Global Credit Data
by banks for banks

Agenda for Methcom Meeting 16 December 2020

Members:

Stephan Jortzik (chair)	Pubudu Premawardena
Patrik Gunnarsson	Ntateko Maimane
Michael Eichhorn	Andrea Buzzigoli
Eric Lin	Rohan Surana
Elizabeth Moleda	Stuart Neilson
Abhishek Kumar	Michael Jacobs
Clemens Mesterom	

Executives: Erik Rustenburg, Nunzia Rainone (minutes), Olivier Plaetevoet, Richard Crecel, Michaël Dhaenens, Hale Tatar, Nate Royal, Nina Brumma, Izelle Kirsten

External: Benjamin Galow

Meeting Start: 1200 CET
Location: by telephone and webex

Item No.	Start time	Item	Responsible	Info or Resolution	Material
		Standing Items:	-		-
1.	1200	Approval Draft M129 Minutes Methcom	NR	R	X
2.	1205	Action points	NR	I	X
3.	1210	Forward looking calendar 2021	NR	I	X
4.	1220	Methcom Calendar and Forward Looking Calendar 2021	NR	R	X
5.	1240	Review Subcommittee Charters	ER OP HT	R	X
6.	1300	Data quality Public Dashboard and Policy	ER BG	R	X
7.	1315	GCD 2021 submission cycles plan	MD	I	X
8.	1330	Review of Country groupings/hierarchy table	NR	I	X
9.	1345	Industry information: GCD stand point	NR	I	X

		Other Items	-		-
10.					

Anti-Trust Warning: participants are warned not to provide sensitive information about their financial institution or customers and are warned not to engage in discussions which might encourage or lead to collusive behaviour. If in doubt then please seek guidance from your own institution's policies or legal counsel.

Item 1

Global Credit Data
by banks for banks

Draft Minutes of the Methcom Meeting held on 27 October 2020

Members Present:	Stephan Jortzik (chair) Pubudu Premawardena (chair) Clemens Mesterom Elizabeth Moleda Michael Eichhorn Michael Jacobs Patrik Gunnarsson Stuart Neilson
Excused:	Abhishek Kumar Andrea Buzzigoli Eric Lin Rohan Surana
External (d-fine):	Benjamin Galow
Executives:	Nunzia Rainone (at the minutes) Erik Rustenburg Michael Dhaenens Nate Royal Richard Crecel

Anti-Trust Warning:

Participants were warned not to provide sensitive information about their financial institution or customers and were warned not to engage in discussions which might encourage or lead to collusive behaviour.

Matters are minuted in the order dealt with by the meeting with the numbering based on the agenda. Material for most of the issues to be discussed had been sent out in advance together with the agenda.

Meeting Start: 12:00 CEST

Quorum is met.

Item 1: Approval Draft 128 minutes Methcom June 23rd 2020

It has been clarified that the submission of Unresolved volumetric spreadsheet is on voluntary base. The minutes were **unanimously approved**.

Item 4: Calibration of statistical PD rating models. Discussion point for members

NR and MJ presented the material.
The following questions/remarks were made:
Most bank are currently calibrating their PD models using a reference masterscale. A Working group on this topic might be of interest for GCD members, with the production of a white paper at the end

of the works. The WG could be also the opportunity to begin the discussion of having a GCD masterscale of reference.
SJ suggested to add to the calibration question, the question on how do we want models to behave, particularly through the cycle.
Covid component might be introduced to rase interest for larger audience.
The Methcom **unanimously approved** the WG initiative.

Item 5: Restructured Facilities
NR presented the material and highlighted the proposed changes.
PG expressed his concern on the complexity of the proposal, not well balanced by the possible advantages. The re-allocation of transactions in the original loan is a complex process. Therefore, methcom members suggested to investigate the current status: how data pool participants are currently submitting restructured loans?
Decision: NR has been instructed to engage in 1 to 1 discussion with LGD data pool participants and investigate how members are currently submitting restructured loan, and the quality of the data they are currently providing. Results of the investigation will be showed and discussed with Methcom.

Item 6: Introduction of Ben Galow long-term senior consultant for GCD (d-fine)
RC Introduce to Methom Ben Galow, long-term senior consultant for GCD. Ben has gained experience on GCD data working on the Unresolved LGD study, closed in June 2020. Ben has a long experience in Credit risk, with a string background in methodology. He will work on specific topics, such us the DQ initiative, item 7 of the meeting agenda.
Methcom welcome Ben.

Item 7: Data quality process and policies: project updates

ER and BG presented the material. Following the compliance WG, GCD has moved in the direction to enforce data quality, measure it and share it transparently with members. The policy will show the GCD data quality management process with a deeper explanation of the data quality dimensions, based on the BCBS239, used to test and monitoring data quality.
In the next meeting the Data quality dashboard and the full document will be shared and discussed with Methcom.

Item 8, 9, 10 and 11 were not discussed as there was no time left. NR invite member to reach out her via mail in case of questions on the remaining items.

SJ closed the call at 14:10.

Item 2

Global Credit Data
by banks for banks

Action points from Methodology Committee Meetings

December 2020

AP No.	Content	Holder	Due Date	Closed	Prio	Status
MC202010	Restructured Loans investigation	NR/ER	H1 2021		H	☺

Item 3

Forward looking calendar 2020 - gray items have been changed from last FLC - **blu items have been addedd since last FLC**

Standing Items			Implementation	Feb	April	Jun	Sep	Oct	Dec
ER	LGD	Outcome of LGD Subcommittee meeting		X	X	X	~~X~~	X	~~X~~
OP	PD &Rating	Outcome of PD & Rating Subcommittee meeting			X				
HT	BP	Outcome of BP Subcommittee meeting			X				
NR	General	Review Subcommittee Charters	H2 2020						X
DT	General	Review Data Pool Regulations	H1 2021		X	X			
Items									
NR/ER	LGD	Restructured Facilities	H2 2020	X	~~X~~		~~X~~	X	
DT	All Platforms	Definition of Default survey	H1 2020	X	X	X			
ER	LGD	Major project: Documentation update	Ongoing		~~X~~			X	
NR	All Platforms	Assessment of Raw Industry Code and LEI implementation	Ongoing	X		~~X~~	~~X~~	X	
NB	LGD	2020 Analytics and Report overview	2020	X	X		~~X~~	X	
Executives	*All Platforms*	Review of Country groupings/hierarchy table/Critical mass rule for region	H2 2020			~~X~~			X
Executives	*All Platforms*	*For all platforms: FAC 3 - Industry code 600 as example for discussion on relation between FAC and Industry codes*	~~H1 2020~~H2 2020			~~X~~		X	X
NB/ER	General	Data Quality Report - part 2	2020			~~X~~		X	X
ER	General	Provide Methcom with a new overview of the number of members who received data back for a certain Data Pool in the last 12 months, including: • Overview of complete years • Years before 2016 • Data replaced since start full submissions • Overview for PD & Rating and BP	H1 2020		X				
Executives	All Platforms	Calendar Methcom 2021	H2 2020						X
Executives	All Platforms	Forward looking calendar Methcom 2021	H2 2020						X

Item 4

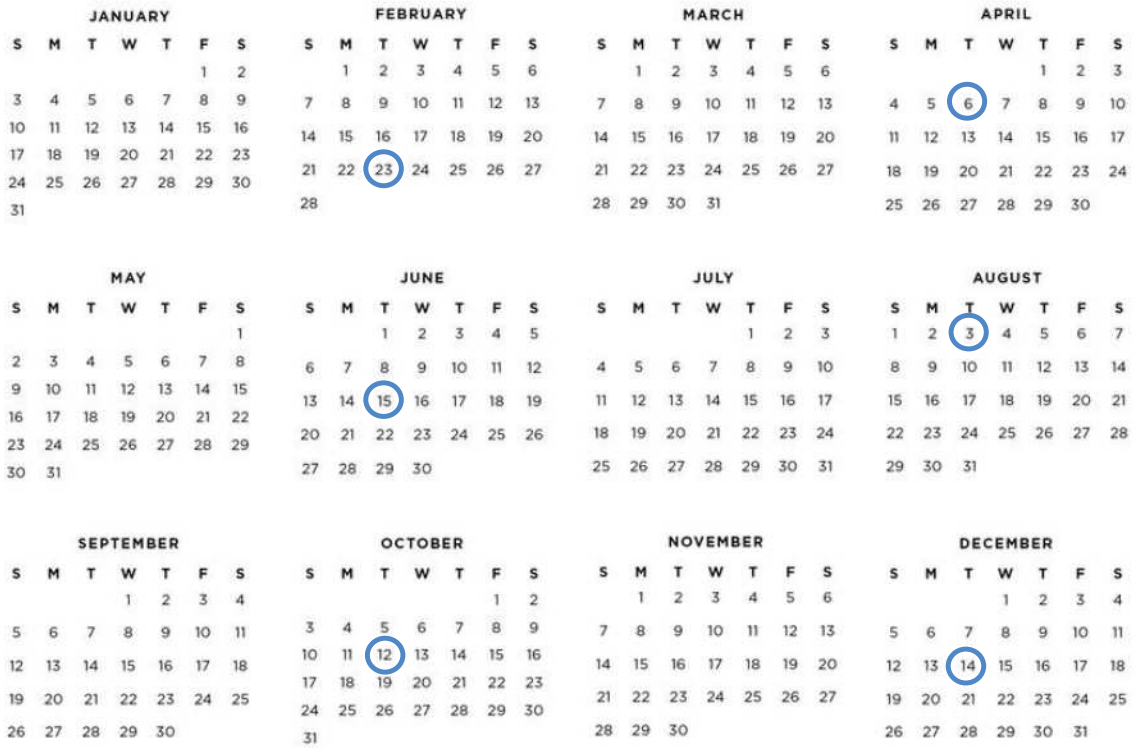
Request to Methcom

Action Requested	• Methcom to **approve** meeting calendar 2021
Background	➢ In December 2020 the calendar is shared for member approval

Copyright The Global Credit Data Consortium 2020 all rights reserved. Confidential

Global Credit Data

Meeting calendar



Mark your calendar
23 February 2021
6 April 2021
15 June 2021
3 August 2021
12 October 2021
14 December 2021

Copyright The Global Credit Data Consortium 2018 all rights reserved. Confidential

Global Credit Data

Which platform do you prefer?

Copyright The Global Credit Data Consortium 2018 all rights reserved. Confidential

Global Credit Data

Forward looking calendar 2021 - gray items have been changed from last FLC - blu items have been addedd since last FLC

Standing Items			Implementation	Feb	April	Jun	Sep	Oct	Dec
ER	LGD	Outcome of LGD Subcommittee meeting		X		X		X	X
OP	PD &Rating	Outcome of PD & Rating Subcommittee meeting		X		X		X	
HT	BP	Outcome of BP Subcommittee meeting		X		X		X	
NR	General	Review Subcommittee Charters	H2 2021						X
DT	General	Review Data Pool Regulations	H1 2021	X	X	X			
Items									
NR/ER	LGD	Restructured Facilities	H1 2021		X				
ER	LGD	Major project: Documentation update	Ongoing	X				X	
NR	All Platforms	Assessment of Raw Industry Code	Ongoing	X					
NB	LGD	2021 Analytics and Report overview	2021		X		X		
NB/ER	General	Data Quality	2021		X				X
ER	General	Overview of members pariticipation in GCD data pools in the last 12 months	H1 2021		X			X	
Executives	All Platforms	Calendar Methcom 2021	H2 2021						X
Executives	All Platforms	Forward looking calendar Methcom 2021	H2 2021						X

Item 5

Global Credit Data
by banks for banks

Charter for GCD PD Subcommittee
(a subcommittee of the Methodology Committee)

Approved by the Methcom/Board/GMM

Background

The Working Group ODF has operated for many years in conjunction with the Methodology Committee to research template, documentation and data control and quality issues for the PD/ODF database. Even though it is named as a working group, this group does not have the specific deliverables, end date or budget that a Working Group is required to have and no decision power is documented, therefore a clarification of its status has been discussed during late 2016.

The governance of the Global Credit Data Consortium (GCD) is laid down in the Articles of Association and the bodies empowered by GCD's statutory documents are the General Meeting, the Board, the Methodology Committee and the Meetings of Data Pool members.

