

# EDW ADJUSTED DATABASE

**28 APRIL 2022**



## ON TODAY'S CALL



**DR. CHRISTIAN THUN**  
CHIEF EXECUTIVE OFFICER

✉ [christian.thun@eurodw.eu](mailto:christian.thun@eurodw.eu)



**LUDOVIC THEBAULT, PHD**  
HEAD OF RESEARCH

✉ [ludovic.thebault@eurodw.eu](mailto:ludovic.thebault@eurodw.eu)



**USMAN JAMIL**  
RESEARCH MANAGER

✉ [usman.jamil@eurodw.eu](mailto:usman.jamil@eurodw.eu)

# **AGENDA**

**INTRODUCTION TO THE EDW ADJUSTED DATABASE**

**EXAMPLES OF DATA CORRECTIONS**

**OVERVIEW OF DATA IMPROVEMENTS**

**EXTENT OF MODIFICATIONS**

**NEXT STEPS**

**Q & A**

# INTRODUCTION

**LUDOVIC THEBAULT, EUROPEAN DATAWAREHOUSE**

# WHAT IS EDW'S ADJUSTED DATABASE?

## The Key Points:

- It is a corrected copy of our ECB database, covering Q1 2013 - Q4 2021
- Adjustments have been made to RMBS, SME, Auto, Consumer, and Leasing asset classes
- Beta version now available

## The Benefits:

- Improved usability
- Reduction of data preparation
- Enhanced decision-making power

## The Goal:

- Copy trustworthy data; or
- Copy corrected data where possible;
- Eliminate incorrect data that cannot be corrected;
- Add useful information where feasible/needed (WIP)

## The Challenge

100% accuracy is not guaranteed but our commitment to continuous improvement is!

# THE DATA QUALITY PROCESS

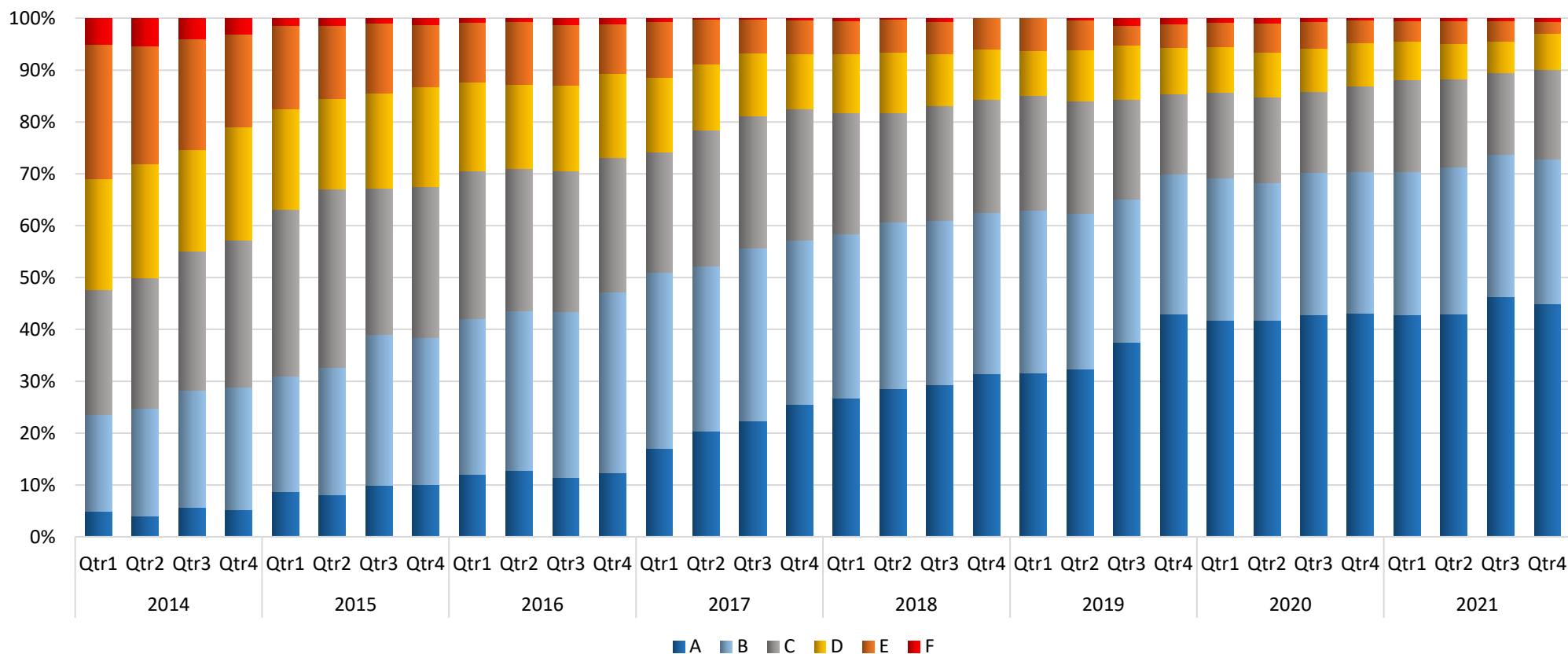
## Overview

- EDW runs hundreds of data quality checks per loan level data (LLD) upload
- EDW uses a ticketing system to track errors until correction of the data
- Data provider/owners typically correct errors for the next submission
- Existing errors tend to stay in the database
- EDW produced data quality scores to track quality

## Good to know

- The oldest data has the most issues
- Frequency of LLD submissions is not always uniform
- Inactive loans are dropped out of later submissions
- Loan identifiers are meant to be static but this is not always the case

# DQS1 (DATA QUALITY SCORES) EVOLUTION FROM 2014 UNTIL Q4 2021 (ECB)



## ADJUSTED DATABASE FEATURES

Adjustments at the field level (e.g. elimination of dummy values)

Adjustments at the EDCODE+PCD+field level (e.g. AR67 / 100 in some LLD)

We correct data errors where possible

Report NULL when data is not usable

Add extra fields where needed (WIP)

Preserve time series since 2013

Exclude the worst data submissions



## EXCLUDED UPLOADS (COMBINATION EDCODE + PCD)

Some submissions (EDCODE + PCD) are excluded from the Adjusted Database

Examples of excluded data:

