

# Data Sharing for Fraud Detection in Insurance: Challenges and Possibilities

Carl Christophe Louis Søylen-Knutsen and Bjørnar Tessem

*Department of Information Science and Media Studies, University of Bergen, Norway*

**Keywords:** Data Sharing, Insurance, Fraud Detection, Innovation.

**Abstract:** Digital development has opened up new tools to enable innovation, one of the options being data sharing among businesses. This paper addresses data sharing in the insurance industry and its innovation potential through a case study from a Norwegian data sharing project. The goal of the studied project is to achieve cross-company data sharing and with that enabling more efficient insurance fraud detection. We look at what requirements need to be fulfilled for data sharing to be implemented and what kind of challenges such a data sharing project meets. We analyse interview data from project participants and systematize their opinions and impressions regarding possibilities and challenges for data sharing. The case shows that data sharing among competitors in the insurance industry is hard to realise, very much due to the lack of trust in how the others will use data, but also due to competition laws and other regulations.

## 1 INTRODUCTION

Data sharing has great potential in business process innovation (Richter and Slowinski, 2019). However, there are in general many obstacles in the way of companies, entrepreneurs and the business community before they will be able to fully exploit the potential that data sharing can offer.

The insurance industry is a particular domain where data sharing may benefit both business and customers. A project led by the Finance Innovation business cluster in Bergen, Norway (Finance Innovation, 2021), is currently underway adopting data sharing among insurance companies with the aim of preventing insurance fraud. The assumption is that a larger pool of data about customers' coverage and claims will improve the ability to uncover fraud. The challenges are many and the biggest challenges may not be technical but rather at the legal level. For example, many of the project participants indicate that Norwegian legislation regarding data sharing is outdated.

Data sharing with personal data is further complicated by the General Data Protection Regulation (GDPR) within the European Union (EU), which also applies to Norway. The GDPR makes it pertinent to question whether the project of Finance innovation can be implemented at all, because of its restrictions on personal data. At the same time, the EU recognizes that data sharing is an important instrument in the de-

velopment of the European economy (Arnaut et al., 2018).

Several authors have addressed data sharing among companies, and they all point to the possibilities in various businesses areas (Huttunen et al., 2019; Celtekliligil and Adiguzel, 2019; Tang et al., 2018). This also includes data sharing for the purpose of insurance fraud detection (Power and Power, 2015). Others are focusing also on the difficulties the companies may encounter (Eckartz et al., 2014). To develop the idea of data sharing as a practical innovation tool there has been suggested several frameworks for organising and enabling data sharing (Grabus and Greenberg, 2017; Richter and Slowinski, 2019; Eckartz et al., 2014). In this research we take a case study approach, and address the issues of data sharing met in Finance Innovation's data sharing project. We use the interview data we have gathered to further develop our understanding of the data sharing concept, and aim to develop a richer understanding of the problems and opportunities related to data sharing.

We start with a background of current research on data sharing, and continue with a description of the Finance Innovation project "Fra data til innsikt" (From data to insight). We go through our data collection approach, and try to draw some new insights in a discussion section, before we conclude.

## 2 DATA SHARING

Arnaut et al. (Arnaut et al., 2018) describe, in their report to the European commission, data sharing as "The process by which a company re-uses data from another company, which is not a direct market competitor, for its own business purposes (excluding contractor-subcontractor relationships). These data were either accessed for free or acquired against some remuneration or other kind of compensation, including the provision of a service." The EU report suggests that such data sharing may contribute significantly value to companies and society, in terms of new business models, new products, and improved internal efficiency in companies. Within a government open data perspective, it has been shown that access to more data gives a small, but clearly significant innovation value (Jetzek et al., 2013). Even though the data sharing concept we are referring to here is not 100% the same thing, authors repeat the message also for business-to-business (B2B) data sharing; there are huge gains to be found (Huttunen et al., 2019).

Power and Power state that insurers around the world can save large sums if data can be shared internally in the industry. To achieve this, it is necessary that the process be streamlined so that sharing can take place as efficiently as possible and fraud can be uncovered more easily (Power and Power, 2015). They suggest that one could use data ranging from individual insurance claims, overall frequency of insurance use, types of insurance activities, individual behavioral data, and individual social media activities. Then probability based analyses of the data should enable targeting of the suspicious cases. There has also been parallel reflections about the possibilities and challenges in health care, mentioning for example the potential in sharing data about rare cases (Tang et al., 2018).