Therefore, the WG must be constituted as Methcom Subcommittee to empower the Working group ODF with specific responsibilities and decision powers and make a clear distinction between a Subcommittee and a Working Group.

This document sets out the mandate, composition of the Subcommittee, position of the Chair and procedures for the PD Subcommittee.

1. Definitions

1.1 **Member Documentation** refers to the Data Input Structure, Data Output Structure, Release notes, User guide and FAQ documents, or any other document with information for the members.

1.2 **Critical Mass Rules** refer to rules that are in place to preserve confidentiality of the Credit Data contributed to a Data Pool, e.g. at least three Pool-participants contribute data of comparable nature.

1.3 **Give to Get rules** refer to the fact that Members share data on the basis of give to get. The principles related to give to get are outlined in the Data Pool Regulations.

2. Mandate

2.1. The purpose of the PD Subcommittee is to work in conjunction with and support the Methodology Committee on topics related to the PD Data Template and data submission (see description of topics in the table below).

2.2. The Methodology Committee delegates decision power on topics where the role of the PD Subcommittee is described as 'Decision' in the table below.

2.3. The PD Subcommittee has the power to take a decision on a topic with the description 'Decision' in the table below and send the paper with the decision taken to the Methodology Committee for information. For the avoidance of doubt, the Methodology Committee may review any Decision and decide on amendments at any time.

2.4. The PD Subcommittee can create a Firm proposal on a topic with the description 'Firm proposal' and send it for approval to the Methodology Committee, meaning the proposal provided to the Methodology Committee has been discussed and endorsed by the PD Subcommittee.

2.5. In the table the responsibilities for the PD Subcommittee on specific topics are listed.

Topic	PD Subcommittee role
Methodological Issues • Product Alignment • Former Member Data • Complexity • Impact on Banks	Firm proposal
Data Model – Methodology Overview • High Level Methodology (includes changes in fields or new fields) • Calculated Fields (incl. aggregation)	Firm proposal
Data Model – Implementation Details • Input Structure • Validation Rules • Output structure • ~~Small t~~Technical, non-methodological implementation details	Decision
Member Documentation	Decision
Critical Mass Rules	Firm proposal
Give to Get Rules	Firm proposal
Scoring	Decision
Reports • Volumetrics and data overview • Peer Comparison Reports	Decision
Surveys	Decision

Formatted: Font: 12 pt

2.6. For the avoidance of doubt the PD Subcommittee does not constitute a meeting of Data Pool Participants as referred to in Article 8 of the Data Pool Regulations. Should the PD Subcommittee require decisions about special expenses (Article 2.6 of the Data Pool Regulations) or publication of Credit Data (Article 6.4 of the Data Pool Regulations) or any other matter that only a Data Pool can decide on, then it shall call a meeting of all Data Pool Participants.

3. Members of the PD Subcommittee

3.1. One delegate from each Member which is a Pool Participant in at least one of the PD Data Pools shall be a member in the PD Subcommittee.

3.2. The Methodology Committee shall appoint the Chair of the PD Subcommittee.

3.3. In case the Chair of the PD Subcommittee is not a member of the Methodology Committee, the Chair of the PD Subcommittee may participate in the Methodology Committee meeting, which is especially encouraged for matters originating from the PD Subcommittee.

3.4. The Chair of the PD Subcommittee can decide to send a representative to present the matters to the Methodology Committee. This can either be one of the members or a GCD Executive who is also participating in the PD Subcommittee.

4. Procedure

4.1. Meetings of the PD Subcommittee shall be held on a regular basis, as frequently as required.

4.2. Meetings may be held by telephone or by other electronic means. Members in the PD Subcommittee shall be entitled to invite guest delegates from their organisation to attend the meeting.

4.3. Notice of meeting shall be given at least a week in advance, with an agenda and relevant material provided as far in advance as possible, but not later than 2 business days before the meeting.

4.4. The quorum for taking a Decision or a decision for a Firm proposal shall be PD Subcommittee members representing 5 Members.

4.5. Decisions shall be taken where there is a quorum present and where a majority of the PD Subcommittee present agree.

4.6. An Executive shall act as secretary to PD Subcommittee meetings and produce a record of decisions made and anything else required by a PD Subcommittee member to be placed on record. These records of PD Subcommittee meetings shall be reviewed by the chairman of the PD Subcommittee and then provided to all PD Subcommittee members as soon as possible.

4.7. To ensure transparency, all agendas, material and minutes of the PD Subcommittee will be made available to all Methodology Committee Members. Material and minutes of the Methodology Committee relevant for the PD Subcommittee will be made available to the PD Subcommittee members.

4.8. The Executive responsible for the PD Subcommittee has the responsibility to share relevant information with the responsible Executive(s) of other subcommittees to ensure alignment of the databases. The Executive will also give an update, based on the material mentioned in paragraph 4.7, during the Methodology Committee meetings.

Approved by the Methodology Committee on ~~13 December 2019~~

Revision cycle: yearly.

Global Credit Data
by banks for banks

Charter for GCD EAD/LGD Subcommittee
(a subcommittee of the Methodology Committee)

Approved by the Methcom/Board/GMM

Background

The Data Quality Working Group has operated for many years in conjunction with the Methodology Committee to research template, documentation and data control and quality issues for the EAD/LGD database. Even though it is named as a working group, this group does not have the specific deliverables, end date or budget that a Working Group is required to have and no decision power is documented, therefore a clarification of its status has been discussed during late 2016.

The governance of the Global Credit Data Consortium (GCD) is laid down in the Articles of Association and the bodies empowered by GCD's statutory documents are the General Meeting, the Board, the Methodology Committee and the Meetings of Data Pool members.

Therefore, the WG must be constituted as Methcom Subcommittee to empower the Data Quality Working group with specific responsibilities and decision powers and make a clear distinction between a Subcommittee and a Working Group.

This document sets out the mandate, composition of the Subcommittee, position of the Chair and procedures for the EAD/LGD Subcommittee.

1. Definitions

1.1 **Member Documentation** refers to the Data Input Structure, Data Output Structure, Release notes, User guide and FAQ documents, or any other document with information for the members.

1.2 **Critical Mass Rules** refer to rules that are in place to preserve confidentiality of the Credit Data contributed to a Data Pool, e.g. at least three Pool-participants contribute data of comparable nature.

1.3 **Give to Get rules** refer to the fact that Members share data on the basis of give to get. The principles related to give to get are outlined in the Data Pool Regulations.

2. Mandate

2.1. The purpose of the EAD/LGD Subcommittee is to work in conjunction with and support the Methodology Committee on topics related to the EAD/LGD Data Template and data submission (see description of topics in the table below).

2.2. The Methodology Committee delegates decision power on topics where the role of the EAD/LGD Subcommittee is described as 'Decision' in the table below.

2.3. The EAD/LGD Subcommittee has the power to take a decision on a topic with the description 'Decision' in the table below and send the paper with the decision taken to the Methodology Committee for information. For the avoidance of doubt, the Methodology Committee may review any Decision and decide on amendments at any time.

2.4. The EAD/LGD Subcommittee can create a Firm proposal on a topic with the description 'Firm proposal' and send it for approval to the Methodology Committee, meaning the proposal provided to the Methodology Committee has been discussed and endorsed by the EAD/LGD Subcommittee.

2.5. In the table the responsibilities for the EAD/LGD Subcommittee on specific topics are listed.

Topic	EAD/LGD Subcommittee role
Methodological Issues • Product Alignment • Former Member Data • Complexity • Impact on Banks	Firm proposal
Data Model – Methodology Overview • High Level Methodology (includes changes in fields or new fields) • Calculated Fields (incl. aggregation) • Data – remapping/removal (Modifications after changes in input structure or VRs/ Obvious omissions in the data)	Firm proposal
Data Model – Implementation Details • Input Structure • Validation Rules	Decision
Member Documentation	Decision
Critical Mass Rules	Firm proposal
Give to Get Rules	Firm proposal
Scoring	Decision
Reports • Volumetrics and data overview • Peer Comparison Reports	Decision
Surveys	Decision

2.6. For the avoidance of doubt the EAD/LGD Subcommittee does not constitute a meeting of Data Pool Participants as referred to in Article 8 of the Data Pool Regulations. Should the EAD/LGD Subcommittee require decisions about special expenses (Article 2.6 of the Data Pool Regulations) or publication of Credit Data (Article 6.4 of the Data Pool Regulations) or any other matter that only a Data Pool can decide on, then it shall call a meeting of all Data Pool Participants.

3. Members of the EAD/LGD Subcommittee

3.1. One delegate from each Member which is a Pool Participant in at least one of the EAD/LGD Data Pools shall be a member in the EAD/LGD Subcommittee.

3.2. The Methodology Committee shall appoint the Chair of the EAD/LGD Subcommittee.

3.3. In case the Chair of the EAD/LGD Subcommittee is not a member of the Methodology Committee, the Chair of the EAD/LGD Subcommittee may participate in the Methodology Committee meeting, which is especially encouraged for matters originating from the EAD/LGD Subcommittee.

3.4. The Chair of the EAD/LGD Subcommittee can decide to send a representative to present the matters to the Methodology Committee. This can either be one of the members or a GCD Executive who is also participating in the EAD/LGD Subcommittee.

4. Procedure

4.1. Meetings of the EAD/LGD Subcommittee shall be held on a regular basis, as frequently as required.

4.2. Meetings may be held by telephone or by other electronic means. Members in the EAD/LGD Subcommittee shall be entitled to invite guest delegates from their organisation to attend the meeting.

4.3. Notice of meeting shall be given at least a week in advance, with an agenda and relevant material provided as far in advance as possible, but not later than 2 business days before the meeting.

4.4. The quorum for taking a Decision or a decision for a Firm proposal shall be EAD/LGD Subcommittee members representing 5 Members.

4.5. Decisions shall be taken where there is a quorum present and where a majority of the EAD/LGD Subcommittee present agree.

4.6. An Executive shall act as secretary to EAD/LGD Subcommittee meetings and produce a record of decisions made and anything else required by a EAD/LGD Subcommittee member to be placed on record. These records of EAD/LGD Subcommittee meetings shall be reviewed by the chairman of the EAD/LGD Subcommittee and then provided to all EAD/LGD Subcommittee members as soon as possible.

4.7. To ensure transparency, all agendas, material and minutes of the EAD/LGD Subcommittee will be made available to all Methodology Committee Members. Material and minutes of the Methodology Committee relevant for the EAD/LGD Subcommittee will be made available to the EAD/LGD Subcommittee members.

4.8. The Executive responsible for the EAD/LGD Subcommittee has the responsibility to share relevant information with the responsible Executive(s) of other subcommittees to ensure alignment of the databases. The Executive will also give an update, based on the material mentioned in paragraph 4.7, during the Methodology Committee meetings.

Approved by the Methodology Committee on ~~13 December 2019~~

Revision cycle: yearly.

Global Credit Data
by banks for banks

Charter for GCD Benchmark Subcommittee
(a subcommittee of the Methodology Committee)

Approved by the Methcom/Board/GMM

Background

The Working Group Benchmark recently started in conjunction with the Methodology Committee to research template, documentation and data control and quality issues for the Benchmarking database. Even though it is named as a working group, this group does not have the specific deliverables, end date or budget that a Working Group is required to have and no decision power is documented, therefore a clarification of its status has been discussed during late 2016.

The governance of the Global Credit Data Consortium (GCD) is laid down in the Articles of Association and the bodies empowered by GCD's statutory documents are the General Meeting, the Board, the Methodology Committee and the Meetings of Data Pool members.

Therefore, the WG must be constituted as Methcom Subcommittee to empower the Working group Benchmark with specific responsibilities and decision powers and make a clear distinction between a Subcommittee and a Working Group.

This document sets out the mandate, composition of the Subcommittee, position of the Chair and procedures for the Benchmark Subcommittee.

1. Definitions

1.1 **Member Documentation** refers to the Data Input Structure, Data Output Structure, Release notes, User guide and FAQ documents, or any other document with information for the members.

1.2 **Critical Mass Rules** refer to rules that are in place to preserve confidentiality of the Credit Data contributed to a Data Pool, e.g. at least three Pool-participants contribute data of comparable nature.

1.3 **Give to Get rules** refer to the fact that Members share data on the basis of give to get. The principles related to give to get are outlined in the Data Pool Regulations.

