Bruno Carreiras
DG-Statistics/AMA/MAM, European Central Bank

Data quality management of entity group data

Relevance and tools to address current challenges

IFC 11th Biennial Conference

Post-pandemic landscape for central bank statistics

25-26 August 2022, Basel
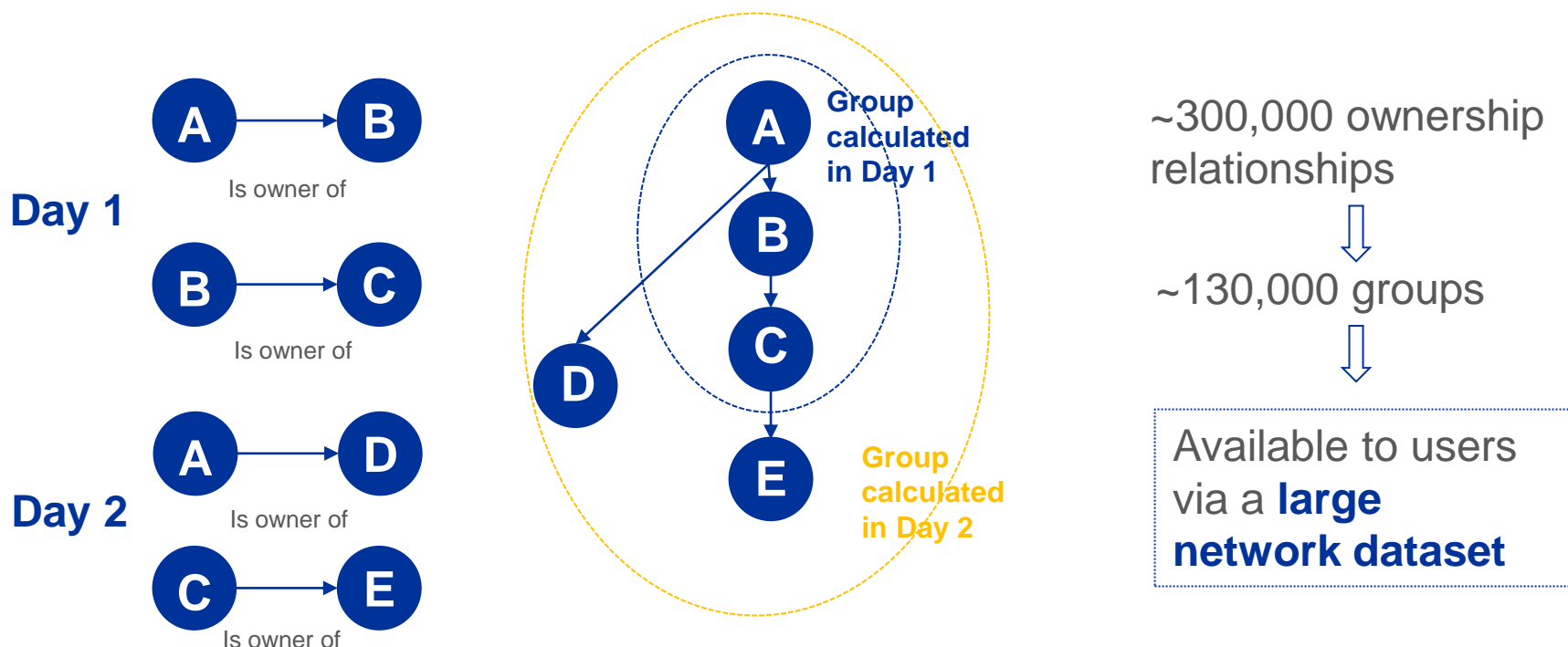
Overview

www.ecb.europa.eu ©

- The pandemic increased the need to **deliver data** to decision-makers **quickly and with <u>quality</u>**.

- Managing these expectations is **challenging**, especially when dealing with **large granular datasets** as the Register of Institutions and Affiliates Data (RIAD).

- Being the ESCB shared entity master-dataset, RIAD is managed within a **highly dynamic environment** where attributes and relationship data can be changed every day.

