



Showcasing strong processes and full transparency that leads to superior Data Quality

Since 2004, GCD has continuously reinforced a framework that is used to measure and monitor Data Quality (DQ). The objective is to achieve high DQ and compliance for the GCD pooled data, as required by global regulations (BCBS 239, ECB Guide to internal models, Fed SR1107).

Data Quality Dimensions (BCBS 239 Compliant)

Timeliness	<1 YR.	Trend**	More than half of data received by GCD is published within one year from the closure of the workout.*
Stability	94%		...of the data remains identical at each update (measured on 10 main LGD drivers). The current trend reflects mild impact of some banks portfolios evolution.
Correctness	99%		...of the data is passing the latest validation rules. (1% legacy data only comply to former validation rules)
Completeness	93%		...of values for mandatory data fields are populated.^
Optional Completeness	48%		...of values for <u>optional</u> data fields are populated.

Freshness of 2022 Data

- Total members active in the LGD Pool: 52
- Fully refreshed submissions performed: 37
- Partially refreshed submissions performed: 15

Substandard data submission not accepted for 2 out of 52 banks.

A full submission concerns a submission of all historical defaults from a member in all Facility Asset Classes a member is contributing to. A partially refreshed submission is when a bank only updates the loans that evolved between two data-collection cycles (e.g. only new defaulted loans are added to the data history). The numbers reflect the refresh of default data over the last three years.

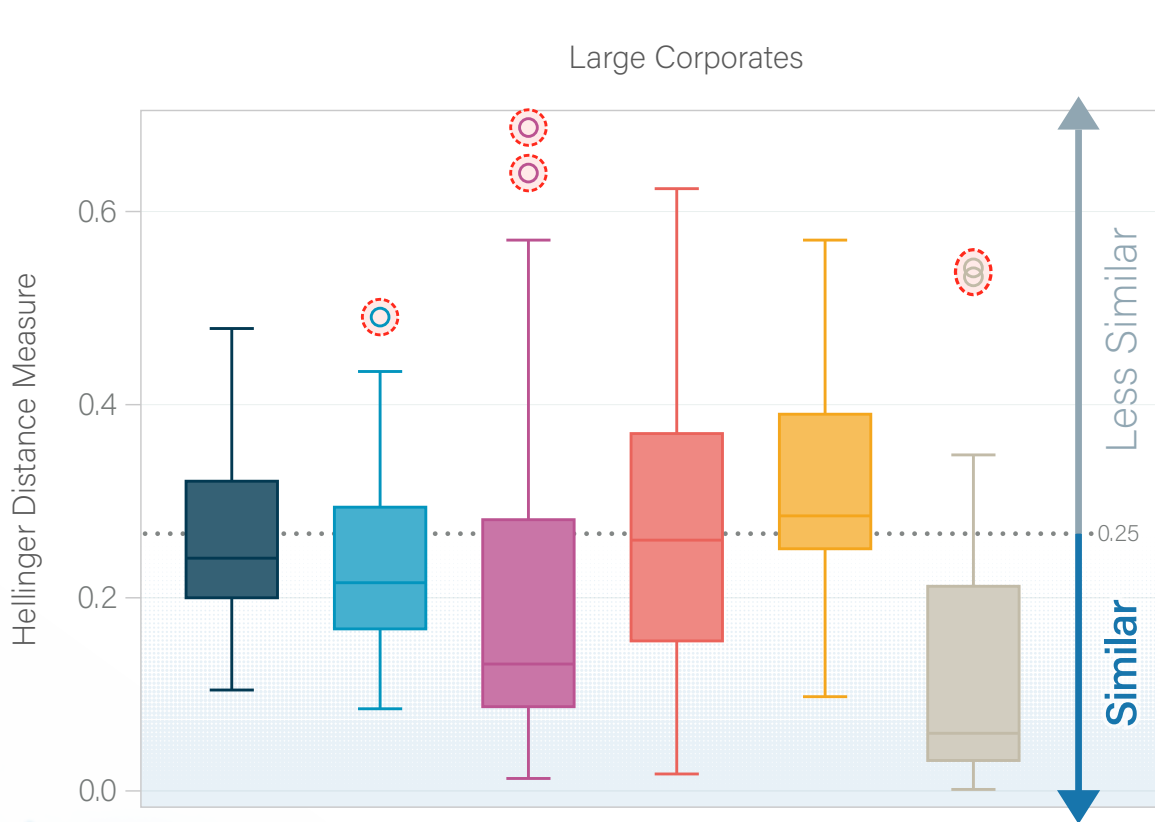
All dimensions are controlled via **450+** quality checks.

* 75% of all closed workouts are received and published within 2 years. ** The YOY trend for the metric. ^ 7% of values for mandatory fields are populated with an escape clause.

Recovery Data is Consistent Across the Majority of Members.

Similarity Analysis

As most of the banks' key risk driver distributions are similar, the Hellinger Distance methodology allows banks to assess representativeness for their model development and benchmarking needs. A few banks have specific business footprints or recovery processes that make them less similar to other banks. Overall, data quality improved for some banks, resulting in more precise classification of the security on loans.



Key Risk Driver	Number of Similar Banks <small>Show the Same Distribution</small>	Number of Less Similar Banks
Time to Resolution	26 <small>Trend</small> ▶	19
Time to Recovery	29 ▶	16
Guarantor Indicator	32 ▶	13
Collateral Indicator	24 ▼	21
Seniority	18 ▼	27
Cure	39 ▶	6

How It Works

GCD compares the similarity of individual banks to the overall data pool using the Hellinger distance measure*. It evaluates various risk drivers to determine the consistency of the data between different banks.

* When the Hellinger Distance Measure for a bank results below 0.25 it is considered as similar profile. This 0.25 threshold is in line with the industry best-practice usages of the measure.

Data Quality Metrics Show Continuous Improvement

GCD monitors global DQ variations by calculating a synthetic LGD distribution indicator based on the Inverted Coefficient of Variation. The metrics (averages and standard deviations) in the chart are calculated at loan level. The data for each submission semester is cumulative.

The chart clearly shows three distinct phases. There is strong volatility of the DQ indicator during the first years of GCD concepts/process alignment between contributing banks up until 2011, followed by the DQ stabilizing at its first plateau through 2016. Finally, the DQ moves into a third period with reinforced DQ management requirements. This shows a new higher, yet still stable, plateau.



GCD's high data quality is the result of a long term investment.

Datapool Volumetrics

Global Credit Data has built up the world's largest non-retail bank loan loss database with over 300,000 defaulted facility observations, totalling over €700 billion in all non-retail Basel asset classes. The LGD/EAD platform, the largest of GCD's databases, was created in 2005 and has grown rapidly ever since. It now gathers detailed information on tens of thousands of defaulted counter parties, particularly data on cash flows and collateral.

LG2022 Datapool Overview by Facility Asset Class

Facility Asset Class	Number of Defaults	Number of Loans	Exposure [in mn EUR]
Small/Medium Enterprises (SME)	104,697	206,737	€128,157
Large Corporates	21,802	48,571	€343,374
Banks & Financial Companies	3,169	5,711	€93,134
Ship Finance	885	1,587	€17,600
Aircraft Finance	364	850	€6,185
Real Estate Finance	16,907	26,786	€80,723
Project Finance	578	1,205	€17,692
Commodities Finance	449	849	€9,156
Sovereigns, Central Banks	142	267	€12,341
Public Services	231	360	€2,502
Private Banking	6,053	9,671	€9,992
	155,277	302,594	€720,856

All metrics in this dashboard are based on results of the H2 2021 LGD data submission cycle.