2. Mandate

2.1. The purpose of the Benchmark Subcommittee is to work in conjunction with and support the Methodology Committee on topics related to the Benchmark Data Template and data submission (see description of topics in the table below).

2.2. The Methodology Committee delegates decision power on topics where the role of the Benchmark Subcommittee is described as 'Decision' in the table below.

2.3. The Benchmark Subcommittee has the power to take a decision on a topic with the description 'Decision' in the table below and send the paper with the decision taken to the Methodology Committee for information. For the avoidance of doubt, the Methodology Committee may review any Decision and decide on amendments at any time.

2.4. The Benchmark Subcommittee can create a Firm proposal on a topic with the description 'Firm proposal' and send it for approval to the Methodology Committee, meaning the proposal provided to the Methodology Committee has been discussed and endorsed by the Benchmark Subcommittee.

2.5. In the table the responsibilities for the Benchmark Subcommittee on specific topics are listed.

Topic	Benchmark Subcommittee role
Methodological Issues • Product Alignment • Former Member Data • Complexity • Impact on Banks	Firm proposal
Data Model – Methodology Overview • High Level Methodology (includes changes in fields or new fields) • Calculated Fields (incl. aggregation)	Firm proposal
Data Model – Implementation Details • Input Structure • Validation Rules	Decision
Member Documentation	Decision
Critical Mass Rules	Firm proposal
Give to Get Rules	Firm proposal
Scoring	Decision
Reports • Volumetrics and data overview • Peer Comparison Reports	Decision
Surveys	Decision

2.6. For the avoidance of doubt the Benchmark Subcommittee does not constitute a meeting of Data Pool Participants as referred to in Article 8 of the Data Pool Regulations. Should the Benchmark Subcommittee require decisions about special expenses (Article 2.6 of the Data Pool Regulations) or publication of Credit Data (Article 6.4 of the Data Pool Regulations) or any other matter that only a Data Pool can decide on, then it shall call a meeting of all Data Pool Participants.

3. Members of the Benchmark Subcommittee

3.1. One delegate from each Member which is a Pool Participant in at least one of the Benchmark Data Pools shall be a member in the Benchmark Subcommittee.

3.2. The Methodology Committee shall appoint the Chair of the Benchmark Subcommittee.

3.3. In case the Chair of the Benchmark Subcommittee is not a member of the Methodology Committee, the Chair of the Benchmark Subcommittee may participate in the Methodology Committee meeting, which is especially encouraged for matters originating from the Benchmark Subcommittee.

3.4. The Chair of the Benchmark Subcommittee can decide to send a representative to present the matters to the Methodology Committee. This can either be one of the members or a GCD Executive who is also participating in the Benchmark Subcommittee.

4. Procedure

4.1. Meetings of the Benchmark Subcommittee shall be held on a regular basis, as frequently as required.

4.2. Meetings may be held by telephone or by other electronic means. Members in the Benchmark Subcommittee shall be entitled to invite guest delegates from their organisation to attend the meeting.

4.3. Notice of meeting shall be given at least a week in advance, with an agenda and relevant material provided as far in advance as possible, but not later than 2 business days before the meeting.

4.4. The quorum for taking a Decision or a decision for a Firm proposal shall be Benchmark Subcommittee members representing 5 Members.

4.5. Decisions shall be taken where there is a quorum present and where a majority of the Benchmark Subcommittee present agree.

4.6. An Executive shall act as secretary to Benchmark Subcommittee meetings and produce a record of decisions made and anything else required by a Benchmark Subcommittee member to be placed on record. These records of Benchmark Subcommittee meetings shall be reviewed by the chairman of the Benchmark Subcommittee and then provided to all Benchmark Subcommittee members as soon as possible.

4.7. To ensure transparency, all agendas, material and minutes of the Benchmark Subcommittee will be made available to all Methodology Committee Members. Material and minutes of the Methodology Committee relevant for the Benchmark Subcommittee will be made available to the Benchmark Subcommittee members.

4.8. The Executive responsible for the Benchmark Subcommittee has the responsibility to share relevant information with the responsible Executive(s) of other subcommittees to ensure alignment of the databases. The Executive will also give an update, based on the material mentioned in paragraph 4.7, during the Methodology Committee meetings.

Approved by the Methodology Committee on ~~13 December 2019~~

Revision cycle: yearly.

Item 6

N.B.

Item 6 is composed by the DQ Management policy and the DQ Dashboard . Below is reported the DQ Management policy.

The DQ Dashboard has been provided as a separated PDF file enclosed to the mail with the material.

Request to Methcom

Action Requested	MethCom **to approve** the Data Quality Dashboard and **to review** the GD data Quality Policy
Background and Status	➢ The Data Quality Project is part of the Compliance Working group. Delivering to members evidences and elements of documentation to proof to members that the GCD data used is compliant with internal and external requirements, e.g., regulatory requirements for banks using the data for IRB modelling purposes ➢ Focus on deliverable on two main elements: ➢ Public Data Quality Dashboard ➢ Work on Data Quality metrics ➢ Finalizing concrete metrics and KPIs ➢ Data Quality Policy ➢ Structuring work is done on the existing data quality processes ➢ Under Methcom review Data quality Dashboard is sent in a separate PDF with the whole deck

Global Credit Data

Data Quality Management Policy

(to be) Approved by the Board of Director
on 12 2020

Global Credit Data
by banks for banks

Table of Contents

INTRODUCTION 4

Key principles, goals and guidelines for GCD DQM Policy 4

References for relevant global DQM regulatory requirements 4

GCD REQUIREMENTS FOR DATA QUALITY MANAGEMENT 5

Why GCD must comply with regulatory standards for DQM 5

GCD Data quality management process 6

Criterions and standards for data quality 6

Reporting and documentation of data quality 6

Requirements from GCD governance 7

DATA QUALITY MANAGEMENT AND GOVERNANCE FRAMEWORK 8

DATA LIFE CYCLE – FROM MEMBER INPUT DATA TO MEMBER OUTPUT DATA AND DATA QUALITY CHECKPOINTS 10

DATA QUALITY DIMENSIONS, DEFINITION AND COMPLIANCE WITH REGULATORY REQUIREMENTS 11

ENSURING THE DATA QUALITY DIMENSIONS 15

Data quality dimensions ensured by processes 15

Data quality dimensions ensured by quantitative DQ checks and metrics 15

ROLES AND RESPONSIBILITIES 17

DATA QUALITY MANAGEMENT – GCD PROCESSES AND PROCEDURES 18

Change process of DQ requirements and checks 18

Trigger/ frequency 18

Involved roles 18

Process steps and description 18

Data quality control process 18

Trigger/ frequency 18

Involved roles 18

Process steps and description 18

Further non-process related data quality control procedures 19

Service level agreements for in- and output data 20

Process to escalate data quality issues 20

Trigger/ frequency 20

Involved roles 20

Process steps and description 20

Interplay of all DQ framework components – continuous and sustainable data quality improvement and monitoring procedure 20

DOCUMENTATION AND REPORTING 21

DQ requirements and checks 22

Documentation of DQ issues 22

Documentation of results and DQ reporting/ scoring 22

MAINTENANCE OF THIS DOCUMENT 23

APPENDIX 1: DATA GOVERNANCE BODIES 24

APPENDIX 2: DQ METRICS AND EXAMPLE VALUES 24

Comparability 24

Timeliness 26

Consistency/ stability 27

Correctness 29

Completeness 29

APPENDIX 3: CORRECTNESS CATEGORIES 30

APPENDIX 4: TOTAL DATA QUALITY SCORE 32

INTRODUCTION

Data quality management (DQM) has been a priority for the Global Credit Data Consortium (GCD) since its creation in 2004. The relentless specification and improvement of DQM rules has allowed for a continuous enrichment of the GCD data bases. As a result, GCD provides the most consistent and detailed data available for industry wide appraisal of defaulted bank loans' recoveries and losses.

This ***GCD DQM Policy*** first states the ***key principles and goals*** for monitoring the quality of data directly collected from member banks and ultimately pooled by GCD. We then review the relevant ***DQM regulatory requirements*** impacting member banks and GCD globally. These elements, together with ***GCD Governance rules,*** are defining each step of the ***GCD DQM operative process***.

By completing additional and readily available material[1] this *GCD DQM Policy* intends to help GCD data users demonstrate the added value and the regulatory compliance of the pooled data.

Key principles, goals and guidelines for GCD DQM Policy

The GCD DQM Policy aims to **continuously and sustainably** pursue the following goals:

G1 Delivery of high-quality and compliant data to member banks with efficient internal DQM
G2 Collection of high-quality data from member banks bound by service level agreements
G3 Continuous improvement of data quality via measurement and feedbacks to member banks
G4 Worldwide promotion/communication of the effective leading quality of the GCD data

While respecting the following key guidelines:

GL1 DQ issues should be monitored as early as possible throughout the data processing chain
GL2 DQ **remediation** on **input data** is **performed exclusively by the members originating the data**

References for relevant global DQM regulatory requirements

We discuss here the DQM dimension in the global regulatory context in which member banks operate and its consequences for GCD's data collection and processing framework as it targets to achieve GCD DQM goals G1 and G2.

Amongst the important consequences from the 2008 Global Financial Crisis was the reckoning of the central importance of information technology and data architecture for effective and comprehensive control of financial risks at institutions. In 2013 the global standard "Principles for the effective aggregation of risk data and risk reporting" was published by the Basel Committee on Banking Supervision (BCBS 239) and shortly after adapted in the data quality management requirements of the "Capital Requirements Regulation" (CRR) of the European Union.

With regard to data quality management, the requirements of the CRR are set out in the "Guidelines on PD estimation, LGD estimation and the treatment of defaulted exposures" (EBA / GL / 2017/16), the "Guidelines for the estimation of LGD appropriate for an economic downturn ('Downturn LGD estimation ') "(EBA / GL / 2019/03) and in the" Regulatory Technical Standards on the specification of

[1] See: ***GCD Compliance Report (FCG)*** and accompanied ***Gross Regulatory List GCD excel file***, and ***GCD Data Pool Regulations***. Available upon your request via your dedicated GCD Executive.

assessment methodology for IRB "(EBA / RTS / 2016/03) of the EBA and, furthermore, in the "ECB guide to internal models" (European Central Bank 2019/10).

Other regulatory authorities around the world have also formulated supervisory guidelines mentioning DQM's requirements for the use of external data, and for live, bank-wide processes to monitor data quality and continuously and sustainably improve it. See, for example, the US Federal Reserve regulation letter SR11-7.

Aware of this context, this policy defines clear and adequate data quality standards established and respected by the GCD and member banks.

GCD REQUIREMENTS FOR DATA QUALITY MANAGEMENT

Why GCD must comply with regulatory standards for DQM

Although GCD itself is not a financial institution subject to being regulated, its main purpose is nonetheless to provide its member banks with compliant and high-quality data eligible for compliant risk modelling purposes under BCBS's IRB regimes and other regulatory and accounting regimes for stress testing, expected credit losses projections, credit-market risks valuations, etc.

In fact, GCD data is used by members for a variety of purposes including and not limited to capital modelling under the Basel rules and Expected Loss Provision modelling under accounting rules (IFRS9, CECL). Yet, the detailed rules around collection of historical data (internal and external) can vary greatly between institutions and therefore **GCD needs also to collect[2] evidences of the rule set applied by each member bank**. GCD data quality standards have been developed by practitioners from our member banks over the past decade to meet the requirements of regulatory, business and accounting purposes. The GCD philosophy is to return granular loan level data to members so that they can deal with the data and make calculations in compliance with the rules affecting them.

The regulatory requirements mentioned in introduction affect all data involved in GCD processes, from data acquisition and collection, throughout data processing by GCD and final distribution back to the member banks. For example the ECB guide to internal models explicitly states that "if institutions use data provided by third parties, the **ECB considers it good practice for them to ensure that the third party has data quality processes** in place to ensure the accuracy, completeness and appropriateness of the data provided."[3]

Therefore, specific functions, roles and responsibilities are designated at GCD and within member banks to ensure compliance with the DQM standards. In the event of non-compliance, appropriate measures are taken. Data quality measures and their effects are adequately documented and reported (per GCD's objective G3).

