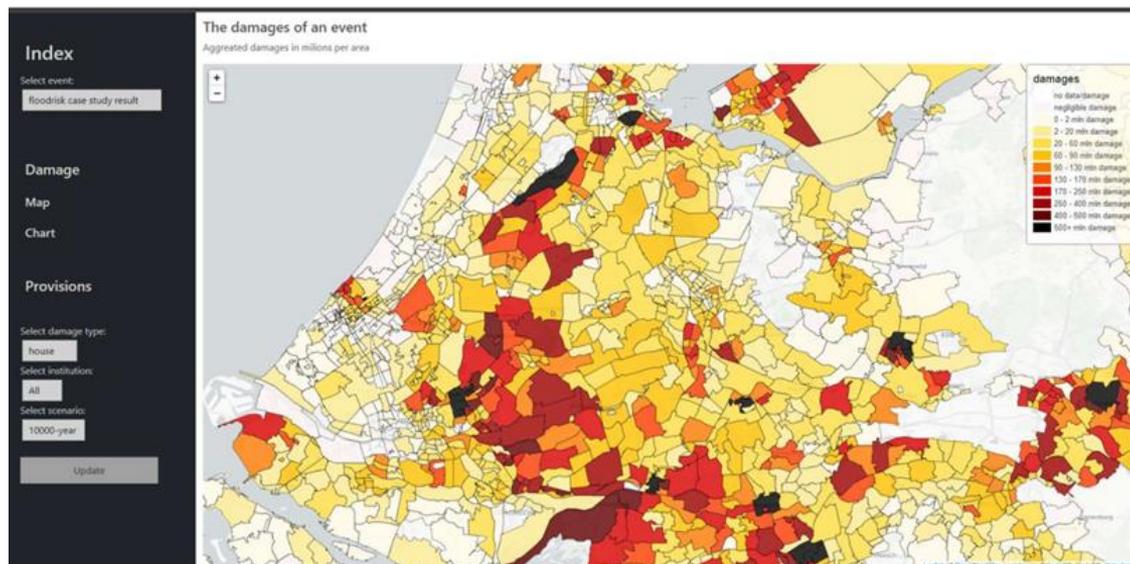
*The project is described in more detail in DNB Analysis  
"Estimating Initial Margins - The COVID-19 market stress" (Van den Boom et al., 2021)*

# Combining internal and external sources

*"Combining internal data with external resources increases the information value of the data, allowing supervisors and policymakers to improve the incorporation of external factors and risks."*

The data you own is much more valuable to you if it is augmented with data owned by others (Mewald, 2023)

Especially these "alternative" sources of data may generate unique insights.



## Project: Digital Twin of Climate Risks

In cooperation with the BIS Innovation Network, the Data Science Hub has developed a digital twin pilot of climate risks. The digital twin was developed to measure the effects of climate events on the financial system via real estate exposures of financial institutions.

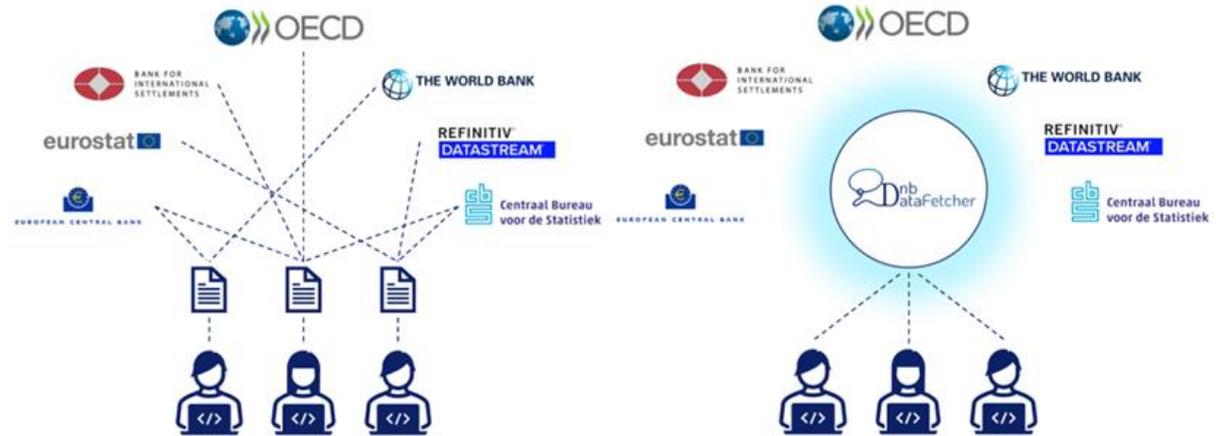
*More information can be found in the BIS note  
"Summary of the workshop on the role of technology in finance" (BIS, 2023)*

# Automating data processes

*"Automating data processes results in more efficient, accurate and highly frequent economic (prediction) models and allows policymakers to focus on modeling rather than preparing the data."*

Manual data wrangling is a labor intensive job, prone to human errors.