- Most submissions dated Q4 2012
- Key data which appears to be incorrect and cannot be rectified (eg. current loan amount)
- Multiple uploads in the same month for a given securitisation – the latest data point is retained as it is usually the correct one
- When a LLD includes duplicate values (all loans appear twice)

# ERROR DETECTION

## Implausible values in relative terms (verification at the field level)

- Dates entered in the future

## Implausible values in absolute terms (can be errors or dummy values)

- Outliers such as:
  - Extremely high incomes
  - Borrower age – we assume to be >18 and <120 years

## Time series comparisons

- For a given EDCODE+field, we compare over time
  - Sum of that field, minimum value, maximum value, average values
  - Variations from one submission to the next
  - Dummies which stand out as min or max values

# EXAMPLES OF DATA CORRECTIONS

LUDOVIC THEBAULT, EUROPEAN DATAWAREHOUSE

# EXTREMELY IMPLAUSIBLE VALUES AND DUMMY VALUES

## Dummy values, description, and correction (replace by NULL)

- Dummy values are extremely implausible values that are used instead of reporting ND
- Dummy values should not be used in calculations as they can have severely distorting effects
- Examples include:
  - AR136 (valuation **amount**) 999999999999
  - AS55 (loan **amount** SME) 99999999999.99
  - AS50 (origination **date** SME) 1900-01-01; 0001-01; 9999-12-01
  - AS37 (SME LGD) 999.99

## Non-dummy value corrections

- Fields with limited options
  - In a Y/N field, anything else is replaced by NULL (in some cases, 1 is replaced by Y and 0 by N)
  - In a field accepting 1,2,3 as an answer, anything else is NULL
- Other sanity checks implemented:
  - Eg. some dates should not be later than the as of date of the data
  - Some clearly excessive income amounts are replaced by NULL

# RMBS ADJUSTMENTS (DATE FIELDS)

18 date fields were verified using 3 logics:

- Known dummy values were removed (replaced by NULL)
- Known incorrect dates were removed (replaced by NULL)
- Dates way into the past or future were removed (replaced by NULL)

**Bureau Score Date**

**Loan Origination Date**

**Date of Loan Maturity**

**Account Status Date**

**Retained Amount Date**

**Further Loan Advance Date**

**Pre-payment Date**

**Interest Revision Date 1**

**Interest Revision Date 2**

**Interest Revision Date 3**

**Final Step Date**

**Valuation Date**

**Current Valuation Date**

**Date of Sale**

**Date Last Current**

**Date Last in Arrears**

**Redemption Date**

**Date of Default or Foreclosure**

# DUMMY VALUES AFFECTING AMOUNTS

Example of decimal point errors affecting field AS55 (loan amount - SME) – dummy is replaced with NULL

EDCODE	PCD	Total	MaxAS55	Amended Total	Fix
SMESES000089101320051	2013-03-28	300,333,921,482	99,999,999,999.99	333,921,482	Excl. dummy
SMESES000089101320051	2013-04-30	300,327,672,535	99,999,999,999.99	327,672,535	Excl. dummy
SMESES000089101320051	2013-07-26	200,309,305,175	99,999,999,999.99	309,305,175	Excl. dummy
SMESES000089101320051	2013-09-23	296,733,295	11,910,747.44	296,733,295	
SMESES000089101320051	2013-10-24	293,560,254	11,910,747.44	293,560,254	

# DECIMAL POINT ISSUES AFFECTING AMOUNTS

Example of decimal point error affecting field AR67 (loan amount) – amounts reported in cents are divided by 100 to adjust

EDCODE	PCD	Total	Minloan	Maxloan	Amendment	Factor
RMBSES000065100820109	2012-12-31	543,666,534	-	5,137,760	543,666,534	1
RMBSES000065100820109	2013-01-31	541,268,537	-	5,137,760	541,268,537	1
RMBSES000065100820109	2013-02-28	533,034,307	-	5,137,760	533,034,307	1
RMBSES000065100820109	2013-03-31	530,458,966	-	5,137,760	530,458,966	1
RMBSES000065100820109	2013-04-30	52,674,246,801	-	513,776,024	526,742,468	0.01
RMBSES000065100820109	2013-05-31	52,421,459,882	-	513,776,024	524,214,599	0.01
RMBSES000065100820109	2013-06-30	52,135,713,808	-	513,776,024	521,357,138	0.01
RMBSES000065100820109	2013-07-31	51,881,453,371	-	513,776,024	518,814,534	0.01
RMBSES000065100820109	2013-08-31	51,669,610,421	-	513,776,024	516,696,104	0.01
RMBSES000065100820109	2013-09-30	51,377,559,409	-	513,776,024	513,775,594	0.01
RMBSES000065100820109	2013-10-31	511,377,065	-	5,137,760	511,377,065	1
RMBSES000065100920081	2012-12-31	837,733,199	-	825,103	837,733,199	1
RMBSES000065100920081	2013-03-31	813,261,111	-	817,275	813,261,111	1
RMBSES000065100920081	2013-06-30	78,804,702,406	-	78,943,414	788,047,024	0.01
RMBSES000065100920081	2013-08-31	77,450,398,292	-	78,419,876	774,503,983	0.01
RMBSES000065100920081	2013-09-30	76,665,850,372	-	78,157,765	766,658,504	0.01
RMBSES000065100920081	2013-12-31	743,841,573	-	773,699	743,841,573	1
RMBSES000065101120061	2012-12-31	581,713,724	-	657,088	581,713,724	1
RMBSES000065101120061	2013-02-28	571,454,162	-	657,088	571,454,162	1
RMBSES000065101120061	2013-05-31	55,579,620,815	-	44,983,108	555,796,208	0.01
RMBSES000065101120061	2013-07-31	54,614,359,961	-	44,645,161	546,143,600	0.01
RMBSES000065101120061	2013-08-31	54,200,701,041	-	44,475,970	542,007,010	0.01
RMBSES000065101120061	2013-11-30	529,959,463	-	439,675	529,959,463	1

## IMPACT OF AMENDMENTS TO THE FIELD AR67 (CURRENT LOAN AMOUNT)

- Differences due to suppression of dummy values
- Correction of the amount originally reported (\*-1; \*100; etc)
- In 4 countries, sum of AR67 shows differences (adjusted vs original)
- Many other amount fields have been similarly amended in the various asset classes