Huttunen et al (Huttunen et al., 2019) did an empirical study documenting B2B data sharing in Finland. 49% of companies in Finland did data sharing, and they had significant business gains from this; in particular from optimised operations. However, the need for data market places was not so accentuated, as data sharing often was established through bilateral agreements. Celtekligil and Adiguzel (Celtekligil and Adiguzel, 2019) has also verified innovation value from data and information sharing in tech industry, in particular if management support is high.

A few frameworks for data sharing addressing different aspect have been suggested. Richter and Slowinsky emphasise the need for data sharing platforms, and suggest to develop a framework for exchange of data similar to the FRAND model for exchange of

patents (Richter and Slowinski, 2019). Eckartz et al. (Eckartz et al., 2014) is more oriented to the practical issues of data sharing and have developed a four step framework for the process of sharing data, in particular addressing constraints. These constraints are Ownership, Privacy, Economic, Data Quality, and Technical. They used three cases from logistics to illustrate their model. Grabus and Greenberg (Grabus and Greenberg, 2017) have collected a number of data sharing agreements, identified commonalities, and developed a framework that could be used to automatically generate agreements taking care of privacy, legal concerns and other restrictions.

## 3 FROM DATA TO INSIGHT

Finance Innovation<sup>1</sup> (FI) is an initiative from the financial industry in Bergen, Norway to start a centre for innovation in Fintech. FI was started in early 2018 and one of their initial actions was to develop a collection of projects. The insurance industry signalled an interest in sharing data, motivated by an assumption that more data would give more precise models for fraud detection. This was especially true of the companies engaged in asset insurance, e.g., the insurance company Tryg<sup>2</sup>, which already had fraud models, but wanted more data. But also within the fields of people insurance (life, health, pension) there was interest. Yearly there is uncovered fraud claims for more than 300 million Norwegian Kroner in Norway, and the insurance business actors think the undiscovered fraud has a value many times higher than that (Finans Norge, 2019).

Fraud models are algorithms that insurers use to assess whether an insurance claim is legitimate or is an attempted fraud. Such models are used by all modern asset insurance companies, and they use machine learning and data to improve their own fraud detection. The assumption is that the more data a fraud model has access to, the better the model will be able to detect an insurance fraud. At the same time, people insurers such as DNB Liv<sup>3</sup> want to develop better computer-based models in life, health and pension insurance. Today, investigations of fraud suspicions against insurance claims in people insurance are something that is initiated based on assessments from case officers only.

From this the project "Fra data til innsikt" (From data to insight) was started with the goal of facilitating the sharing of claims data and fraud data from

<sup>1</sup><https://financeinnovation.no/>

<sup>2</sup><https://www.tryg.no>

<sup>3</sup><https://www.dnb.no/forsikring/personforsikring>

the various insurance companies (Finance Innovation, 2021). The project started with participants from the finance industry, including Tryg and DNB Liv, consulting companies such as Webstep and PwC, and researchers from the University of Bergen, the Norwegian School of Economics and Business Administration and Western Norway University of Applied Sciences. Engagement has been high at times, but at present the activity is low. For the time being, there is work going on with a so called minimal viable prototype (MVP) within car insurance, involving only two companies.

Within the project, it immediately became clear that there were challenges on several fronts. There were business and legal challenges, technological challenges, challenges with what data can be shared, and challenges with analytical methods. Over the past two years the project worked on the various issues, and received strong support from the subdivision for insurance in Finans Norge<sup>4</sup> (collaboration organisation for the Norwegian finance industry), and inside the participating finance industry companies.

However, in the last year and longer, the project has had little progress. Thus, the project can hardly be said to be a success. But perhaps that particular fact makes it more interesting to uncover what factors contributed to today's status of the project. Hopefully a case study of the project can provide important learning and understanding to aid future innovation projects with data sharing.

Thus, through a case study (Benbasat et al., 1987) of this project we may provide some valuable answers to these questions: What challenges in data sharing for fraud detection does the insurance industry have to address? What solutions may contribute to realize the innovation potential of data sharing for fraud detection?

## 4 DATA REGARDING THE PROJECT

We have gathered only qualitative data about the project. The most important data are interviews with project participants, but there is also some documentation in the shape of internal project emails.

The emails used in the analysis are provided from one of the authors who had an active role in the project. These emails also had attached some files (texts, presentations) relevant to the project. These emails and the attachments were used to get an overview of the earlier stages of the project and the direction the

project took.

The main data for our analysis are semi-structured interviews with project participants. The participants have been selected on the basis of the second author's knowledge of the project and on suggestions from Finance Innovation's project managers. It was a goal to have a diverse professional background in the respondent group. All the requested respondents were positive to participating in the interviews. The interview guide is found in the Appendix.