[2]Evidences of banks' internal data quality processes as they may impact data provided to GCD if applicable, e.g. internal DQ policies, DQ sign-off, internal DQ audit report relevant for GCD.
[3]https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.guidetointernalmodels_consolidated_201910~97fd49fb08.en.pdf ECB guide to internal models

GCD Data quality management process

Pursuing objective G3 for continuous assessment and monitoring of the data quality a precise scope as well as adequate measurements and standards for data quality are transparently defined. Competencies and responsibilities in roles in order to monitor the data quality and its reporting are nominated. Furthermore, based on defined criteria and threshold values, the conformity of the data with the member banks' data quality standards is determined, and appropriate measures to improve the data quality initiated are implemented.

For a description of the data management processes at GCD, we refer to *chapters Ensuring the data quality dimensions* and *DQM process and procedures*.

Criterions and standards for data quality

GCD is formulating adequate criteria and standards relating to data quality. They apply to internal, external as well as pooled data that are involved in credit risk measurement. The criteria should be determined in a sensible and expedient manner and are adequately placed in the context of the GCD data architecture which is presented with all relevant data flows and processing steps.

The standards for data quality established in this way serve transparency and are accessible and documented. In addition, DQ assessment in the form of controls and measurements are setup.

For details on the GCD criterions and standards for data quality, we refer to the chapter

Data quality dimensions, definition and compliance with regulatory requirements.

Reporting and documentation of data quality

The data quality determined based on the specified measurements, criteria and standards are documented and reported at regular intervals. In addition, the reports on data quality can be created and made available on an ad hoc basis. In general, the facts reported are presented transparently and in a way that is appropriate for the target group. The scope of the GCD reports, findings of data quality deficiencies as well as recommended measures for their elimination should be sufficiently specified, documented and included in the data quality reports. It must be ensured and documented that data quality deficiencies are reported to the responsible persons. In addition, the measures taken, and their effects should be shown in the data quality reports.

For details on reporting and relevant documentation at GCD, we refer to this policy chapters *Data quality control process* and *Process to escalate data quality issues*. On the side of the business and IT divisions roles and the associated responsibilities and competencies within the data quality management process are defined. Among other things, these roles are responsible for monitoring and reporting the data quality, as well as addressing and initiating suitable measures in the event of detected deviations in the data quality from the standards of the bank.

For details on roles and responsibilities at GCD, we refer to chapter *Roles and responsibilities*.

To continuously reinforce GCD's consolidated DQM, **GCD is considering the possibility to also collect DQM internal evidences** directly from members via appropriate internal measurements and reports.

Requirements from GCD governance

The non-profit association Global Credit Data is governed by two main documents, the Global Credit Data Consortium **Articles of Association**[4] and the Global Credit Data Consortium **Data Pool Regulations**[5]. Particularly the latter governance document formulates requirements with respect to data quality in its **Article 4.8**. Each member shall:

(a) "use a due standard of care in the preparation and submission of Credit Data to the Association and ensure that data submitted shall be at minimum up to the standard of the data used by the member for their own risk measurement and management".
(b) take all commercially reasonable efforts to gather and submit all the Credit Data that meet the criteria of the Data Pool in which it is a Pool-participant.
(c) ensure that the Credit Data submitted, if incomplete for a given year for a given geographic area, is not biased in respect of PD, LGD or EAD by reason of such incompleteness.
(d) control their own sharing of Credit Data and methods to the Association and to other members of the Association such that their company's confidentiality rules are not compromised and their portfolio position is not unduly exposed

Non-functional requirements

By design, GCD DQM framework is meeting the following non-functional DQM requirements:

- Evaluability: The results of DQ-checks/ reports are available for systematic evaluation.
- Homogeneity: Test results are homogeneous, i.e. the same information should be provided for the different tests.
- Central storage and historization of DQ-check results: DQ-check results and underlying data are centrally persisted in order to be able to guarantee transparency and traceability.
- Flexible enhancement of DQ-checks: It must be possible to enhance or remove existing DQ-checks in a flexible way without high implementation effort.

[4] See: **Articles of Association The Global Credit Data Consortium**. Available upon your request via your dedicated GCD Executive.
[5] See: **GCD Data Pool Regulations**. Available upon your request via your dedicated GCD Executive.

DATA QUALITY MANAGEMENT AND GOVERNANCE FRAMEWORK

To ensure the quality of the data GCD established and implemented an effective data quality management framework which developed since its existence over the last 15 years. This complies to the regulatory standard as described in this policy's chapter Why GCD must comply with regulatory standards for DQM. Below the overview of the data quality management framework is given in Figure 1. The key components of the DQM framework are:

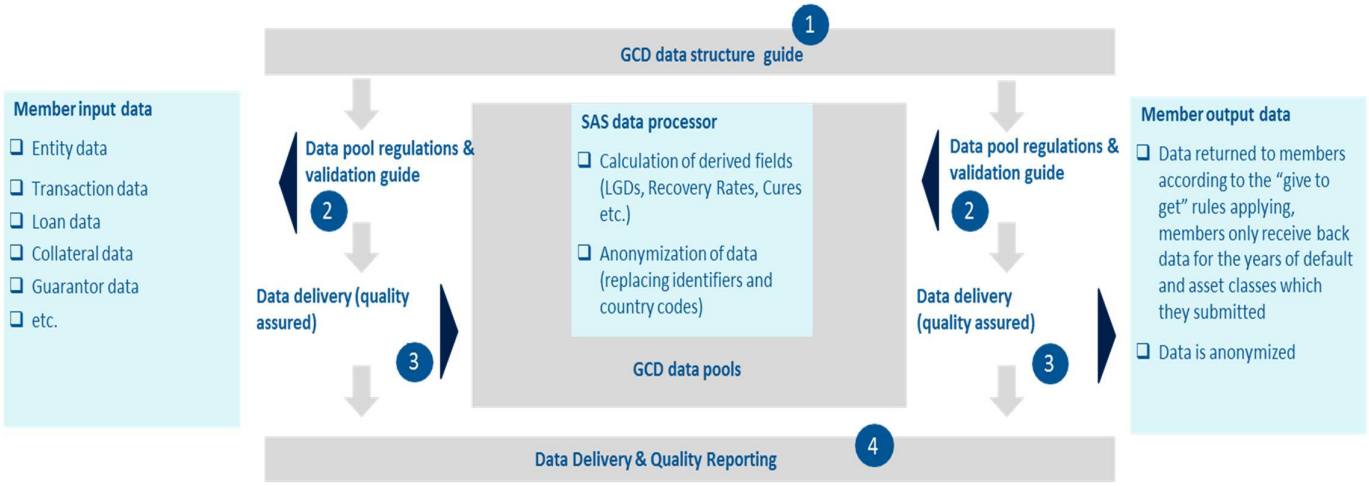


Figure 1 – Overview data processing and data management framework

1. In the **GCD data structure and validation guide** [Excel document] the following information is defined:
    a. Entities (loan, transaction, etc.) and their attributes at the business level
    b. Metadata for entities and attributes (mandatory, classification...) and for specific information required for input and output fields

2. **Data pool regulations** are used to define requirements and responsibilities between member banks and GCD (see Service level agreements for in- and output data):
    a. Requirements are related to the quality of data, delivery times, etc.
    b. Requirements are centrally recorded by GCD in the **validation guide** and approved by Methodology Committee (read more about the role of the Methodology Committee in chapter ROLES AND RESPONSIBILITIES and also GCDs data governance bodies in Appendix 1)

Appropriate **data quality checks** are defined, implemented and carried out at well-defined **checkpoints** (see

3. DATA LIFE CYCLE – FROM MEMBER INPUT DATA TO MEMBER OUTPUT DATA AND DATA QUALITY CHECKPOINTS), particularly at data input (validation rules (VR) and audit) and output to member banks (DQ scoring):

Data Quality checks are steered by Data Quality Dimensions (see chapter

Data quality dimensions, definition and compliance with regulatory requirements)

a. The member banks ensure the requirements according to the data pool regulations
b. Information on the data quality checks and audits is stored
c. Data quality checks are clearly assigned to the requirements

4. **Reporting** ensures the necessary transparency with regard to data deliveries and especially with regard to their quality
    a. Metadata and the results of data quality checks and audits are integrated for this purpose
    b. The reporting is distributed to all stake holders including member banks (audit letter, DQ scoring)

The data quality governance framework is depicted in Figure 2[6]. It shows all involved stakeholders, their interactions and roles. The interaction with the GCD board in case of escalation of DQ issues is faded out because this is only a formal measure which did not occur throughout the history of GCD.

For further details on roles and responsibilities and processes we refer to Chapter Roles and responsibilities and chapter Data quality management – GCD processes and procedures, respectively.

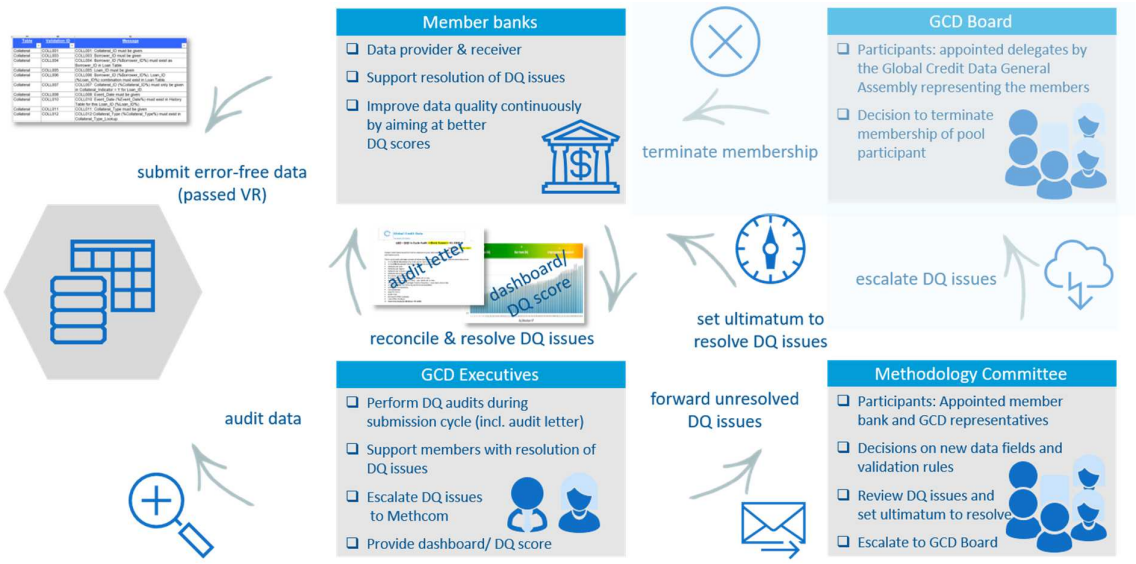


Figure 2 – Overview data quality governance framework

[6] Note: This is not a process chart and, hence, has no starting/ending point.

DATA LIFE CYCLE – FROM MEMBER INPUT DATA TO MEMBER OUTPUT DATA AND DATA QUALITY CHECKPOINTS

In the Data Quality Management Framework to come from the Member Input data to the Member output data there are data quality checks, data flows and processing steps in place which needs to be considered in detail. The guideline is followed that data quality issues should be detected and corrected as early as possible in the processing chain. Hence, an upload of member data is only possible after passing an encompassing set of validation rules followed by an audit by GCD Executives.Pre-submission audit reports are delivered to members for their preparation of the data submission.