	AR67 (Sum Adjusted / Sum Original)									
	2013	2014	2015	2016	2017	2018	2019	2020	2021	
BE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
ES	<b>10.8%</b>	<b>100.2%</b>	<b>100.3%</b>	<b>100.3%</b>	<b>100.4%</b>	<b>100.4%</b>	<b>100.4%</b>	<b>100.4%</b>	<b>100.4%</b>	<b>100.3%</b>
FR	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	<b>100.6%</b>	<b>101.0%</b>	<b>101.1%</b>	
IE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
IT	<b>26.9%</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
NL	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
PT	<b>115.3%</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
SE			100.0%							
UK	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%



# DECIMAL POINT ISSUE AFFECTING PERCENTAGES

Decimal point error in field AR109 (current interest rate) – value reported as 0.0338 instead of 3.38 is multiplied by 100 to adjust

edcode	pcd	Min	Max	Average	loans	Amendment	Factor
RBMUK000551100120075	2016-01-31	0.0069	0.0669	0.04	40,748	3.62	100
RBMUK000551100120075	2016-03-31	0.0069	0.0669	0.04	39,688	3.60	100
RBMUK000551100120075	2016-07-31	0.0069	0.0669	0.04	35,820	3.59	100
RBMUK000551100120075	2016-10-31	0.0035	0.0639	0.03	37,047	3.38	100
RBMUK000551100120075	2017-01-31	0.3500	6.3900	3.35	34,572	3.35	1
RBMUK000551100120075	2017-04-30	0.3500	6.3900	3.33	32,115	3.33	1
RBMUK000551100120075	2017-07-31	0.2500	6.3900	3.20	47,088	3.20	1
RBMUK000551100120075	2017-10-31	0.2500	6.3900	3.18	44,525	3.18	1
RBMUK000551100120075	2018-01-31	0.5000	6.3900	3.24	42,355	3.24	1
RBMUK000551100120075	2018-04-30	0.5000	6.3900	3.15	44,106	3.15	1
RBMUK000551100120075	2018-07-31	0.50	6.39	3.04	47,245	3.04	1
RBMUK000551100120075	2018-10-31	0.75	5.60	3.11	44,467	3.11	1
RBMUK000551100120075	2019-01-31	0.75	5.29	3.10	41,917	3.10	1

# STATIC FIELD CORRECTION

## When implausible values are identified in a static field, we:

- Assume that the last available submission is the correct one
- Verify whether foreign nationals are a minority
- Check the OC - if there is a substantial proportion of FN, it should be mentioned in the OC

## Example: Identifying change of pattern

**AR16** Optional field Y/N “Foreign National” indicating whether the borrower is a national of the country in which the property and mortgage loan resides. If no data available, refer to Taxonomy for inputs.

EDCODE	PCD	AR16	AR16 corrected	Sum Before correction	Sum After Correction
RMBSES000065101820066	2016-02-29	N	Y	311,483,323	40,518,220
RMBSES000065101820066	2016-02-29	Y	N	40,518,220	311,483,323
RMBSES000065101820066	2016-05-31	N	Y	305,618,091	40,271,123
RMBSES000065101820066	2016-05-31	Y	N	40,271,123	305,618,091
RMBSES000065101820066	2016-08-31	N	N	40,021,575	40,021,575
RMBSES000065101820066	2016-08-31	Y	Y	298,690,571	298,690,571
RMBSES000065101820066	2016-11-30	N	N	39,722,088	39,722,088
RMBSES000065101820066	2016-11-30	Y	Y	291,711,348	291,711,348

# DATA QUALITY AFFECTING PERFORMANCE FIELDS (1)

## Corrected performance field errors (examples)

- A defaulted loan - current amount (AR67) is reported in the default amount field (AR177) and zero is reported in the current loan amount field:
  - The improperly reported value in the default amount is used to populate the current loan amount
  - Error affecting some french and Spanish deals
- Days in arrears:
  - One SME data provider reported the loans in arrears with a lag of 90 days (we add 90 days to the number of days reported in this case)
  - For RMBS, the same data provider reported loans in arrears with a lag of 3 months (add 3 months to the number of months reported)

## Remaining errors affecting performance fields (and how to deal with them)

- Not all errors and inconsistencies affecting the performance fields are rectified
- The default amount should be a static field, but sometimes is not static
- A date of default should be provided if the loan is in default, but is often not provided

## DATA QUALITY AFFECTING PERFORMANCE FIELDS (2)

Loan outstanding amount becomes 0 when loan reaches default status. We use the default amount as a proxy.

In this case, we use the amount reported in the default field is indeed a „dynamic amount“ ...

NON ADJUSTED DATABASE						ADJUSTED DATABASE					
EDCODE	PCD	AR3	AR67	AR166	AR177	EDCODE	PCD	AR3	AR67	AR166	AR177
RMBSES000	2019-07-09	490,004,103,044,881	0	3	84,393.94	RMBSES000	2019-07-09	490,004,103,044,881	84,393.94	3	84,393.94
RMBSES000	2019-10-09	490,004,103,044,881	0	3	83,286.14	RMBSES000	2019-10-09	490,004,103,044,881	83,286.14	3	83,286.14
RMBSES000	2020-01-09	490,004,103,044,881	0	3	82,373.01	RMBSES000	2020-01-09	490,004,103,044,881	82,373.01	3	82,373.01
RMBSES000	2020-04-07	490,004,103,044,881	0	3	81,949.04	RMBSES000	2020-04-07	490,004,103,044,881	81,949.04	3	81,949.04

In this case, we use the amount reported in the default field is indeed a “static amount”, the adjusted value we enter for AR67 is therefore a proxy.