- Very large datasets combined with regular changes will necessarily need measures to **monitor and ensure data quality efficiently:**

> This paper tries to respond to that need by presenting a metric, the **Significance Multiplier**, applied to the <u>RIAD Group Networks</u> that aims at **prioritising potential data quality issues** by evaluating movements in group structures.

How are groups created in RIAD?

In RIAD, **relationship data is collected from NCBs and NCAs** (based on national data sources) and disseminated to users **daily.**
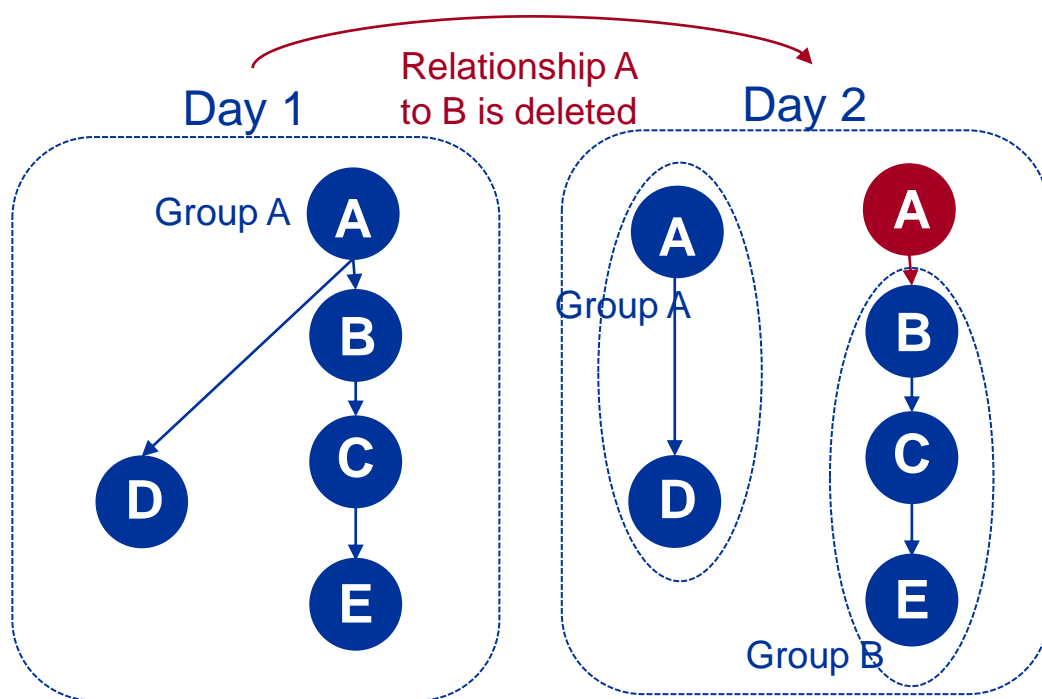– Based on such relationships, **group network trees are <u>calculated</u> each day (dynamic)**

Day 1

A → B

Is owner of

B → C

Is owner of

Day 2

A → D

Is owner of

C → E

Is owner of

Group calculated in Day 1

A
B
C
D
E

Group calculated in Day 2

~300,000 ownership relationships

⇩

~130,000 groups

⇩

Available to users via a **large network dataset**

Ensuring groups' data quality: group movements

DQM for groups is performed in two ways:
– Consistency checks on the relationship data (pre-calculation)
– **Monitoring movements in group structures: real business changes vs erroneous changes (post-calculation)**

Relationship A to B is deleted

Day 1

Day 2

Group A

A

B

C

D

E

Group A

A

D

A

Group B

B

C

E

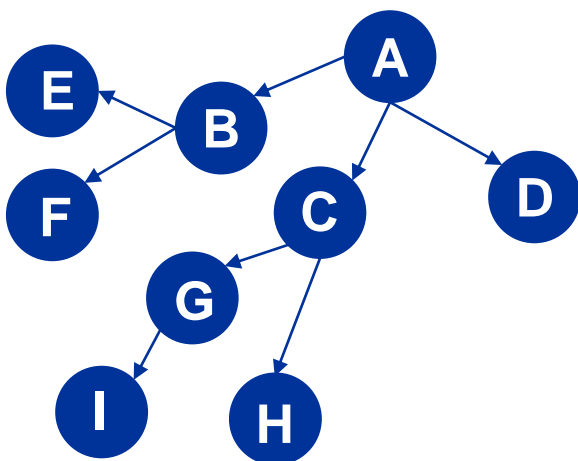**B is a leaver of Group A and Group B is generated**