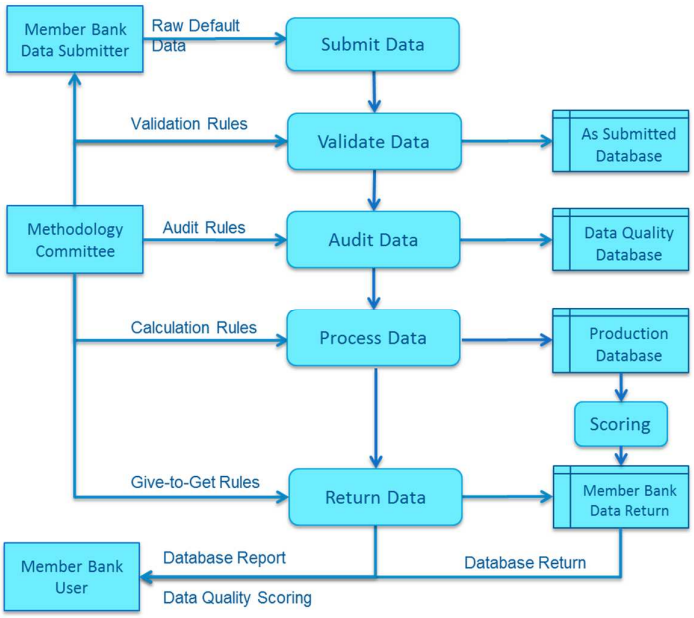


Figure 3 – Internal GCD Data flow and processing, after data submission by member

Figure 3 depicts the data flow and processing. After a member submits data (input data), an automated execution of validation rules is triggered. Further processing is only possible after all validation rules are successfully passed. Afterwards an audit by GCD executives is performed and data inconsistencies are communicated to the members and requested to resolve. After a successful audit the data gets aggregated via a standardized and stable ETL and derived fields are calculated. Before the data is returned to the member banks (output data) a DQ scoring based on metrics capturing the relevant data quality dimensions is created. A comparison with other member banks data is included. For further details we refer to chapter *Data quality dimensions, definition and compliance with regulatory requirements.*

The Member Bank, Methodology Committee and GCD executives and the processes & procedures around validating, auditing and scoring the data, play an important role in the process of continuously improving the data quality. Please refer to the summary in *Interplay of all DQ framework components – continuous and sustainable data quality improvement and monitoring procedure.*

Data quality dimensions, definition and compliance with regulatory requirements

GCD adheres to the best practices by defining key principles for DQM in consistence with the BCBS 239 recommendation:

- Principle 3: **Accuracy and Integrity**: Risk data needs to be accurate and reliable, with controls as strong as those for accounting data. Risk data needs to be reconciled with sources and validated with a "balance between automated and manual systems".
- Principle 4: **Completeness**: All material risk data needs to be aggregated. Banks need to monitor their risk data for completeness.
- Principle 5: **Timeliness**: Risk reporting and aggregation needs to be produced in time to meet risk management needs of the bank.
- Principle 6: **Adaptability**: Risk data aggregation capabilities need to be flexible enough to meet ad hoc requests including scenario analysis or country specific analysis or risk.

GCD has identified that all the above principles are broadly applicable, by analogy, to GCD's LGD/EAD data collection, aggregation, returns and reports. Specific comments on GCD's compliance with principles 3 to 6 are covered below:

- Principle 3: **Accuracy and Integrity**: GCD's data integration is performed in an accurate and reproducible way. The process accurately transforms the new input data into the total data pool, without losing any data points. GCD's principles of not amending, adding or removing any information, but only allowing such changes to be made by the members, makes the entire process reconcilable.

  Accuracy of calculation is ensured by the documentation and transparency of the calculations made. In the LGD/EAD database every calculation made by GCD, e.g. LGD or Time to Resolution, can be replicated by the banks who receive the data in return. GCD is not able to confirm the accuracy of the data input by the members, as only their auditors can do this. However, GCD does perform both hard technical rule checks and softer expert reasonableness checks on the data input and then rejects the non-compliant data and asks banks to re-input their data, (see later sections).

- Principle 4: **Completeness**: An interpretation of what completeness means for GCD's LGD/EAD data could be that GCD should receive a complete delivery of all defaulted cases in each facility asset class from each of the banks delivering to that facility asset class. In addition, the data fields for each of the contributed borrowers and loans should be complete. GCD ensures this completion by setting certain fields as mandatory and not accepting data which does not contain sufficient information.
- Principle 5: **Timeliness**: GCD's process workflow ensures a timely aggregation of all input data provided by the member banks.
- Principle 6: **Adaptability**: GCD has a well-defined data architecture, which includes a detailed data model that allows for flexible adhoc analyses such as by country code, etc.

In conclusion, GCD's data architecture and process workflow ensures by its design the compliance with BCBS 239 w.r.t. data aggregation.

In the ECB guide to internal models the following data quality dimensions have been explicitly mentioned:

(a) Completeness - Values are present in the attributes that require them
(b) Accuracy - Data is substantially error free
(c) Consistency - A given set of data can be matched across different data sources of the institution
(d) Timeliness - Data values are up to date
(e) Uniqueness - Data is free from any duplication from filters or other transformations of source data
(f) Validity - Data is founded on an adequate and rigorous classification system
(g) Availability/Accessibility - Data is made available to the relevant parties
(h) Traceability - The history, processing and location of the data under consideration can be easily traced

Though there is some overlap with BCBS 239, the data quality dimensions in ECB guide add further aspects of data quality. Also w.r.t. to the further data quality dimensions, GCD's data architecture is compliant:

- (c) Consistency – GCD ensures by its data processing steps that data is always calculated the same way within a dataset and also across different data sets over time.
- (e) Uniqueness – There are automated checks for duplicates performed during upload of member input data. Duplicates are not allowed to enter the system.
- (f) Validity – Further there are automated checks that ensure validity of data for example w.r.t. specified value ranges for attributes.
- (g) Availability/Accessibility – GCD has a clear process that provides availability both to responsible executives at the different stages during the submission process and to member banks.
- (h) Traceability – Also w.r.t. traceability the data processing process ensures that at each step it is clear how information is being transformed and which information enters into which data aggregation.

For the use of external data the ECB guide further states that for the use of external pool data, comparability with own processes and data needs to be ensured.

The US Federal Reserve in its supervisory regulation letters (SR11-7) makes clear demands on the use of external data (and models) for benchmarking in https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf

> *"Benchmarking is the comparison of a given model's inputs and outputs to estimates from alternative internal or external data or models. It can be incorporated in model development as well as in ongoing monitoring. For credit risk models, examples of benchmarks include models from vendor firms or industry consortia and data from retail credit bureaus. Pricing models for securities and derivatives often can be compared with alternative models that are more accurate or comprehensive but also too time consuming*

to run on a daily basis. Whatever the source, benchmark models should be rigorous and benchmark data should be accurate and complete to ensure a reasonable comparison."

Combining and enhancing all requirements from BCBS 239, ECB guide to internal models and SR 11/7, one arrives at the following set of data quality dimensions and definitions:

Data quality dimensions	Definition	Regulatory source	Is GCD compliant?
Completeness	Completeness measures the degree of population of data values in a data set. Values are present in the attributes that require them.	BCBS 239, ECB guide, SR 11/7	Yes, Yes, Yes
Accuracy	Accuracy measures the degree of error-freeness of a data set. Further, Accuracy measures the degree to which data factually represents its associated real-world object, event, concept or alternatively matches the agreed upon source(s).	BCBS 239, ECB guide, SR 11/7	Yes, Yes, Yes
Consistency/Stability	Consistency refers to the fact that data values are consistently represented within a dataset and between datasets and consistently associated across data sets. It also refers to the composition of datasets across time (stability over time).	ECB guide	Yes
Timeliness	Timeliness is a measure of time between when data is expected versus made available. Data values are up to date.	BCBS 239, ECB guide	Yes, Yes
Uniqueness	Data is free from any duplication from filters or other transformations of source data. Uniqueness states that no entity exists more than once within the dataset. A key value relates to each unique entity and only that specific entity within the dataset.	ECB guide	Yes
Validity	Validity measures whether a value conforms to a preset standard (e.g. value	ECB guide	Yes

	range, precision, aggregation level). Data is founded on an adequate and rigorous classification system.		
Availability/Accessibility	Accessibility measures how easy it is to acquire data when needed, how long it is retained, and how access is controlled.	ECB guide	Yes
Traceability	The history, processing and location of the data under consideration can be easily traced.	ECB guide	Yes
Integrity	Integrity measures the structural or relational quality (referential integrity) of data sets.	BCBS 239, ECB guide	Yes, Yes
Adaptability	Covers the appropriate information needed for certain type of analyses. Data to meet a broad range of on demand, ad hoc risk management reporting requests.	BCBS 239	Yes
Comparability and adequacy	Measures how data compares to other banks (e.g. average LGDs, time to recovery etc.) and how data from a certain pool is compatible with own data.	ECB guide	Yes
Degree of data quality culture established at member bank	Has the member bank established a BCBS 239 compliant data governance and data quality management framework? Does the member bank share details on processes such as definition of default? Does the member bank certify the completeness of provided data? Does the bank respond to audit requests?	-	Yes

Following closely the regulatory data quality dimensions ensures compliance with regulatory requirements. Furthermore, enhancements w.r.t. comparability and adequacy or the degree of data quality culture established at a member bank allows to consider the pooled nature of GCD data.

ENSURING THE DATA QUALITY DIMENSIONS

In this Chapter it is described how data quality is ensured at GCD.

The basis for ensuring data quality are the data quality dimensions defined before. The data quality dimensions are either ensured by processes itself or concrete data quality checks and metrics.

Data quality dimensions ensured by processes

One way to ensure data quality is by implementing processes which automatically cover some of the data quality dimensions introduced in the previous chapter. The following dimensions are covered by processes:

- **Accuracy:** Since one of GCD's principle for data processing is to do not modify any data along its processing workflows, it is implicitly ensured that all output data matches accurately the input data without any changes. Furthermore, all aggregations used to derive quantities such as ex-post LGDs or recovery rates went through a thorough testing process and are well-established among the member banks.
- **Availability/Accessibility:** GCD ensures availability and accessibility with its anonymization and data pool concept. Before data is returned to each member bank it undergoes an anonymization process such that information about other pool contributors is not visible. Furthermore, the give to get concept ensures that each member bank will receive only data from the pool the contributed to.
- **Traceability:** As shown in Figure 3, GCD's data processing workflow has at each processing step quality gates in order to follow guideline GL1 **Error! Reference source not found.** Hence, by construction of the data processing workflow, traceability is ensured.
- **Adaptability:** GCD provides a large number of attributes with a data history of over 20 years. Since the data model including every attribute underwent a thorough confirmation process by MethComm, it is ensured that all necessary information for highly individual risk modelling analyses in the context of LGD and EAD is available with a sufficient data granularity and history.

Data quality dimensions ensured by quantitative DQ checks and metrics

Besides a processual ensuring of data quality dimensions, GCD uses quantitative checks and metrics to ensure and measure data quality. For the following data quality dimensions DQ checks and metrics are used.

- **Uniqueness, Validity and Integrity:** GCD ensures and measures correctness of data by automated validation rules (the respective GCD DQ dimension is called Correctness). To enter the GCD database data must go through 450+ validations. This real-time validation process ensures consistency across the data pool, promotes transparency, shortens the learning curve and most importantly keeps the level of data quality high. Validation rules and methodologies have been developed and constantly reviewed by the Methodology Committee. These validations are documented extensively for easy consultation during the data submission process and reconciliation with the data definitions and objectives. Validation rules are built

into every field of every table in the web portal where information goes in and gives banks automated feedback on every rule broken and all warnings about data they have either filled in or forgotten to fill in. For example, simple but crucial checks are if the format is correct (e.g. in a date field) or the correct drop-down items from the lookup tables are used. But it is also checked if related fields are consistently filled. One example is if banks have submitted that they have real estate collateral they must fill in what type of real estate it is. Some validation rules are only triggered if another field is filled with certain information. For example: The transaction table records all cash flows between default and resolution. This can be incoming cash flows (e.g. principal payments or interest payments), outgoing cash flows (e.g. principal advances) or write -offs. For entries which are incoming cash flow types another field indicating the source of the payment must be filled.

- **Completeness:** The degree of completeness w.r.t. population of non-mandatory fields is evaluated calculating coverage ratios.
- **Consistency and Stability:** For GCD consistency and stability means how key data fields of resolved loans w.r.t. LGD modelling remain stable between two subsequent submissions. In order to measure consistency and stability two submissions need to be compared with each others. For details on the exact definition of the metric, we refer to [APPENDIX 2].
- **Timeliness:** For GCD timeliness means how timely resolved loans will be submitted. In order to measure timeliness two submissions need to be compared with each others. For details on the exact definition of the metric, we refer to [APPENDIX 2].
- **Comparability and Adequacy:** GCD uses the Hellinger distance measure to compare distributions of typical key risk drivers for LGD modelling such as time to recovery or cure rate of individual member banks with the overall pool. The resulting DQ scorings of different banks are anonymously compared with each others and returned back to the members. For details on the exact definition of the metric, we refer to [APPENDIX 2].
- **Building the Data Quality Culture at member bank, and at GCD:** GCD conducts regular surveys involving questionnaires with yes/no questions to assess how member banks improve their DQ culture and processes and how they manage to instill good practices and strong systems of internal controls built to ensure the integrity of data. This is supporting GCD's and member banks' changes to improve their culture of data management and quality.