NON ADJUSTED DATABASE						ADJUSTED DATABASE					
EDCODE	PCD	AR3	AR67	AR166	AR177	EDCODE	PCD	AR3	AR67	AR166	AR177
RMBSFR000	2019-01-31	10207GPR08708249	213,722	3	213,722	RMBSFR000	2019-01-31	10207GPR08708249	213,722	3	213,722
RMBSFR000	2019-02-28	10207GPR08708249	213,722	3	213,722	RMBSFR000	2019-02-28	10207GPR08708249	213,722	3	213,722
RMBSFR000	2019-03-31	10207GPR08708249	-	3	213,722	RMBSFR000	2019-03-31	10207GPR08708249	213,722	3	213,722
RMBSFR000	2019-04-30	10207GPR08708249	-	3	213,722	RMBSFR000	2019-04-30	10207GPR08708249	213,722	3	213,722
RMBSFR000	2019-01-31	10807GPR08713489	50,792	3	50,792	RMBSFR000	2019-01-31	10807GPR08713489	50,792	3	50,792
RMBSFR000	2019-02-28	10807GPR08713489	50,792	3	50,792	RMBSFR000	2019-02-28	10807GPR08713489	50,792	3	50,792
RMBSFR000	2019-03-31	10807GPR08713489	-	3	50,792	RMBSFR000	2019-03-31	10807GPR08713489	50,792	3	50,792
RMBSFR000	2019-04-30	10807GPR08713489	-	3	50,792	RMBSFR000	2019-04-30	10807GPR08713489	50,792	3	50,792
RMBSFR000	2019-01-31	11315P0004261997	22,219	3	23,453	RMBSFR000	2019-01-31	11315P0004261997	22,219	3	23,453
RMBSFR000	2019-02-28	11315P0004261997	22,081	3	23,453	RMBSFR000	2019-02-28	11315P0004261997	22,081	3	23,453
RMBSFR000	2019-03-31	11315P0004261997	-	3	23,453	RMBSFR000	2019-03-31	11315P0004261997	23,453	3	23,453
RMBSFR000	2019-04-30	11315P0004261997	-	3	23,453	RMBSFR000	2019-04-30	11315P0004261997	23,453	3	23,453

# DATA IMPROVEMENTS

LUDOVIC THEBAULT, EUROPEAN DATAWAREHOUSE

# DATA CAN ALSO BE IMPROVED

## Data improvements

- All NDs turned into NULL
- Originator field for various asset classes is populated with originator name
  - Sometimes, a number, code, or initials are provided instead of name
  - Identifiers are often inconsistent (can be international, eg. SWIFT, or national, or data provider-specific)
  - Code is replaced by the actual originator name
- Extra information is produced to improve data user experience (gain time & accuracy)

## Extra fields (Work In Progress)

- Quarter identifier field to identify data corresponding to a specific quarter
  - Some deals submit quarterly, others monthly
- Region names
  - A region name is easier to interpret than a NUTS code or the first digits of postcode
  - Mapping has been established to produce a NUTS 2 or NUTS 3 region name depending on info provided

# RMBS IMPROVEMENTS TO AR5 (ORIGINATOR)

- AR5 populated either by actual originator name or a code (international, national, or an internal identifier)
- We enter the actual name of the originator in the database
- The result when running a “SELECT DISTINCT AR5” query for an Italian multi-originator deal:

**Select distinct query outcome  
(Non-adjusted database)**

	AR5
1	06160
2	03059
3	06080
4	03240
5	06315
6	06070
7	01025
8	06340
9	06260
10	06385
11	CARIPLO
12	01010
13	Banca Intesa Spa
14	05748
15	06225
16	06065
17	06010
18	06280
19	06345

**Select distinct query outcome  
(Adjusted Database)**

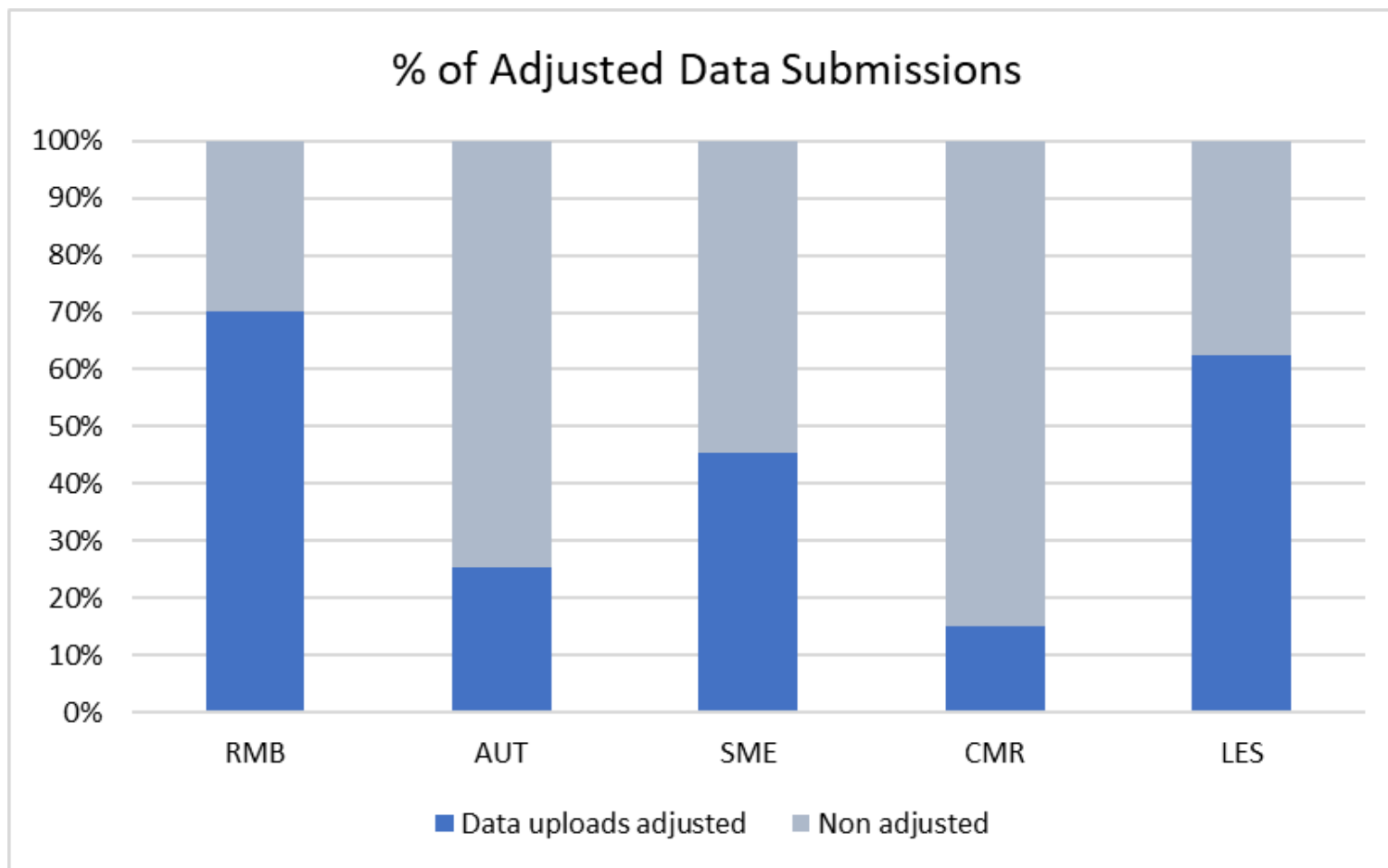
	AR5
1	SANPAOLO IMI S.P.A.
2	CASSA DEI RISPARMI DI FORLI' E DEL
3	CARIPLO S.P.A.
4	BANCO DI NAPOLI S.P.A.
5	CASSA DI RISPARMIO DELLA PROVINCIA
6	CASSA DI RISPARMIO DI PISTOIA E DEL
7	CASSA DI RISPARMIO DEL FRIULI VENEZ
8	CARIPLO
9	BANCA DI CREDITO SARDO S.P.A.
10	Banca Intesa Spa
11	CASSA DI RISPARMIO IN BOLOGNA SPA
12	BANCA DI TRENTO E BOLZANO
13	BANCA DELL'ADRIATICO S.P.A.
14	CASSA DI RISPARMIO DEL VENETO
15	CASSA DI RISPARMIO DI VENEZIA SPA
16	CASSA DI RISPARMIO DI FIRENZE SPA
17	CASSA DI RISPARMIO DI RIETI SPA
18	CASSA DI RISPARMIO DI ASCOLI PICENO
19	CASSA DI RISPARMIO DI SPOLETO SPA