Possible mistakes impact users → **movements need to be assessed**

In a universe of 120,000 groups that may change every day, it is important to **evaluate and prioritise** the monitoring and verification of movements

Finding the most relevant movements

In the context of the RIAD groups: (1) movements always start in a single event on a group member and (2) if a member leaves it takes all its direct and indirect children.

At the baseline, the relevance of a group member can be given by the count of members affected by an event on that member or, **Significance Multiplier** → the size of **(sub)group headed by the affected entity**.

Group member	Significance Multiplier
A	9
B	3
C	4
D	1
E	1
F	1
G	2
H	1
I	1

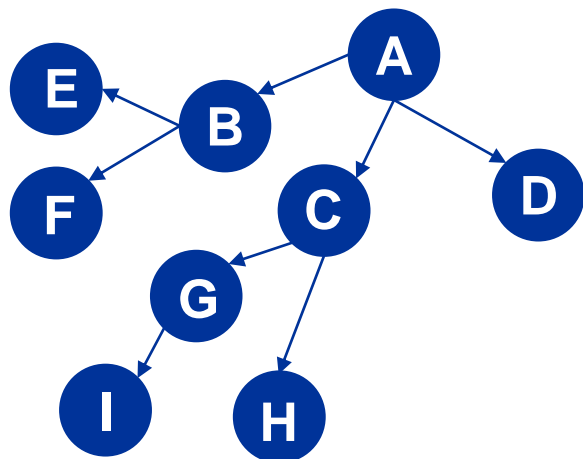👍 Measure of relevance for group member and movements intra and cross group

👎 Larger groups and members with more children are always considered more relevant

What is relevant for you?

RIAD Groups are used by multiple stakeholders → important to account for their priorities and needs in any data quality procedure and related DQM prioritisation.

- Extension: the **valuation should be based on a criterion selected by the interested business area (α)** rather than a pure count of the movements. The Significance Multiplier stops being the count of all the members of a subgroup but the sum of its α (αSM)

Group member	SM	αSM
A	9	$\alpha_A + \alpha_{B,\ldots,I}$
B	3	$\alpha_B + \alpha_E + \alpha_F$
C	4	$\alpha_C + \alpha_G + \alpha_H + \alpha_I$
D	1	α_D
E	1	α_E
F	1	α_F
G	2	$\alpha_G + \alpha_I$
H	1	α_H
I	1	α_I

👍 **Prioritisation ranking** for data quality is based on a defined criterion and not on a naïve count

👍 By-product: Users can **evaluate every group member** for their business purpose (can be calculated daily)