ROLES AND RESPONSIBILITIES

Global Credit Data takes its responsibilities in line with GCD Data Pool Regulations and data quality seriously and requires its members to do the same. Article 4A of the GCD Data Pool Regulations sets out an assurance of data quality within the activities of the organisation.

Role	Responsibilities
Member bank	• Members are obliged to **ensure** that data they deliver to the Association is complete and of **good quality** • Each member shall carefully **consider the outcome of audits** and perform updates where required to deliver a high level of data quality • Where possible each member of a Data Pool is required to **re-submit all credit data** previously submitted at least **every three years** • Each member shall carefully **consider DQ scores** provided with pool data and work **to continuously improve the data quality**
GCD executive	• Executives have the right and the obligation to **perform in cycle and out of cycle audits** and share the results with the members in an audit letter. Review during the audits consists of both technical and expert review, including special analytical audits • Executives **support members to resolve DQ issues** • Executives **provide a dashboard and DQ score** indicating the data quality of the member • Executives may **refer unresolved/ critical DQ issues to the Methodology Committee** in such a way as to ensure anonymity of the pool-participant and objectivity of the Methodology Committee
Methodology Committee	• **Clarify escalated DQ issues** and **put in place sanctions/ warnings** if necessary • **Approve new or removed field** (values) in the data template **and** new or amended **validation and audit rules**
GCD Board	• **Decide** on suggestion by Methodology Committee **to terminate membership of a pool-participant on possible ground of longstanding unresolved GQM issues**

DATA QUALITY MANAGEMENT – GCD PROCESSES AND PROCEDURES

In this Chapter all processes and procedures associated with DQM are described. Processes do always have a sub section structure containing the sub sections Trigger/ frequency, Involved roles and Process steps and descriptions. Sections on procedures do not have such a sub section structure.

The Chapter is closed with bringing all process, procedures and elements together to arrive at a continuous procedure to monitor and improve data quality.

Change process of DQ requirements and checks

Trigger/ frequency

On demand: The change of DQ requirements is triggered by new fields or, in general, by measures to increase data quality and completeness, like, e.g., a further field becomes mandatory.

Involved roles

- GCD executives
- Methodology committee

Process steps and description

1. **Design**: A subcommittee existing of **GCD executives** carries out the design of **new/ changed validation/ audit rules or data fields**.
2. **Approval**: The **Methodology Committee** gives approval

Data quality control process

The elements for controlling data quality comprise both GCD's formal processes and coherence checks on the aggregated data outcome by members.

Trigger/ frequency

Each data submission cycle (twice a year). On demand audits are also possible.

Involved roles

- Member banks
- GCD executives

Process steps and description

1. **Data Delivery**: The first step in the GCD quality process is in the data delivery in which the **member banks** are only able **to submit their data** to GCD if it passes many strict quality tests (**validation rules**). The tests of course check for correct field formats but in addition to this they also perform completeness checks based on borrower, loan, collateral and guarantor level. There are also logic checks on each submission to make sure that the cash flows and accounting entries reported add up. For further information on validation rules, we refer to Data quality dimensions ensured by quantitative DQ checks and metrics.
2. **Data Audit**: Next step in the GCD quality process is the audit, this occurs after the delivery and before return of the data. In this phase a **GCD executive looks at the delivery for each bank** and reviews it for reasonableness and completeness. The executive then contacts banks where necessary to request clarifications or amendments. They also **issue an audit let**ter for each bank highlighting their weaknesses. It is only by using this experienced human expertise that data gaps can be determined and fixed. **Member banks** shall support to **resolve DQ**

issues. Examples include banks delivering data in which every field and every borrower passes the filters but which, when the patterns are examined, shows that they are entirely missing any reference to guarantors, for example. The member bank is then asked whether in their lending policy and country they do not use guarantors. A satisfactory answer must be received.

3. **Data Scoring**: Then **GCD executives continue with a data scoring system** focussing on "bad points" comprising for example missing "non-fatal" fields, e.g. industry code, and other common data quality issues that are measurable by the metrics mentioned in section Data quality dimensions ensured by quantitative DQ checks and metrics. The number of bad points (by reference to proportions of the delivery so as not to penalise large banks) are totalled and a ranking of all banks is made. **Member banks receive a detailed scoring** of their bad points as well as a list of scores by other banks (anonymised of course) both overall and for individual issues. In this way, a member bank can gauge the quality of the data by reference to their own known data quality and can judge the quality of individual elements with which they are concerned in their modelling. GCD uses this scoring and ranking to encourage banks to improve and to gauge the overall improvement in data quality over time.

Further non-process related data quality control procedures

Cleaning during RDS: Member banks are also advised and encouraged to examine the detailed raw data they receive in the twice-yearly data return and to produce their own "representative data set" or RDS which comprises a filtering of the data to ensure that it matches the portfolio of the bank which will be using it. This is normally done by limiting the data to only countries of interest, asset classes of interest and often particular drivers such as industry, seniority or security. During this process the final data cleaning is done such that data which is possibly wrongly classified or which may have data errors or which may be incomplete for this analysis, is removed.

Regular review: Furthermore, the GCD Articles of Association make clear that the organisation has the right to remove any data from the database if the GCD executives feel that it is not delivered in good faith and indeed as an ultimate sanction, GCD has the right to expel any member. This latter sanction has never been employed however GCD has in its history had the case of a member bank voluntarily removing data of insufficient quality after discussions with the executive.

Member data quality overview: All GCD member banks are subject to Basel II regulatory scrutiny. Any bank which uses GCD data for either model building or benchmarking is required to prove to its regulator that the data is of good quality and is appropriately representative. GCD data and the RDS produced therefrom, has been submitted by many banks to many regulators with very good results. The regulators are aware of the default levels of their regulated banks and can see whether the bank has submitted sufficient data to GCD. GCD data's granularity, size and source makes it the best quality pooled data available for PD, EAD and LGD.

On-site missions at member banks: GCD carries out on-site support to further enhance data quality by direct comparison with input data and/or interrogating responsible persons at member banks and reviewing internal/external audit reports [7].

[7] See: GCD Compliance Report (FCG). Available upon your request via your dedicated GCD Executive.

Service level agreements for in- and output data

Each member bank has to formally agree upon the *Articles of Association*[8] and the *Data Pool Regulations*[9], which are the central governance documents of GCD. Particularly in the Data Pool Regulations responsibilities w.r.t. to data quality are formulated. Article 4A of the GCD Data Pool Regulations sets out an assurance of data quality within the activities of the organisation. Members are obliged to ensure that data they deliver to the Association is of good quality. To support members with this, executives conduct audits to check the data before it is put in the database.

Furthermore, the GCD structure and validation guide sets up requirements on the attribute level w.r.t. to data format, validity, etc.

Process to escalate data quality issues

Trigger/ frequency

On demand: If the contribution of a member is not of sufficient quality.

Involved roles

- GCD executives
- Methodology Committee
- GCD Board

Process steps and description

1. **Inform Methodology Committee: GCD executives** will **report insufficient data quality of a member to the Methodology Committee** in such a way as to ensure anonymity of the Pool-participant and objectivity of the Methodology Committee [REFERENCE Data Pool Regulations 4A + 3.3]. Only if it is necessary for formal decision on data pool membership or at the request of the Pool-participant involved will the name of the Pool-participant be disclosed to the Methodology Committee.
2. **Review and set deadline to member bank**: If the **Methodology Committee finds**, at any time, **that a Pool-participant does not duly fulfil its Membership Obligations** with respect to the contribution of credit data, it shall give written notice thereof to the Pool-participant concerned, **demanding full compliance within a reasonable period of** time to specified in such notice but in any event within six months after the notice is given.
3. **Sanctions**: If after a reasonable period of time the Pool-participant still fails to fully comply with its Membership Obligations concerned, the **Methodology Committee shall propose** to the Board **to terminate the participation of the Pool-participant** in the Data Pool.

Interplay of all DQ framework components – continuous and sustainable data quality improvement and monitoring procedure

In this section it is described where the Data Quality Framework and its processes and policies come together in a continuous and sustainable improvement cycle of the data quality. In the figure below the cycle is shown.

[8] See: Articles of Association The Global Credit Data Consortium. Available upon your request via your dedicated GCD Executive
[9] See: GCD Data Pool Regulations. Available upon your request via your dedicated GCD Executive.

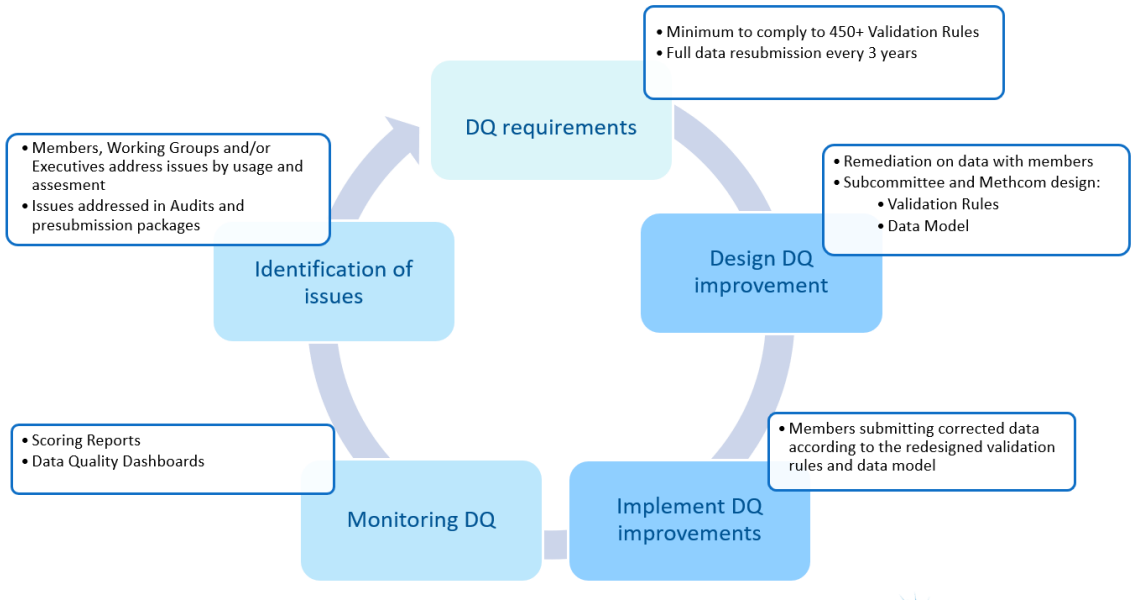


Figure 4 – GCD's data quality monitoring and improvement cycle.

- Starting with the *Identification of Issues by Audits* this occurs after the delivery and before return of the data. In this phase a GCD executive looks at the delivery for each bank and reviews it. The executive then contacts banks where necessary to request clarifications or amendments. They also issue an audit letter for each bank highlighting their weaknesses. (for details, please see *Data quality control process*)
- While members, *working groups and/or executives* are using the data, potential issues on the validation rules could be found and should be addressed to the executives.
- When, from these exercises, gaps in the data model and/or validation rules are observed. An improvement action will be initiated by the executives to start making a change to validation rules or the data models (see Section *Change process of DQ requirements and checks*). A change will impact the *minimum requirements* on the validation rule with which members have to comply.
- GCD executives provide support to the members on improving the data quality in the submission. The members work towards a submission complying up to the latest standards.
- Then GCD moves on to *monitoring* via the data scoring system which has been described in Section. For more details on the concrete DQ scoring reports, we refer to Section Documentation of results and DQ reporting/ scoring.
- The scoring and other measures are reflected on the Data Quality Dashboards for monitoring purposes.

DOCUMENTATION AND REPORTING

In this chapter documentation is described for the DQ requirements and checks as well as DQ findings and issues.

Further the reporting of DQ back to the member banks is detailed.

DQ requirements and checks

DQ requirements and checks are documented in the Data Structure and Validation Guide[10]. Here, the rules are listed on attribute level with a description and a trigger. See table below.

