



Limits to data quality?!

Solvency II requires data for Technical Provisions and for Internal Models to be ‘complete, accurate and appropriate’¹. DNB (2017) provides guidance “to ensure data quality”². These are ambitious requirements.

In real life, there are data quality issues. Many data quality issues have ‘solutions’ (Osborn (2024), Rao (2024) and Elahi (2022)). In this article, we present a few ‘limits to data quality’. Issues that we feel are root-causes, hard to resolve. We suggest that they deserve more attention and may well influence data quality expectations.

DNB (2017) describes the requirements for a data quality management framework in various areas. We identify root causes for data quality issues in three of these: (1) data architecture, (2) external providers and (3) objectives. We end with concluding comments.

DATA ARCHITECTURE

Two (related) root-cause issues related to ‘Single Source of Truth’ (see e.g. CRO Forum, 2020).

Issue 1: Redman (2016) writes:

The data they need has plenty of errors, and in the face of a critical deadline, many individuals simply make corrections themselves to complete the task at hand. They don’t think to reach out to the data creator, explain their requirements, and help eliminate root causes.

This is not an easy issue to resolve. The root-cause could be lack of time for the proposed feedback loop to the data creator.

Issue 2: A Reddit question: “What are the most challenging data quality issues you face frequently?”³ Wiki702 replies:

Zero schema enforcement with little to none existent governance.

Not all internal users feel fully comfortable surrendering their data models (‘schema’) to a central authority. Wang at al. (1991) identifies ‘believability’ as the first criterion for quality data. Building an effective central database takes time, to ensure that the user trusts the database.

DATA QUALITY POLICY – EXTERNAL PROVIDERS

Data quality of external data is a challenge. DNB (2017) writes:

“With external data providers, a data delivery contract has been agreed. In this contract, it described at data element level what data is to be provided, the way quality requirements are ensured and steps to be taken if the quality requirements are not met.”

Issue 3: As for ‘what data is to be provided’, the devil is in the details. Wiki identifies a host of ‘metadata’ dimensions⁴. ISO (2015) specifies ‘syntactic quality’ as a dimension of data quality. We suggest that the use of external data also requires a helpline to the ‘scientists’ driving the measurement processes. If there is an outlier, you may need to consult the scientists.

Issue 4: As to ‘quality requirement are not met’. Data (quality) requirements are dynamic, in line with changing data demand and changing data supply. How much lead-time does a supplier give when changing their data definitions? How much influence does the insurer have on the data definitions they can get from the supplier? That influence is often limited.

DATA QUALITY POLICY – OBJECTIVES

To make business decisions on data quality requires measures / scores of data quality. When are data ‘good enough’?

DNB (2017) recommends:

From the risk assessment, you may conclude required data are missing. [...] We expect that you as insurer, on the basis of this risk assessment, quantify the impact on your Solvency II reports.

The term ‘on your Solvency II reports’ is rather broad. A specific impact focus would be the Solvency Ratio. Delegated Acts (Article 222) goes one level deeper:

change or error in the outputs of the internal model, including the Solvency Capital Requirement, or in the data used in the internal



model shall be considered material where it could influence the decision-making or the judgement of the users of that information, including the supervisory authorities.

When a breach of the Solvency Capital requirement is threatened, data need to be particularly accurate. These ‘top-level criteria’ can then be cascaded down depending on the business decision.

Issue 5: It is, almost by definition, impossible to quantify (the impact of) data completeness / accuracy. We do not have the impact of missing / inaccurate data on Solvency II reports. We do not have the impact of missing / inaccurate data on decision-makers.

For such cases, article 82 of the Solvency Directive reads:
appropriate approximations, including case- by-case approaches, may be used

Yes, we can construct approximations (say scaling) to handle ‘completeness’. And, yes, we can do sensitivity analyses to assess the risk of missing / inaccurate data. But we do not know how ‘appropriate’ these approximations are, or exactly what sensitivity analyses to do. That is why we need data. The CRO Forum (2020) writes ‘Data quality is [...] a risk in itself: an operational risk’.

CONCLUDING COMMENTS

DNB (2017) writes: “*DNB expects that at all times insurers strive to attain the data quality requirements for critical data elements*”. This could be viewed as a dynamic, ‘holy grail’ interpretation of data quality requirements.

Data quality is a journey, with the following lessons learned:

- In developing new regulatory requirements, regulators and insurance companies need to cooperate to ensure that they are ‘doable’. It helps if data used for reporting can be used for steering, as interests of regulators and insurance companies are aligned.
- It is important to have a lead-time for new regulations to be implemented. This is needed to implement the new regulations, and to manage expectations about what can be delivered when. Initially, approximations may need to be used.
- If regulator / insurance company can identify data that are more ‘complete, accurate or appropriate’, the insurer is expected to use these data (‘comply or explain’).

As for the end result of the journey: we will be in data heaven, full of perfect quality data. ■

References

Atlan, 2024, “Data Quality vs Data Accuracy: 15 Key Differences To Know”, Data Quality vs Data Accuracy: 15 Key Differences To Know.

CRO Forum, 2020, “Data Quality in the Insurance Sector”.

De Nederlandsche Bank, 2017, “Guidance beheersing Solvency II data kwaliteit door verzekeraars”.

Elahi, E., May 23, 2022, 12 Most Common Data Quality Issues and Where Do They Come From, 12 Most Common Data Quality Issues – Data Ladder.

ISO (2015), “Data quality – Part 8: ‘Information and Data Quality: Concepts and Measuring’”. ISO 8000–8.

Osborn, T., Updated May 08 2024, “8 Data Quality Issues and How to Solve Them”, 8 Data Quality Issues And How To Solve Them.

Rao, S., Nov 29, 2024, 10 Common Data Quality Issues (And How to Solve Them), 10 Common Data Quality Issues (And How to Solve Them).

Redman, T., 2016, “Bad Data Costs the U.S. \$3 Trillion Per Year”, Bad Data Costs the U.S. \$3 Trillion Per Year.

Wang, Y.R., and L.M. Guarascio, 1991, “Dimensions of Data Quality: toward Data Quality by Design”, IFSRC Discussion Paper #CIS–91–06.

1 – The Directive (2009 /139/EC, Articles 82 and 121) contains data quality requirements. The Delegated Acts (2015/35) develops these in Article 222 and others (see CRO Forum, 2020).

2 – Quotes translated from Dutch.

3 – What are the most challenging data quality issues you face frequently? : r/dataengineering
https://www.reddit.com/r/dataengineering/comments/1c8dy1r/what_are_the_most_challenging_data_quality_issues/

4 – Metadata – Wikipedia
<https://en.wikipedia.org/wiki/Metadata>

T. van der Meer AAG works as Senior Actuary at Achmea. The article is written on a personal title.