Automating data collection processes results in more efficiency and accuracy.



## Project: the DataFetcher

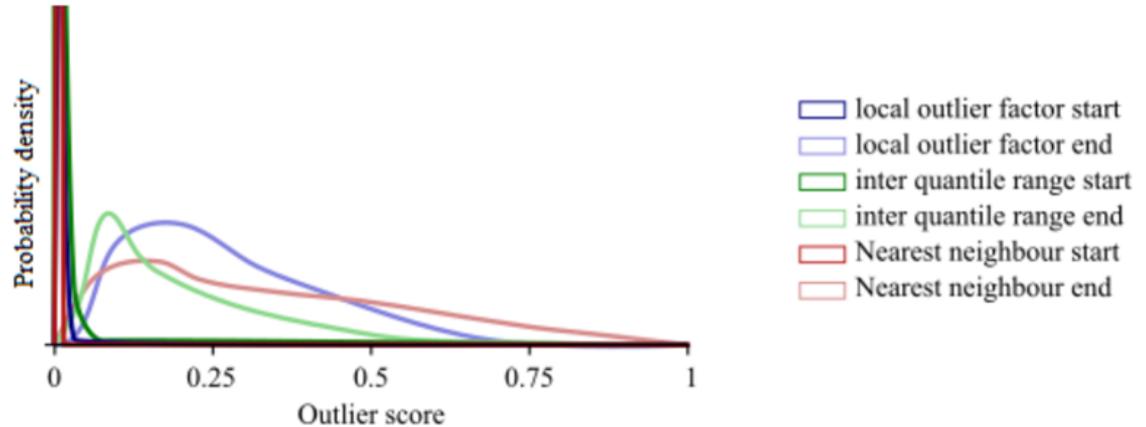
The DataFetcher is a package developed by the DSH and the figure illustrates its purpose. Multiple processes within the central bank use external data sources, resulting in colleagues collecting (the same) data manually or via ad hoc scripts. This may also result in cases where different (or even outdated) versions of the same data set are used within DNB. This is the left panel. In an ideal situation, i.e., the right panel, colleagues within the same institution have immediate access to the same data while restrictions dictated by privacy and confidentiality should be respected.

*The package is available to ESCB NCBs.*

# Start simple

*"Machine learning has great potential and outlier detection models have already proven to be effective innovations in supervision. Machine learning is, however, not a prerequisite for a successful data science project. In many cases, a simple model is a great place to start."*

An outlier detection approach should be able to optimally use the knowledge gained from the verification of the ground truth and adjust accordingly.



## Project: Outlier detection with reinforcement learning

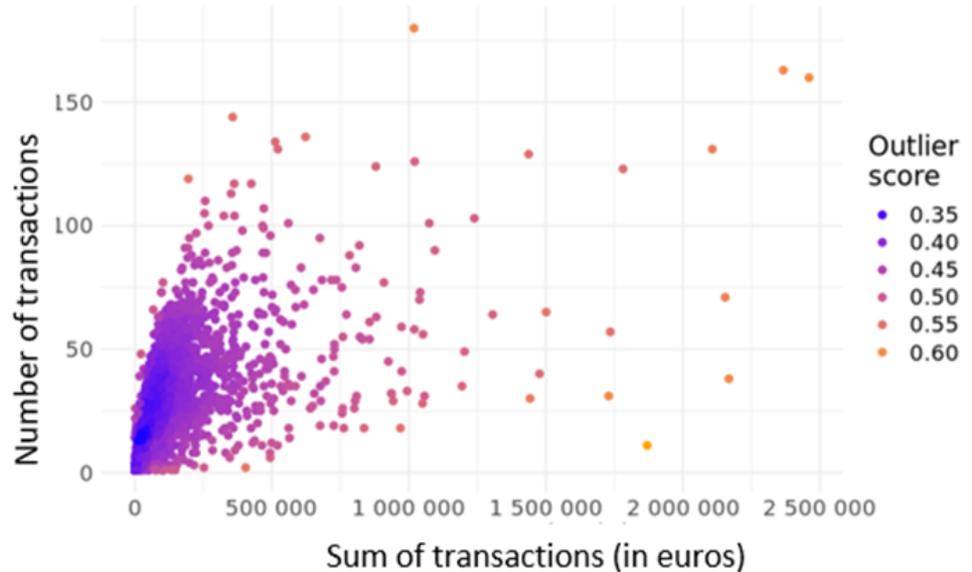
In this project we applied a reinforcement learning on a statistical outlier detection approach. Using an ensemble of proven outlier detection models in combination with reinforcement learning we were able to tune the coefficients of the ensemble with each additional bit of data. The figure shows the distribution of outlier scores for the three different methods. In a reinforcement learning approach, an algorithm is not just trained and applied, but in each iteration, the algorithm gets feedback on its performance.