# EXTENT OF DATA MODIFICATIONS

LUDOVIC THEBAULT, EUROPEAN DATAWAREHOUSE

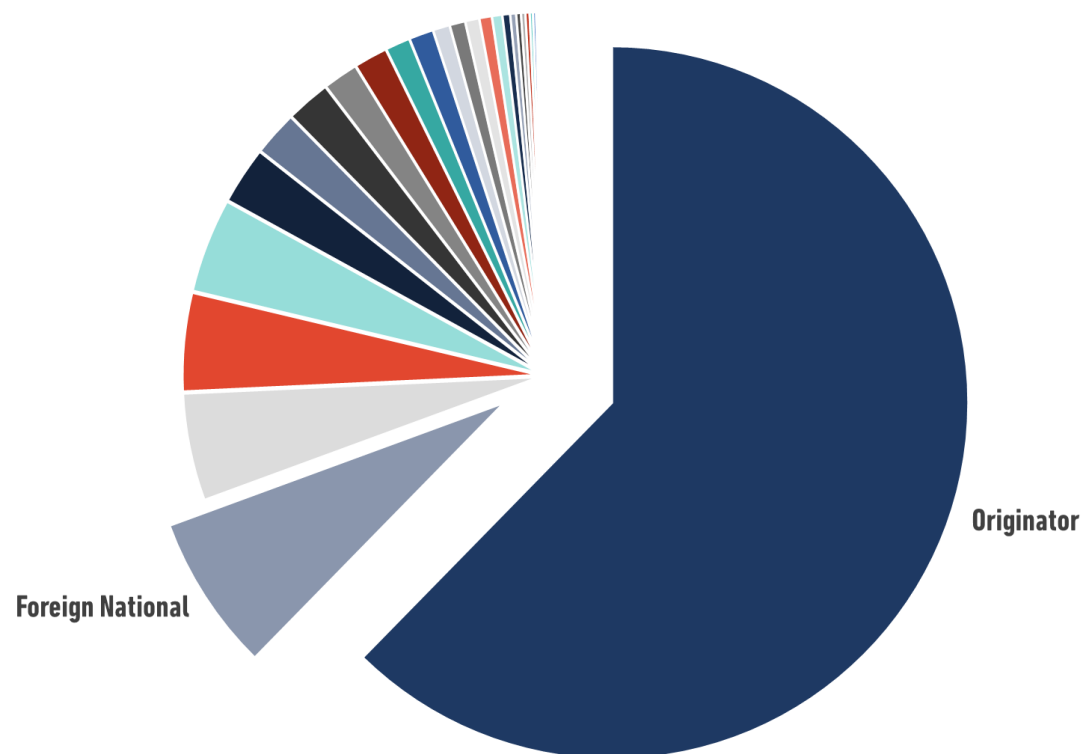


# MOST RMBS AND LEASING SUBMISSIONS HAVE AT LEAST ONE ADJUSTMENT



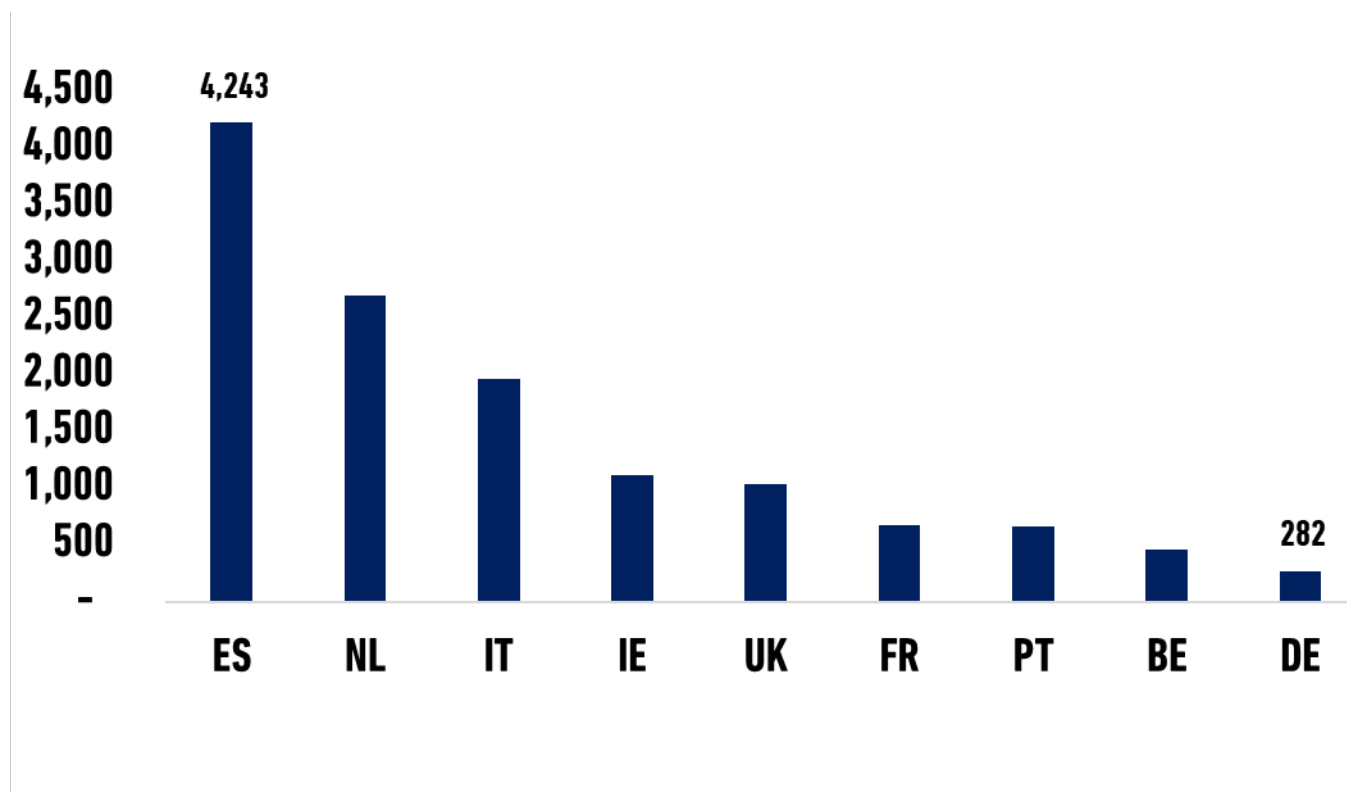