Furthermore audit rules that are used by executives during their data audits are documented in the same Data Structure and Validation Guide[11].

Table	Validation ID	Data field	Trigger	Message	Type
Entity	ENT015	Entity_Asset_Class	Entity_Asset_Class not in Entity_Asset_Class_Lookup	ENT015: Entity_Asset_Class (%Entity_Asset_Class%) must exist in Entity_Asset_Class_Lookup.	Error
Entity	ENT016	Public_Private_Indicator	Public_Private_Indicator is empty	ENT016: Public_Private_Indicator must be given.	Warning
...		...	...	...	...
History	HIST004	Loan_ID	Count {Loan_ID, Event_Type=2} > 1	HIST004: Loan_ID (%Loan_ID%) must have only one One Year Prior to Default event.	Error
...		...	...	...	...
Loan	LOAN024	Facility_Type	Facility_Type not in Facility_Type_Lookup	LOAN024: Facility_Type (%Facility_Type%) must exist in Facility_Type_Lookup.	Error

Documentation of DQ issues

DQ issues directly arising from validation rules during submission process will be documented in terms of status reports and send directly to the submitting member bank such that the data can be corrected because only data passing all validation rules is ready to be further processed.

DQ issues found in data audits will be addressed to member banks in form of audit letters.

Documentation of results and DQ reporting/ scoring

Based on its data submission each bank is assigned a single score. The scoring ranges from 0 (good) to 6 (bad). The scoring allows monitoring improvement on data quality by item, asset class and bank. 40% of the score, represents the correctness of the data; whether the data complies with the most recent validation rules. If an entry is new or resubmitted, the score is automatically 0. Another 40% of the score covers completeness checks. It scores if a bank submits optional fields or optional tables. The final 20% scores comparability. It compares banks' LGDs, Time to Recovery and Cure Rates and detects if a bank's submission is significantly different from the pool. Naturally, this must be a softer criterium as not all the banks are expected to be the same but it triggers a discussion on completeness of the submission if levels are very low or high compared to peers.

The scoring is a tool where the absolute numbers are not so important but rather tracking over time shows improvement in the overall dataset as well as by individual members.

[10] https://www.globalcreditdata.org/userguides/lgd-ead-platform-documentation LGD-EAD Platform Data Structure and Validation Guide.xlsx
[11] https://www.globalcreditdata.org/userguides/lgd-ead-platform-documentation LGD-EAD Platform Data Structure and Validation Guide.xlsx

MAINTENANCE OF THIS DOCUMENT

This document is subject to a regular review and update process in order to guarantee future compliance with international regulatory requirements and standards.

APPENDIX 1: DATA GOVERNANCE BODIES

Global Credit Data is governed by a Board, whose detailed responsibilities are defined in the GCD Articles of Association. The members of the Board are individuals appointed by the Global Credit Data General Assembly delegates representing the Members.

Specific activities around the collection, analysis and use of data are controlled by the Methodology Committee whose members are appointed by the Management Board. The detailed responsibilities of the Methodology Committee are further described in the GCD Articles of Association and GCD Data Pool Regulations [chapter reference]

APPENDIX 2: DQ METRICS AND EXAMPLE VALUES

Comparability

In accordance with Art. 24 and 25 of EBA GL 2016/17 institutions have to analyse the distribution and range of values of key risk characteristics statistically to ensure that the data used to develop a model is representative for the application portfolio.

Regulatory requirements like the above stimulate the necessity to include a measure for comparability as a separate data quality dimension. Comparability for GCD means how the distribution of key risk drivers w.r.t. LGD modelling purposes on a member bank's portfolio compare with the overall pool.

Data ensemble: In order to perform the comparison it is crucial to choose a reasonable data ensemble. For this purpose, only loans from active members, which are part of GCD's standard RDS will be considered. Further, the distribution properties will be compared per pool participant and facility asset class to arrive at a more comparable data ensemble.

Note: As already explained for the definition of GCD's standard RDS [REFERENCE], the above definition of the data ensemble has not necessarily to be the natural choice of an individual member bank for its purposes. Hence, a lower DQ score w.r.t. comparability for an individual member bank only indicates that there has to be elaborated a careful definition of selection criteria for a subsample of the overall data pool for its applications.

Metric: In order to assign a continuous DQ score, the Hellinger distance serves as a fit-for-purpose metric. Let for a risk driver the (discrete) probability distribution of values for a member bank be represented by $P = (p_1, p_2, \dots, p_k)$ and for the overall pool by $Q = (q_1, q_2, \dots, q_k)$ with probabilities (relative frequencies) p_i, q_j. The Hellinger distance is then calculated as:

$$\mathrm{H(P,Q)} = \frac{1}{\sqrt{2}} \sqrt{\sum_{\mathrm{i=1}}^{\mathrm{k}} \left(\sqrt{\mathrm{p_i}} - \sqrt{\mathrm{q_i}}\right)^2}$$

The resulting values are between 0 and 1. H = 0 means that the distributions of both samples (member bank and overall pool) are identical, while H = 1 means that they are perpendicular. In general values exceeding 0.25 indicate a larger difference in distributions.

Note: For risk drivers having continuous value ranges, the distribution needs to be discretised first before the Hellinger distance metric can be applied. Further tests like the Kolmogorov-Smirnoff test might fail while the Hellinger distance score is still below 0.25 because of discretization effects. Nevertheless, the Hellinger distance still allows for comparison between different member submissions as long as the same discretization has been used for calculation.

Considered risk drivers: The following risk drivers have been selected based on its nature of typically impacting the LGD value:

- Continuous risk drivers:
  - Time to resolution
  - TTRec
- Discrete risk drivers:
  - Seniority code
  - Cure
  - Guarantee indicator
  - Collateral indicator

Translation into DQ score: In order to arrive at a meaningful DQ score per member bank and facility asset class the following procedure is carried out:

1. Calculation of Hellinger distance for each risk driver per member bank and facility asset class
2. Map the individual Hellinger scores H_i to a DQ score. Here the fact is used that distributions are considered similar for Hellinger scores below 0.25. Hence, Hellinger scores below 0.25 will be mapped to 100%. For all other Hellinger scores the respective DQ score is obtained by adding 0.25 to (1 - Hellinger score). This avoids a gap in values between 100 and 75%.
3. Calculate per member bank and facility asset class the equally weighted average over the DQ scores of all risk drivers.

Example values: Based on H1/2020 data example values have been calculated (see summary in Table 1 – Example values for comparability metric for H1/2020 data). Since the score is derived from the Hellinger distance, where values below 25% symbolize similar distribution properties between a lender and the pool, the achieved DQ scores are very satisfying.

Lender ID	Facility Asset Class	Score [%]
A	1	94.43
A	2	86.69
A	3	93.22
A	6	98.12
A	7	90.69
A	9	98.35
A	10	97.33
A	11	92.41
B	2	98.72
B	3	85.88
B	4	94.88
B	5	95.5
B	6	83.63
B	7	91.58
B	8	97.88
B		91.5
	1	99.78
	2	97.84
	3	88.34
	5	91.62

	6	100
	7	99.14
	8	89.04
	9	96.68
	11	90.99
	1	89.41
	2	92.6
	3	90.22

Table 1 – Example values for comparability metric for H1/2020 data

Timeliness

For GCD timeliness means how timely resolved loans will be submitted. In order to measure timeliness two submissions need to be compared with each others.

Metric: The following steps will be carried out:

1. Determine set of all resolved loans that have been newly added to a submission (which have not been part of the previous submission)
2. Determine degree of delay between resolution date of a loan and inclusion into submission. There are five zones identified.
   - zone i Loans submitted one year after resolution but before two years
   - zone ii Loans submitted later than two years after resolution but before three years
   - zone iii Loans submitted later than three years after resolution but before four years
   - zone iv Loans submitted later than four years after resolution but before five years
   - zone v Loans submitted later than five years after resolution
3. Per aggregation level (e.g. member bank, facility asset class) the relative frequency (relative to the total amount of new loans) for each zone will be calculated.

Translation into DQ score: In order to arrive at a meaningful DQ score per aggregation level the following procedure is carried out:

1. *Calculation of weighted average over all relative frequencies for all zones. The weights reflect higher penalties for later submission:*
   - zone i 5% weight
   - zone ii 10% weight
   - zone iii 15% weight
   - zone iv 30% weight
   - zone v 40% weight
2. *The afore calculated weighted average yields a value between 0 and 1. The lower the better the data quality. This can be simply inverted (1 - weighted average of relative frequencies) to arrive at a DQ score with higher percentage values indicating a better DQ score.*

Example values: Based on H1/2020 and H2/2019 data example values have been calculated (see summary in Table 2 – Example values for timeliness metric for H1/2020 and H2/2019 data). There are relatively many lower scores here. This has several origins like the submission of a larger chunk of older loans from some lenders. In order to have a better control here and avoid low scores, the execution of the DQ score is already included in the data audit phase, such that loans with known/explainable reasons can be manually excluded from scoring.

Lender ID	Score [%]
A	99.95
B	99.71
C	65.42
D	100.00
E	100.00
F	70.00
G	100.00
H	63.40
I	97.99
J	69.03
K	100.00
L	100.00
M	75.20
N	67.23
O	96.19
P	100.00

Table 2 – Example values for timeliness metric for H1/2020 and H2/2019 data

Consistency/ stability

For GCD consistency/ stability means how key data fields of resolved loans w.r.t. LGD modelling remain stable between two subsequent submissions. In order to measure consistency/ stability two submissions need to be compared with each others.

Data ensemble: Loans from active members, which have been resolved.

Metric: Comparison of different data fields between two subsequent submissions. Determine relative frequencies of differences for all key data fields per member bank.

Considered key data fields: The following data fields have been selected. After resolution these fields should not be subject to further changes:

- Time to resolution
- TTRec
- LGD nominal 2 (capped)
- Seniority code
- Cure
- US Segment
- Country of residence
- Facility asset class
- Collateral indicator
- Guarantee indicator

Note: It is known that further cash flows may be added after resolution and, hence, cashflow sensitive values like LGD_2 or TTRec will change. Nevertheless, the standard process should be to only resolve a loan after all expected cahflows are available. Only real post-resolution payments may be added later but they are explicitely exempted from the LGD calculation and, hence, do not lead to changes

in value. Hence, an inclusion of LGD and TTRec values into stability measure fosters improvements in the submission of resolved loans.

Translation into DQ score: In order to arrive at a meaningful DQ score per aggregation level the following procedure is carried out:

1. Calculation of equally weighted average of the relative frequencies obtained for all data fields under consideration.
2. *The afore calculated weighted average yields a value between 0 and 1. The lower the better the data quality. This can be simply inverted (1 - weighted average of relative frequencies) to arrive at a DQ score with higher percentage values indicating a better DQ score.*

Example values: Based on H1/2020 and H2/2019 data example values have been calculated (see summary in Table 3).

Lender ID	**Score [%]**
A	94.63
B	96.41
C	99.28
D	95.68
E	97.19
F	98.38
G	87.09
H	98.70
I	95.94
J	99.02
K	95.80
L	95.76
M	91.19
N	97.77
O	99.67
P	99.17
Q	95.27

Table 3 – Example values for consistency/ stability metric for H1/2020 and H2/2019 data

Correctness

The correctness measure covers the Uniqueness, Validity and Integrity dimensions. The correctness score looks at all validation rules that are an "error" and look at correctness. Based on the percentage of entries in the current dataset that pass the validation rule, it will assign a score per rule. Scores will be segmented and weighted to certain areas which are deemed to be more important to determine an overall correctness score. The total score is then translated to a percentage.

In total 400 correctness checks are done. Which could be divided into 27 categories (See Appendix 3: Correctness Categories). These checks are based on the Validation Rules which trigger an error on the fields in the 8 tables of the datamodel. Each check is weighted to the total Correctness Score. The result is the formula shown in the equation below:

$$\text{Correctness Score} = \frac{Count\ of\ Relevant\ Records - Count\ of\ Erroneous\ Records}{Count\ of\ Relevant\ Records} * \text{Weight}$$

The first step is to find and count all relevant records. This means to select all records which are restricted to the conditions applicable for the relevant field. For example, in the History table, the Loan Status value 'Default' can not be provided when the Event Type is 'Resolved'. The number of records with Event Type – Resolved will be counted as relevant records to measure the correctness of certain loan status.