Defining a priority ranking for DQ

SM for <u>priority ranking</u> → relevance of groups and group members gives an indication on which movements should be assessed first.
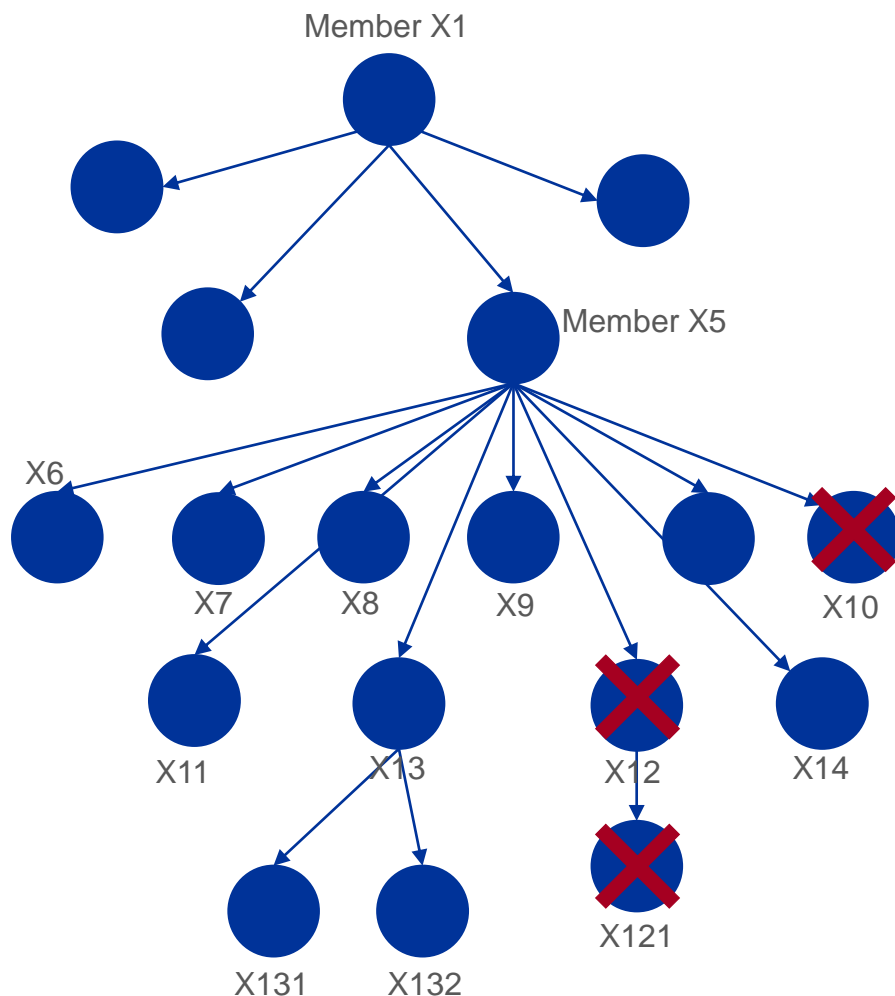
By group

Group Head	SM	αSM (Bln €)	SM Ranking	αSM Ranking
Banking group 5	696	>2,000	5	1
Banking group 7	363	>2,000	7	2
Banking group 8	348	>2,000	8	3
Banking group 17	137	>1,000	17	4
Banking group 15	186	>1,000	15	5
Banking group 6	348	>500	9	6
…	…	...	…	…
Banking group 233	50	>100	36	12

By group member

Group Member	SM	αSM (Bln €)	SM Ranking	αSM Ranking
Member 320	21	>2,000	>120	1
Member 11	194	>2,000	11	2
Member 321	2	>2,000	>500	3
Member 323	1	>2,000	>500	4
Member 324	25	>1,000	>120	5
Member 325	5	>1,000	>500	6
…	…	...	…	…

Movements are given the right priority!
→ movements on **relevant branches are assessed first**

Applied to: RIAD Banking groups, end-May
Business specific criteria (α): Individual Balance Sheet

Valuation of corporate changes – the Banking Group X

Member X1

Member X5

X6
X7
X8
X9
X10
X11
X13
X12
X14
X131
X132
X121

X12 and X10 branches are closed in RIAD reflecting real world changes!

What is the **impact of each of the changes** on banking group X balance sheet?

$$\alpha SM_{X12} = \alpha_{X12,Jan22} + \alpha_{X121,Jan22} > 1,500 \, Mln \, €$$

$$>$$

$$\alpha SM_{X10} = \alpha_{X10,Jan22} < 1,500 \, Mln \, €$$

Changes are ranked by relevance intra-group!

Valuation is already available before the event (Jan)!

Applied to: RIAD data on banking group X, end-Jan vs end-May
Business specific criteria (α): Individual Balance Sheet

- The **Significance Multiplier** provides benefits in two ways:
  – Provides a **prioritisation measure** for data quality management of changes in group networks;
  – Provides an **evaluation technique** for corporate changes for assessing the impact of movements in *real time*.

- Already being used to prioritise data quality work on RIAD group networks.

- **Possible extensions:**
  – Application to other business cases: climate change (carbon intensity) or credit data;
  – Application to other network datasets under the same business conditions (dynamic environment).

Thank you! Questions?

Bruno.Carreiras@ecb.europa.eu