# RMBS ADJUSTMENTS

211 Million out of 338 Million loan record adjustments were made on the Loan Originator field (AR5)



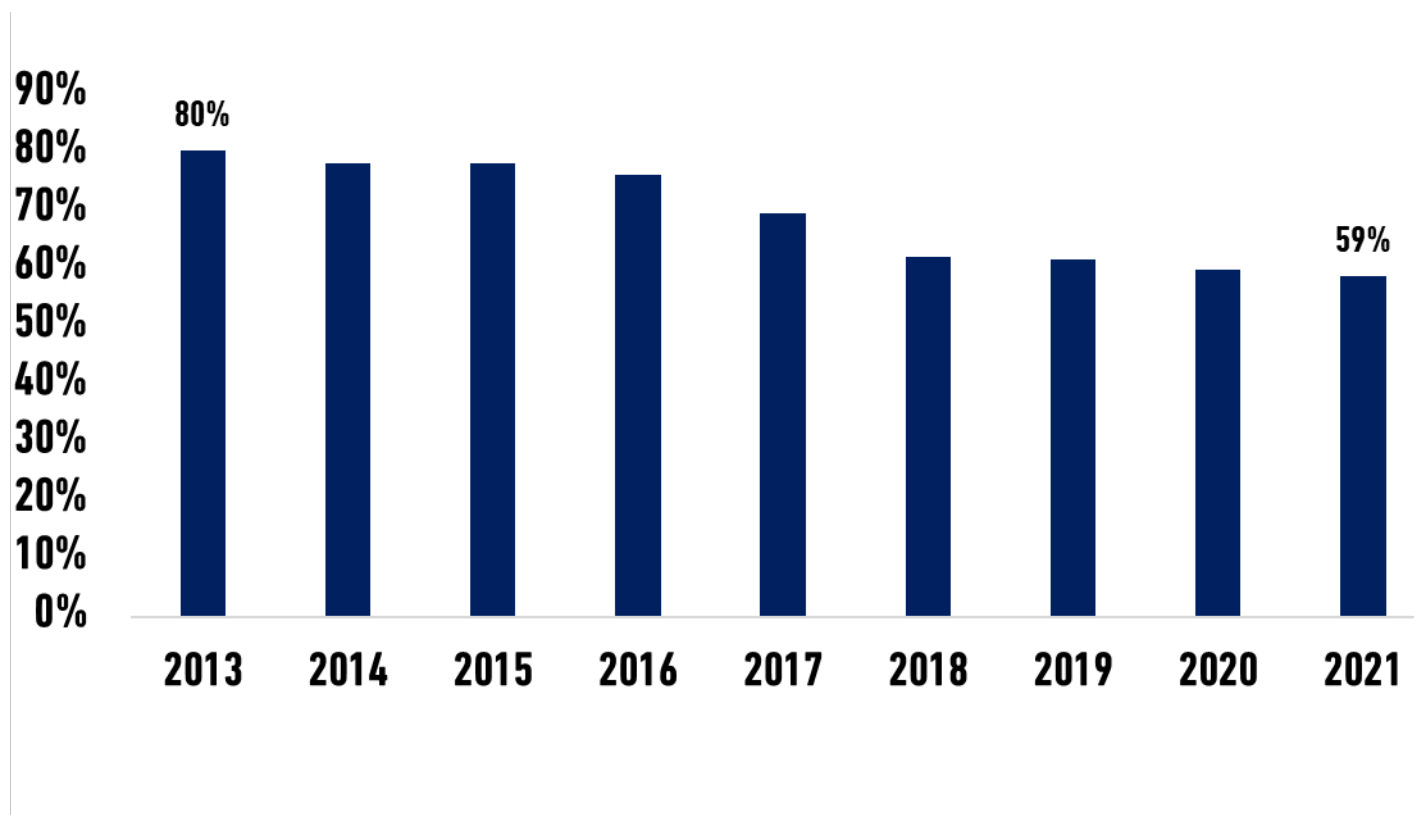
# RMBS ADJUSTMENTS (EXCLUDING AR5 - ORIGINATOR)