*The project is described in more detail in the paper "[Outlier Detection with Reinforcement Learning for Costly to Verify Data](#)" (Nijhuis and van Lelyveld, 2023)*

# Domain knowledge

*"The real value of data science lies in the combination of data science techniques and domain knowledge, and therefore perfectly illustrates the importance of domain experts in data science."*

Domain knowledge is not only important as input for the model, but also to interpret model outcomes.



## Project: Know Your Customer

In this project outlier detection is used to identify potentially fraudulent client transactions from a sample of bank clients. With the model, we were able to effectively select clients with an abnormal transaction profile.

This project clearly shows the importance of domain knowledge. What are fraudulent transactions? It is relatively easy to classify an outlier as "fraudulent", while it is not. The graph shows outlier detection scores for bank clients, plotted against two of the client characteristics. The results of the outlier detection model resulted in the identification of new risks and efficiency gains, since supervisors are now able to consider all transactions instead of considering samples.

More information on this DSH project can be found in Cambridge Suptech Lab's ["State of Suptech Report"](#).

# Adoption by the business

*"The value of data science applications depends on the adoption by the business and therefore user-friendly interfaces to integrate the data science solution into the daily workflow are as important, if not more important, as technical excellence."*

Successful data science solutions can be both be one-off solutions or applications implemented in the workflow.

	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	
	Fluorescence	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	
DNB	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	
		Tape Area									
		Soil									
		Tear Size									
		Hole Size									
	Graffiti										
	Corner Fold										
	Corner Missing										
	Tape Decision										
	Fluorescence										

Cash handler

## Project: False Unfit Banknotes

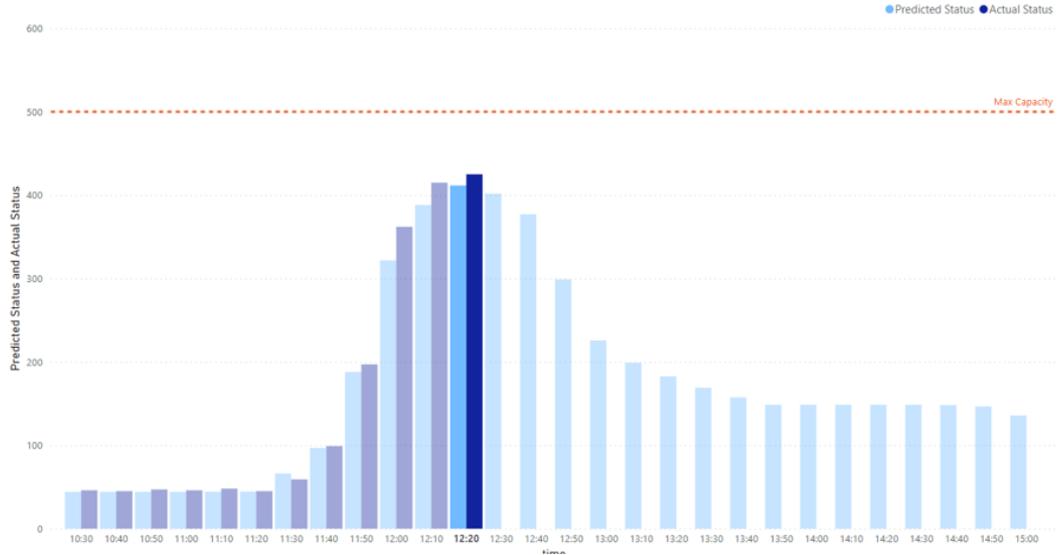
The aim of this project was to reduce the number of False Unfit Banknotes (i.e. banknotes classified as unfit by the cash handler, while deemed fit by DNB). While it is easy to compare the consequences of adjusting just one of the rules (e.g., tape decision or dirt), it quickly becomes more complicated once multiple rule settings are adjusted simultaneously. Machine Learning was applied to arrive at the optimal combination of multiple rule adjustments. Reducing the number of unfits can save a lot of effort and expense, and this project resulted in a set of recommendations for our Payments division to achieve these cost reductions.

# Data science for the entire organization

*"Last but not least, data science has value for the entire organization, including HR and business operations, and should therefore be in the 'heart of the organization.'"*

Don not only consider the traditional banking topics.

Data science can bring value for, for example, HR, internal services or the legal department as well.



## Project: Sensor data

We are currently experimenting with motion sensors in our office building to predict how busy our cafeteria will be. Figure 9 shows the results of our prediction model in the current state.<sup>9</sup> Such forecasts can help our catering to plan capacity and our staff make a more informed choice to time their lunch. Sensor data can, however, be used for other purposes, for example, to monitor the no-shows for meeting room reservations.