We interviewed seven respondents. Of these, six are still active in the project, while one has been out of the project for a couple of years. Almost all interviews were conducted digitally using Microsoft Teams due to the risk of coronavirus. One interview was conducted over the phone. Participants were made aware that they were recorded on audio or video recordings.

The respondents were

1. Project manager, Finance Innovation
2. Manager, tech consultancy company
3. Manager, business intelligence, insurance company
4. Data scientist, insurance company
5. CEO, Finance Innovation
6. Lawyer, consultancy company
7. Project manager (previous), Finance Innovation

## 5 ANALYSIS OF INTERVIEWS AND EMAILS

In the following the findings from the interviews are organised into seven topics:

- misaligned expectations
- GDPR and other regulations
- public support
- technical solutions
- facilitating the project
- the future of fraud
- other opportunities

During analysis, statements were coded with terms matching topics in the interview guide, but we also coded for other issues. Topics that emerged as important during the analysis were *misaligned expectations*, *technical solutions*, and *facilitating the project*. The other four topics came out of issues addressed directly in the interview guide. During interviews the questions about GDPR, public support, fraud in the

<sup>4</sup><https://www.finansnorge.no/en/>

future, and new business opportunities got the richest answers.

### 5.1 Misaligned Expectations

In the project's earlier phases, it was difficult to get the actors to agree on the scope of the project. The project initiative came from Finance Innovation, and to a less extent from the participants. The earlier meetings was meant to create some consensus, and data sharing was from the start the main theme, and then insurance fraud was seen as one possible arena for data sharing.

There was a lot of unstructured discussion about how things could be done and it was difficult to reach agreements, and the emails show that the same topics are recurring from meeting to meeting. Respondent 2 said, "there was a working group discussing and there somehow never became anything real out of it". The various participant companies sent people with different background. For example, one insurance company sent a data analyst, whereas another sent a business responsible. One aspect was the varying size of the companies. Some are market leaders, and others are smaller niche companies. One respondent pointed to the lack of funding to the project from the participating companies: "We need money, we need financial commitment from the actors, that's what it's really about".

### 5.2 GDPR and Other Regulations

All respondents considered GDPR to be a challenge to the project's feasibility. However, it was not considered the biggest challenge among all the respondents. But no one wanted to push the regulatory limitations, and this meant that important parts of the data would need to be excluded from data sharing. Respondent 1 said: "We have sacrificed quite a bit of explanatory power in order to stay on the right side of the law". They also mentioned other issues, like competition regulations in Norway, which state that it is not legal for businesses to cooperate if they have more than 30% market share. There are also laws that restrict the sharing of information among finance industry companies.

It was stated that both the GDPR and the Norwegian competition laws are unclear and that this presents significant challenges for the project. A similar project has not been done earlier in Norway. This means that the interpretation of laws lacks precedents, further complicating the understanding of the laws and consequently the progress of the project.

### 5.3 Public Support

During the interview the government's role in the project was discussed, and several of the respondents mentioned Datatilsynet<sup>5</sup> (Norwegian Data Protection Authority). They have set up a regulatory sandbox where businesses can try out solutions without being afraid of breaking regulations. This project would be a good case for the sandbox, and would ensure that all the issues with regulations could be put aside for a while. Respondent 3 said, "the regulatory sandboxes are a place where you can test out a project, without fear of reprisals, you might say. Where you work in cooperation with the regulatory authorities." One of the respondents mentioned that this project has so high societal value that it should allow for exemptions from regulations.

### 5.4 Technical Solutions

Recently the project has decided to focus on creating a MVP solution for car insurance. Only two of the initial insurance companies are involved in this (at the last time we checked), although one would have wanted more to make a solution viable. The envisioned, so called black box, solution is one where the companies will be able to see their own data only. But they may learn fraud models using all the available data. It is assumed that this will solve the regulatory challenges. However, this solution is according to respondent 6 semi-optimal. He said: "The solution, considering legal issues, would be to anonymise the data sets. Then you can do whatever you want with data and you can keep them for how long you want. But they were not able to do that. So I think the black box solution was a very good way to be able to actually give that AI enough data to train a model". Respondent 4 was more positive, and focused on how the MVP reduced chances of data misuse, but with a realistic twist: "It is quite difficult to completely remove the ability to misuse the data and we have talked about different solutions to that."