4,243 out of a total 13,460 RMBS uploads (EDCODE + PCD) adjusted are from Spain



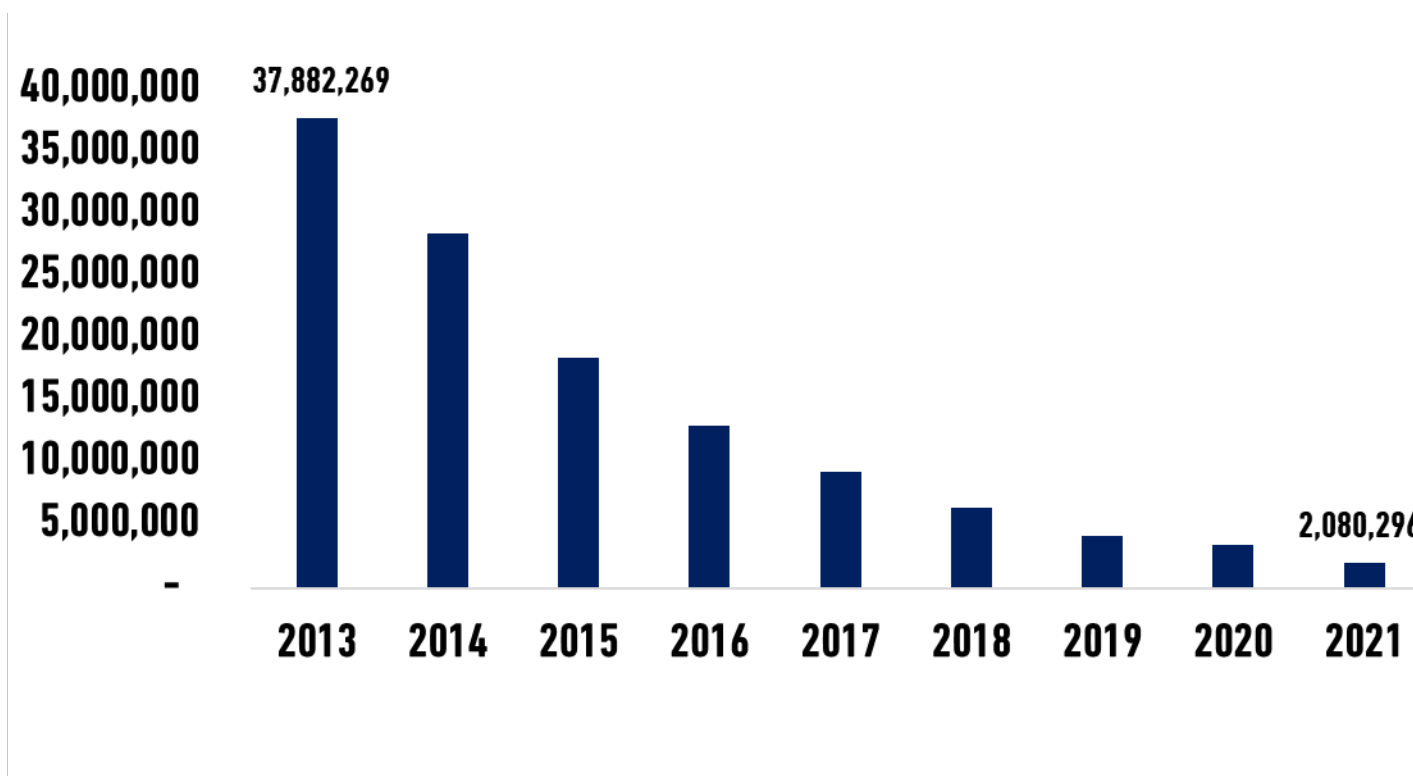
# RMBS ADJUSTMENTS (EXCLUDING AR5 - ORIGINATOR)

Portion of total LLD uploads that required adjustments decreased following each Pool Cut-off Date



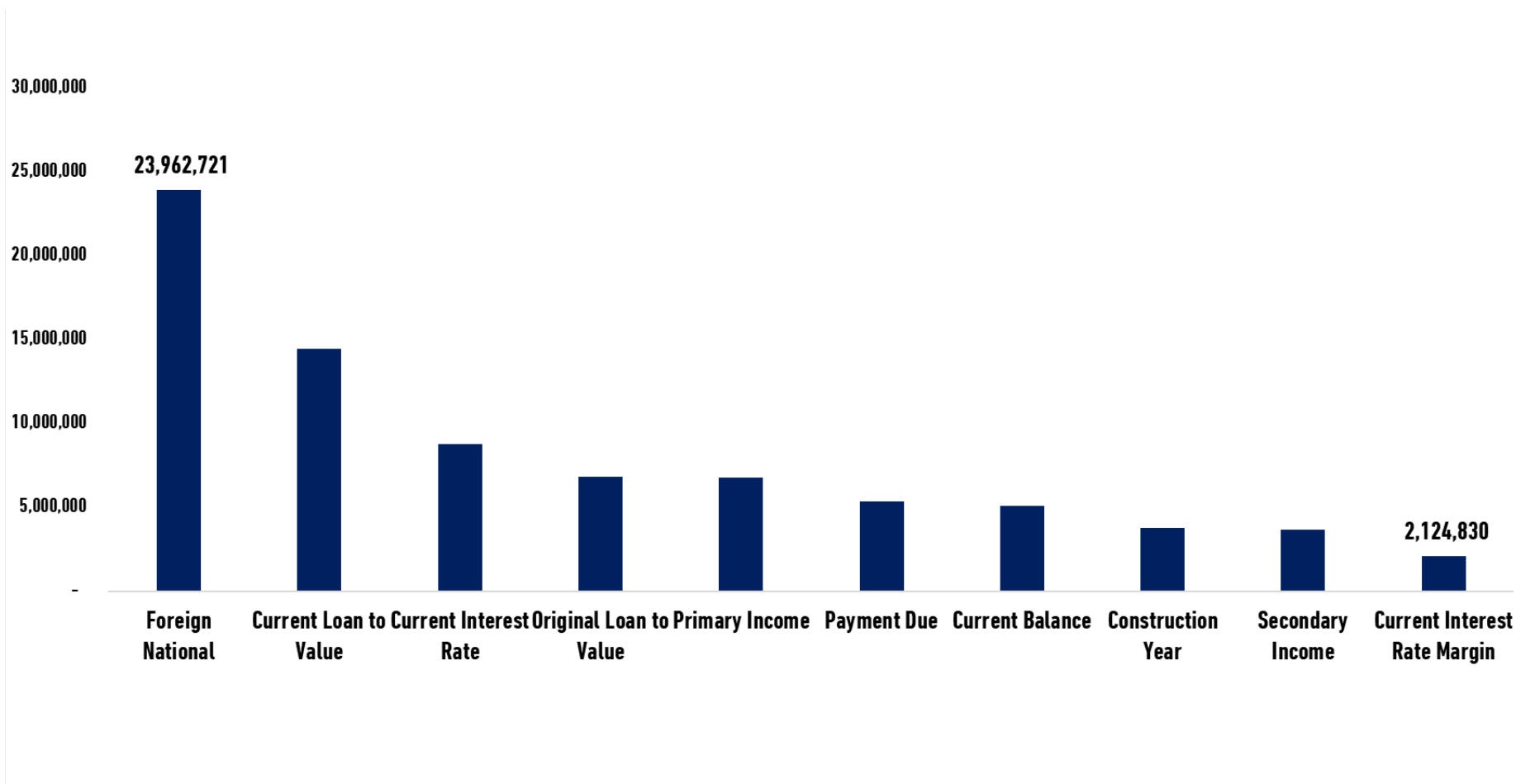
# RMBS ADJUSTMENTS (EXCLUDING AR5 – ORIGINATOR)