# Five essentials

## Dataloop

- Started as a pilot in 2018...
- Launched as an application to be used by both supervisors and supervised entities in 2023.

What are the essentials to integrate data science in the organization?

NEWS ITEM SUPERVISION

## New application: Dataloop

Read aloud

This autumn, we are launching a new application for Solvency II insurers in My DNB<sup>®</sup> Dataloop. The application offers you an improved user experience to interact with us, and it enables us to assess the data quality of reports more efficiently.

Published: 05 September 2023



© iStock

My DNB is the platform where we offer various services for the industry. Dataloop is an application for assessing the data quality of supervisory reports and related communication with insurers, and it will be part of the Reporting Service (formerly: Digital Reporting Portal).

## 1. Common view

*"Embracing new data science methods and using them throughout the organization requires sufficient appetite for experimentation -- especially at senior levels -- and therefore a common vision is key."*

## 3. Responsible coding

*"Responsible coding is needed to ensure the replication and reproducibility of the work."*

## 4. Data governance

*"A mature framework for data governance is needed to work responsibly, and this should be embraced by the organization."*

## 2. Right mix of skills

*"A data science function should be able to combine many different activities, and this requires the right mix of skills."*



## 5. Close contact with IT

*"A well-established IT environment is needed to facilitate data scientists in all their needs."*

# Way of Working at the Data Science Hub



## **Measurable goals**

Have a clear mandate and measurable KPIs



## **Inspiration & communication**

Inspire people at all levels, make available all project documentation and communicate about successes.



## **Project design**

Work in a structured way and agree upon division of tasks with onboardings and offboardings.



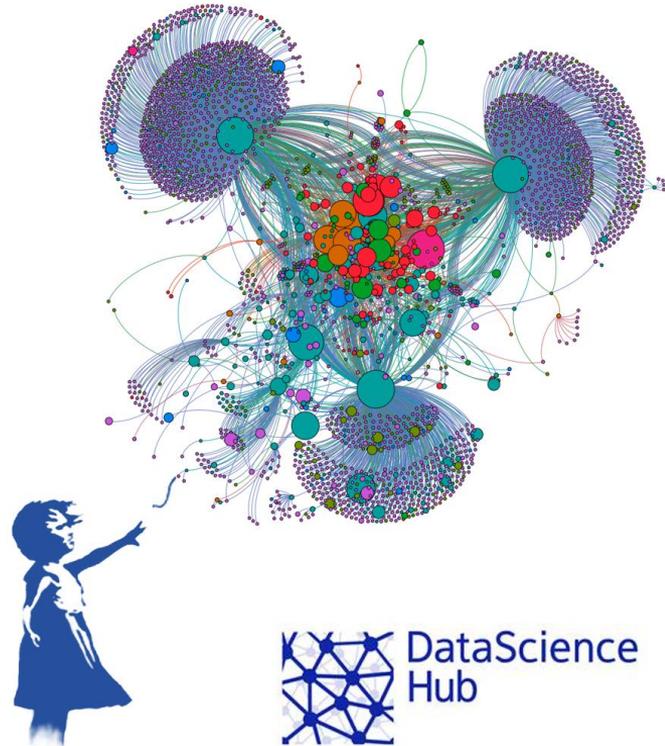
## **Activities and trainings**

Build a data science community by offering regular and ad-hoc events.

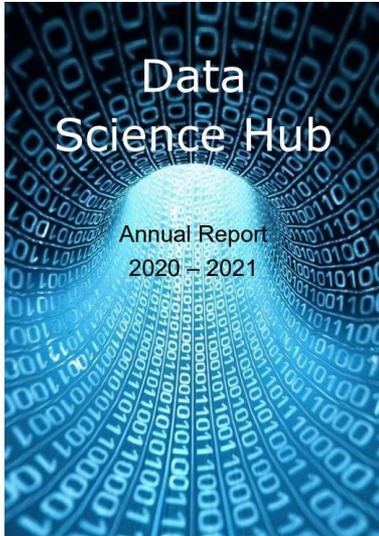
# Conclusion

- A well-defined strategy, mandate, goals, and a structured way of working definitely helped us to apply data science within the organization.
- Just start! And don't think that getting data science to work for an organization can be achieved by hiring a few smart data geeks and having them develop "AI" in a remote corner of the organization.
- Share! *"Instead of sharing shiny PowerPoint presentations, we could share the functionality that allows us to replicate the analyses of the others with our own data"* ([Majoor, 2022](#))
  - A quick win is to start sharing code, as we do through [DNB's GitHub](#).

**Data science has not only great potential but is already valuable for central bankers and supervisors.**



# Relevant links



[Annual Report  
2020-2021](#)



[Annual Report  
2022](#)



[DNB GitHub](#)



[data\\_science@dnb.nl](mailto:data_science@dnb.nl)