The next step is to count data records with possible errors. These are the records among identified relevant records that violated the selected data quality rule. In our example the records with Loan Status value 'Default' and having Event_Type value 'Resolved' are counted.

The last step is to weight the result of the check. And sum all check results in the total Correctness Score.

Completeness

The field completeness score looks at all applicable optional or conditional optional fields and based on the percentage of fields filled, assign a score. Scores will be added and are weighted if certain areas are deemed to be more important to determine an overall completeness score. The total score is then translated to a percentage.

In total there are 81 Completeness checks done. These checks are done on fields which have the Optional feature to fill. In fact, the selections are Escape Clause (Conditionally) Mandatory Fields, Optional Fields, Conditional Optional Fields and Optional Tables.

$$\text{Completeness Score} = \frac{Count\ of\ Relevant\ Records - Count\ of\ Incomplete\ Records}{Count\ of\ Relevant\ Records}\ \text{x Weight}$$

The first step is to find and count all relevant records. This means to select all records which are restricted to the conditions applicable for the relevant field. For example, in the Collateral table, the field Real_Estate_Type_Class can only be filled if the Collateral_Type is a Real Estate. The latter fields are counted as relevant records.

The next step is to count data records with incomplete fields. These are fields missing or containing the Escape Clause (-1) identified relevant records that violated the selected data quality rule. In our example the records missing a Real_Estate_Type_Class and where the Collateral_Type is a Real Estate are counted.

The last step is to weight the result of the check. And sum all check results in the total Completeness Score.

APPENDIX 3: CORRECTNESS CATEGORIES

	Data Quality Dimension	Check Level	Category	Example
1	Integrity	Cardinality	ID's check between different ID fields	Loan_ID exists in current dataset or submitted data but with different Borrower_ID
2	Integrity	Referential Integrity	Missing ID value	Borrower_ID is empty
3	Integrity	Referential Integrity	ID's check between different tables	Collateral_Guarantor_Link not in Entity Table as Entity_ID
4	Integrity	Referential Integrity	Business key check between different tables	{Borrower_ID, Loan_ID} not in Loan Table
5	Uniqueness	Controlled redundancy	Too many records provided for an element	Count {Transaction_Type=310, Loan_ID} > 1
6	Uniqueness	Controlled redundancy	Too many records provided in one table	Event_Type 4 and Event _Type 5 exist for this loan
7	Uniqueness	Element Uniqueness with dataset	Duplicate element combination	Count {Loan_ID, Event_Type=2} > 1
8	Uniqueness	Element Uniqueness with dataset	Duplicate date combinations	Count {Borrower_ID, Loan_ID, Collateral_ID,Event_Date} combination > 1
9	Uniqueness	Element Uniqueness with dataset	Different value for one ID	Aircraft_Type different per Collateral_ID
10	Validity	Domain of Predefined Values	Non Escape Clause Conditional Mandatory field is empty	Bank_Or_Financial_Company is empty and Entity_Asset_Class = 4 or 8
11	Validity	Domain of Predefined Values	Lookup Check	Bank_Or_Financial_Company not in Bank_Or_Financial_Company_Lookup
12	Validity	Domain of Predefined Values	Non Escape Clause Mandatory field is empty	Entity_Type is empty

13	Validity	Domain of Predefined Values	Old Lookup Values	Collateral_Type= 810
14	Validity	Values Conform to Business Rule	Check Amounts	(SUM(Transaction_Amount) when Loan_Status = 6 at Event_Type= 5 and Transaction_Type=300) < Sum (LOA at Event_Type 3 + TT 400 + TT410 + TT420 + TT450 + TT480 + TT500 + TT600 + TT700 + TT800)
15	Validity	Values Conform to Business Rule	Incorrect value filled	Bank_Or_Financial_Company <> 1 to 5 and Entity_Asset_Class = 4
16	Validity	Values Conform to Business Rule	Non expected field is filled	Bank_Or_Financial_Company is filled and Entity_Asset_Class <> 4 or 8
17	Validity	Values Conform to Business Rule	Logic order of different date fields	Date_Of_Last_BB_Audit => Event_Date
18	Validity	Values Conform to Business Rule	Mandatory Record level missing in one table	Event_Type 3 missing for Loan_ID
19	Validity	Values Conform to Business Rule	Conditional Mandatory Record level missing	Loan_Status = 3 at Event_Type= 5 and count(TT300) = 0
20	Validity	Values Conform to Business Rule	Duplicate information in same record	Secondary_Industry_Code = Primary_Industry_Code
21	Validity	Values Conform to Business Rule	Logic order of different date records	Logical order mismatch in: Date (Event_Type= 1) < Date (Event_Type = 2) < Date (Event_Type = 3) < Date (Event_Type = 4 or 5).
22	Validity	Values Conform to Business Rule	Balanced Cash Flow	Balanced Cashflow: (TT100 + TT200 + TT250 + TT300 + TT490 + TT900 + LOA at Event_Type =5) - (LOA at Event_Type =3 + TT400 + TT420 + TT450 + TT480 + TT500 + TT600 + TT700 + TT800)) / (LOA at Event_Type =3 + TT400 + TT420) < -0.05
23	Validity	Values Conform to Format	Date fields Format	Event_Date not a valid date
24	Validity	Values in Specified Range	Within Minimum and Maximum value	Balloon_Percentage < 0 or Balloon_Percentage > 100
25	Validity	Values in Specified Range	Values depending on other values in table	Collateral_Value = -1 AND Total_Collateral_Value <> -1 OR Collateral_Value = 0 AND Total_Collateral_Value <> 0
26	Validity	Values in Specified Range	Incorrect date filled according to submission date	Event_Date > Today
27	Validity	Values in Specified Range	Incorrect date filled according to submission date and condition	Event_Date at Event_Type 4 > today + 365

APPENDIX 4: TOTAL DATA QUALITY SCORE

The total data quality score of the data quality criteria takes the average of the individual quality dimension scores (as previously mentioned – Stability, Timeliness, Completeness, Correctness, Comparability). This average is weighted by data quality dimension importance, meaning that each data quality criteria contributes it's individual score to the total data quality score. The weights are as follows:

Data Quality Dimension	Weight to Total Score
Timeliness	15%
Comparability	15%
Stability	20%
Correctness	30%
Completeness	20%

Item 7

Request to Methcom

Action Requested	MethCom **to be informed** about the 2021 submission cycle calendar
Background and Status	➢ GCD is informing methcom about the 2021 submission cycles for the three data pool: ➢ LGD/EAD ➢ PD and Rating ➢ Benchmarking platform

Global Credit Data

H1/2021

Week	Dates	BP Names	BP Clusters	LGD	PD & Rating	ICC	GCD Conference
2	4/1 - 10/1	CL				S	
3	11/1 - 17/1	CL				S	
4	18/1 - 24/1	CL				S	
5	25/1 - 31/1	CL			PP	S	
6	1/2 - 7/2	S			S	A	
7	8/2 - 14/2	S			S	A	
8	15/2 - 21/2	D			S	A	
9	22/2 - 28/2	A	S		S	A	
10	1/3 - 7/3	RS	S		S	A	
11	8/3 - 14/3	DA	S	PP	S	A	
12	15/3 - 21/3	DR	S	S	D	RS	
13	22/3 - 28/3	CL	S	S	A	RS	
14	29/3 - 4/4	CL	S	S	A	RS	
15	5/4 - 11/4	CL	D	S	A	DA	
16	12/4 - 18/4	CL	A	S	RS	DA	
17	19/4 - 25/4	CL	A	S	DA	ICC PR	
18	26/4 - 2/5	CL	A	D	DA	ICC PR	
19	3/5 - 9/5	CL	RS	A	DA	ICC PR	
20	10/5 - 16/5	S	DA	A	DR	ICC PR	
21	17/5 - 23/5	S	DA	A	DR	ICC PR	
22	24/5 - 30/5	D	DR	RS		ICC PR	
23	31/5 - 6/6	A	DR	DA			
24	7/6 - 13/6	RS		DA			
25	14/6 - 20/6	DA		DA			
26	21/6 - 27/6	DR		DR			
27	28/6 - 4/7	CL		DR			
28	5/7 - 11/7	CL					
29	12/7 - 18/7	CL					
30	19/7 - 25/7	CL					
31	26/7 - 1/8	CL					

H2/2021

Week	Dates	BP Names	BP Clusters	LGD	PD & Rating	ICC	GCD Conference
32	2/8 - 8/8	CL					
33	9/8 - 15/8	CL					
34	16/8 - 22/8	CL	S				
35	23/8 - 29/8	S	S	PP			
36	30/8 - 5/9	S	S	S			
37	6/9 - 12/9	D	S	S			
38	13/9 - 19/9	A	S	S			
39	20/9 - 26/9	RS	S	S			
40	27/9 - 3/10	DA	D	S			
41	4/10 - 10/10	DR	A	S			
42	11/10 - 17/10	CL	A	D			
43	18/10 - 24/10	CL	A	A			
44	25/10 - 31/10	S	RS	A			
45	1/11 - 7/11	S	DA	A			
46	8/11 - 14/11	D	DA	RS			
47	15/11 - 21/11	A	DR	DA			
48	22/11 - 28/11	RS	DR	DA			
49	29/11 - 5/12	DA		DA			
50	6/12 - 12/12	DR		DR			
51	13/12 - 19/12	CL		DR			
52	20/12 - 26/12	CL					
1	28/12 - 3/1	CL					
2	4/1 - 10/1	CL					
3	11/1 - 17/1	CL					
4	18/1 - 24/1	CL					
5	25/1 - 31/1	CL					
6	1/2 - 7/2	CL					
7	8/2 - 14/2	CL					
8	15/2 - 21/2	CL					
9	22/2 - 28/2	CL					

PP	Presubmission Package & Put in Production and Realease Note of small Enhancement
S	Submissions (Test or Firm)
D	Firm Submission Deadline
A	Audit & Resubmission
CL	Send to GCD the New Name you want to inlcude in central List

RS	Revised Submission
DA	Data Aggregation and final DQ checks
DR	Data and Reports Deliverable
ICC PR	ICC Public Report

Please keep in mind that the portal is always open for use.

Item 8

N.B

The Geographic_Country_Hierarchy is provided as a separate Excel file enclosed in the mail with the general material. In the following page the usual Methcom request slide.

Request to Methcom

Action Requested	MethCom **initiate the discussion** about reviewing the Geographic_Country_Hierarchy table
Background and Status	➢ The Geographic_Country_Hierarchy table is used to assign a geography region to country where the critical mass rule is not reached out. ➢ In GCD there are four levels: Country, Geo_1, Geo_2 and Geo_3. Whether the critical mass rule of three lender per country is not reached out the upper-level region is assigned to the loan. ➢ For some countries there were no geo 1 nor geo 2. The united nation classification has been used to fill in the missing geographic level. Some missing still remain. ➢ Material is available in the excel sent via mail with the material deck

Global Credit Data

Item 9

Request to Methcom

Action Requested	• Methcom to **initiate the discussion** on Industry type and its relationship with Facility Asset Class
Background	➢ Currently, there is no clear guidance on how to assign the industry type to the borrower. **Do GCD members want industry type at obligor or at group level**? ➢ Do GCD members want to differentiate the two, entity and group industry types? ➢ In GCD there is no mandatory association between Facility Asset Class and Industry types, except for industry type "Finance and Insurance" (*Primary_Industry_Code=600*) which is mandatory associated with FAC Banks and Financial Institution. Is this association meaningful? ➢ Example provided in following slide

Copyright The Global Credit Data Consortium 2020 all rights reserved. Confidential

Example n.1

A big automotive company (group level) has subsidiaries in the sector of Finance and insurance for financing the purchase of automotive products.

One of these subsidiaries defaults. At entity level, the borrower is identified with the Facility Asset Class Banks and Non-Banks Financial Institution (FAC=3).

What industry type should be assigned, the entity or the group industry type?

Entity level: With the current Validation Rules we force the lender to assign **industry type "Finance and Insurance"** (*Primary_Industry_code=600*).

Group level: Does it make sense to assign the group **industry type "Manufacturing - Automotive"** (*Primary_Industry_code=2501*)?

Copyright The Global Credit Data Consortium 2018 all rights reserved. Confidential