### 5.5 Facilitating the Project

As mentioned the project was initiated from FI, as a response to a brain storming process among finance industry actors. And having a facilitator like FI was essential for the project to get going. The importance of Finance Innovation being a non-profit organization in the middle was pointed out as a trust builder among participants. For example, using the premises of FI

<sup>5</sup>www.datatilsynet.no

allowed companies to meet on neutral ground and cooperate. It would have been different if such a meeting were to take place at one of the companies. In particular, trust was seen as important because there are fears that the data may be misused by dishonest forces. One of the respondents said that the weekly meetings and were an important contributor to creating trust.

But still the project progressed slowly. Respondent 3 suggested this was due to the lack of an early "minimum solution, good enough to demonstrate to the managements of the various insurance companies that this is something that will bring great value. Which will be worth investing in." The minimal solution was hard to do because of lack of example data from the various companies. In the end, they seemed to not dare to share data for an experimental solution. The companies had a strong scepticism towards sharing data because of a fear that competitors might get insights into their business models. It also was clear that it was difficult for them to set aside the people resources to prepare data. This again relates to the financial commitment from the actors that we also mentioned previously in the analysis.

## 5.6 The Future of Fraud

The respondents were invited to reflect on what the effect of such data sharing could be. Respondent 3 said "There is quite a few hundred million a year in costs so there is clearly a potential in preventing fraud. Thus in the utmost consequence you give a cheaper insurance to the honest customers". But there is no indication that such sharing of data will actually lead to more uncovered fraud. This remains to be seen. Still, some speculated about the possibility from data sharing to uncover new and more advanced fraud attempts. They were sceptical about the customers' ability to change their behavior as more fraud would be uncovered. The arguments was that they do not have insight into the algorithms, but also that insurance fraud is mostly not done by professional criminals.

## 5.7 Other Opportunities

Some respondents envisioned that data sharing also could give other opportunities than improved fraud detection. Respondent 2 reflected on the possibility to compete with international actors who have a large advantage in amounts of data, like Google and Facebook: "this is an opportunity for the Norwegian insurance industry to meet competition with any international players who come with a large data advantage".

Although related to data sharing only indirectly, automating decision processes was mentioned by several respondents. But they did not see any other business opportunities or new products.

## 6 A NOT SO SUCCESSFUL DATA SHARING PROJECT

The "From Data to Insight" project was started with a diverse group of participants, with the aim of sharing data among companies in insurance. The project decided to focus on sharing data for the detection of insurance fraud, based on assumptions about a potential similar to what Power and Power argued for (Power and Power, 2015). That is, more and richer high quality data shared among competitors would enable automated fraud detection, and significant reductions of losses. The project was not successful compared to the initial expectations, and has ended up with only two partners sharing data, aiming towards a somewhat restricted black box solution.

If we look at EU's definition of data sharing (Arnaut et al., 2018), we notice immediately that this does not include data sharing among competitors, which the FI project actually aims to do. And perhaps this is also the reason why the project never really took off. Eckartz et al.'s (Eckartz et al., 2014) list of constraints had **Ownership** on the top, and most likely there was some lack of trust (or not enough trust) among the companies regarding how others would use the data to infer business models. The companies actually saw a big problem in sharing these data. This, together with the regulations imposed by Norwegian competition laws put in effect due to being competitors, actually created some Ownership hurdles in the start that the project never managed to pass. Privacy was also considered an important issue, but since actually no data was really shared in the sense that competitors could see each other data, it was less of a problem than the competing company issues.

Another aspect that hindered the development of the project, was the lack of technical solutions showing its viability. This most likely caused less supporting management in the participating companies. They wanted to see how value could be realised before committing to the data sharing idea. An early regulatory sandbox solution demonstrating the value would probably have helped in engaging management more. Celteklicil and Adiguzel (Celtekligil and Adiguzel, 2019) have shown that, as in most information system projects, management support and involvement is critical to success. The insurance business is conser-

vative, and wants to see clear value from initiatives.

Huttunen et al. (Huttunen et al., 2019) showed that shared spaces for data sharing are not necessarily needed, bilateral solutions may give just as high value. The advantage is that those may result in less complex agreements, and remove some of the more challenging competition and privacy issues. Eventually this was the outcome of this project, two companies collaborating on data about car insurance.

In the initial research questions for the project we aimed to get insights into what kind of solutions were possible in such a project. What we observed is that beyond the restricted black box solution, which is not yet operative, there are basically no solutions found yet. Very much stopped due to the inability to handle the initial constraint of ownership. Perhaps, if one had a focused on a demonstration solution from early on, there would be more engagement in trying to solve the ownership issue.

This project is a case study with a limited set of interview respondents, and will of course not generalise to other situations, and in particular to other projects on data sharing for insurance fraud detection. However, it illustrates some of the problems such a project may meet, and emphasises the need to focus on early demonstration of a viable system, focus on handling trust among data sharers, and commitment among managers to assign the necessary people resources. As mentioned, insurance fraud has large societal costs, and solutions to reduce it will have great value for companies, customers, and society in general.

## 7 CONCLUSION

This paper describes a case study of a data sharing project in the insurance industry, where data sharing was hoped to be a driver for innovation. The project was in large parts unsuccessful. We have observed how data sharing among competitors becomes problematic due to several forms of regulations, divergent interests, lack of trust, lack of management support, and other issues. The case study shows that data sharing among companies is difficult to realise, even though there is a large potential for economic gains for all parties. One observation is that early demonstrators of solutions, for example, in a regulatory sandbox, could contribute to increased trust, management support, and engagement.

## REFERENCES

- Arnaut, C., Pont, M., Scaria, E., Berghmans, A., and Leconte, S. (2018). *Study on data sharing between companies in Europe*. everis Benelux.
- Benbasat, I., Goldstein, D. K., and Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. *MIS Quarterly*, 11(3):369–386.
- Celtekligil, K. and Adiguzel, Z. (2019). Evaluation of Data Sharing in Production Firms and Innovation Orientation in The Effect of Management Capability on Operational Performance. *Procedia Computer Science*, 158:781–789.
- Eckartz, S. M., Hofman, W. J., and Van Veenstra, A. F. (2014). A Decision Model for Data Sharing. In Janssen, M., Scholl, H. J., Wimmer, M. A., and Bannister, F., editors, *Electronic Government*, Lecture Notes in Computer Science, pages 253–264, Berlin, Heidelberg. Springer.
- Finance Innovation (2021). Detection of Insurance Fraud. <https://financeinnovation.no/innovation-projects/detection-of-insurance-fraud>. Accessed: 2021-08-17.
- Finans Norge (2019). Forsikringssvindel i Norge - svikstatistikk 2019 (Eng: Insurance fraud in Norway - fraud statistics 2019).
- Grabus, S. and Greenberg, J. (2017). Toward a Metadata Framework for Sharing Sensitive and Closed Data: An Analysis of Data Sharing Agreement Attributes. In Garoufallou, E., Virkus, S., Siatiri, R., and Koutsomihia, D., editors, *Metadata and Semantic Research*, Communications in Computer and Information Science, pages 300–311, Cham. Springer International Publishing.
- Huttunen, H., Seppälä, T., Lähteenmäki, I., and Mattila, J. (2019). What Are the Benefits of Data Sharing? Uniting Supply Chain and Platform Economy Perspectives. Technical Report 93.
- Jetzek, T., Avital, M., and Bjørn-Andersen, N. (2013). Generating Value from Open Government Data. *ICIS 2013 Proceedings*.
- Power, D. and Power, M. (2015). Sharing and Analyzing Data to Reduce Insurance Fraud. *MWAIS 2015 Proceedings*.
- Richter, H. and Slowinski, P. R. (2019). The Data Sharing Economy: On the Emergence of New Intermediaries. *IIC - International Review of Intellectual Property and Competition Law*, 50(1):4–29.
- Tang, C., Plasek, J. M., and Bates, D. W. (2018). Rethinking Data Sharing at the Dawn of a Health Data Economy: A Viewpoint. *Journal of Medical Internet Research*, 20(11):e11519.

## APPENDIX

### Interview Guide

- Tell me about your professional background and how you got into the finance innovation project? What kind of tasks do you have on your project?
- What do you see as the biggest challenges facing the project today?
  - Do you think of GDPR as a challenge to your project? Why? Why not?
  - What challenges does the GDPR present?
  - Do you think that Norwegian regulations present challenges for the project? In case, what challenges?
  - Do market participants pose any challenges? In case, what challenges?
- What do you think it takes for the project to be realized?
  - Do you think trust is an important factor?
  - Why is trust important? How can trust be improved?
  - Can the insurance industry learn anything from the banking industry in terms of data sharing? The have managed to cooperate on systems such as BankID. What do you think you can learn from it? What can be transferred from the banking industry to the insurance industry?
    - Do you think this is something the Norwegian authorities can do to realize the project? In case what?
- What opportunities do you think the project can give the insurance industry?
  - Do you think that the insurance industry will be more efficient with data sharing?
  - Do you think it can give the insurance industry new business opportunities? In case what?
- How do you think the project could help uncover insurance fraud?
  - To what extent do you think it could help uncover insurance fraud?
  - How will it support the process of solving insurance fraud?
  - Are there any special types of fraud that are going to be more easily uncovered than others?
  - Could we see new forms of insurance fraud if the project is successful? In case what?