Number of loan records that required adjustments decreased drastically each year



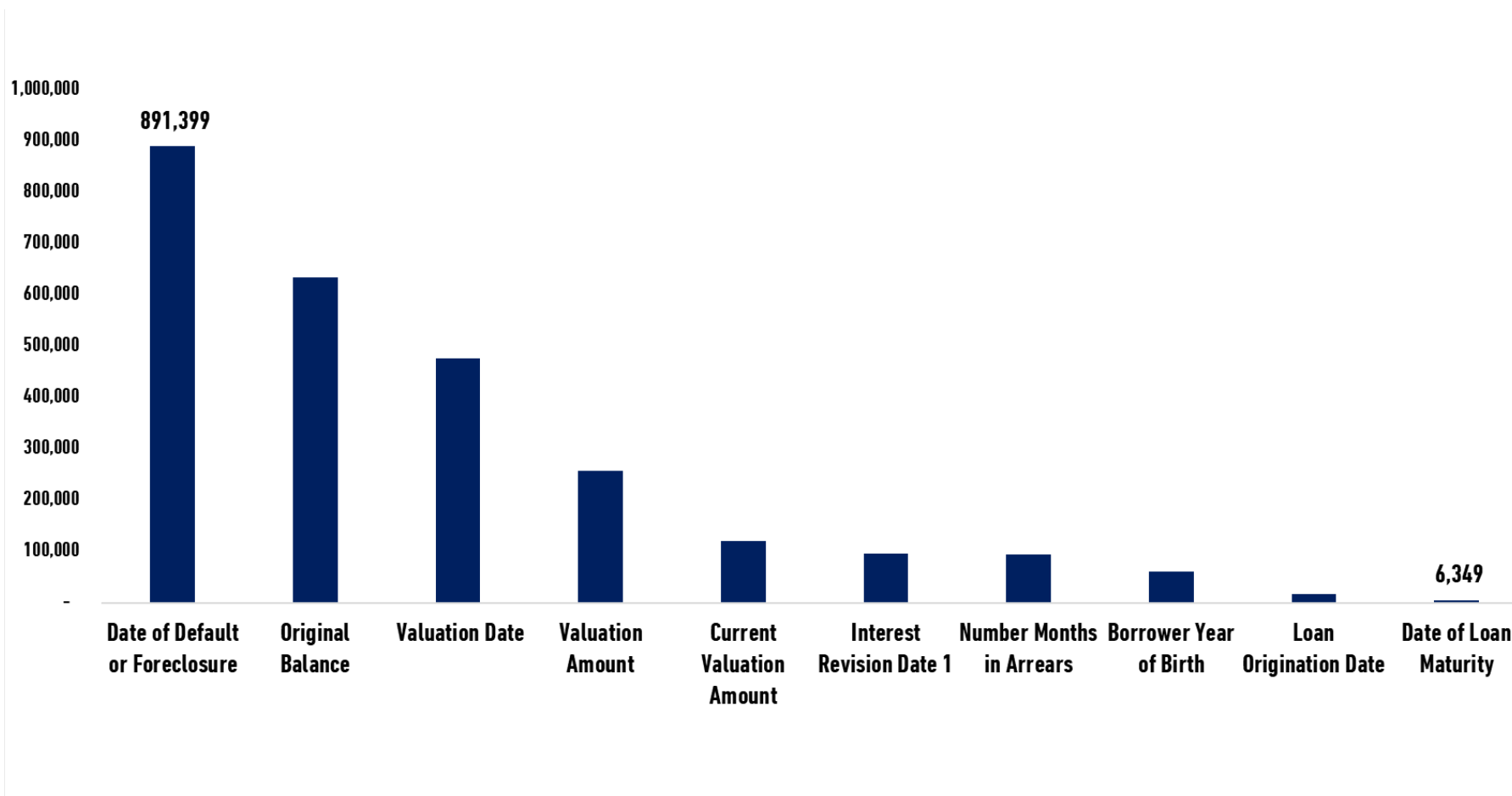
# RMBS ADJUSTMENTS

Examples of fields where >1M loan records were adjusted



# RMBS ADJUSTMENTS

Examples of fields where <1M loan records were adjusted



# NEXT STEPS

LUDOVIC THEBAULT, EUROPEAN DATAWAREHOUSE



# NEXT PROJECT: AN “ALL IN ONE” DATABASE

## Extra fields

- A work in progress to improve data

## Ongoing improvements

- This is a beta version, further corrections or improvements to existing corrections will be added where possible

## Since 2021, LLD is received via different sources

- ESMA format will eventually replace the ECB-format data
- Data is also received via the UK database
- Some data has also been received in „unstructured“ format (not yet databased)

## An “all in one” solution

- So far, we have worked mainly with ECB data, for which we have up to nine years of data
- For some queries, we need to use data from several sources
- An integrated database is planned (ESMA data + time series from the adjusted database)
- An ESMA-format database will be back-populated with adjusted ECB data

## For now...

- The Adjusted Database will be available to research users as a beta version, covering 2013-2021

**Please let us know if you are interested!**



**Q&A**

# NEXT RESEARCH WEBINAR

## SAVE THE DATE:

## 2 JUNE 2022 @ 16:00

[REGISTER NOW](#)

# THANK YOU//CONTACT US

## EUROPEAN DATAWAREHOUSE GMBH

**Walther-von-Cronberg-Platz 2**

**60594 Frankfurt am Main**

 **[www.eurodw.eu](http://www.eurodw.eu)**

 **[enquiries@eurodw.eu](mailto:enquiries@eurodw.eu)**

 **+49 (0) 69 50986 9017**

This presentation (the “Presentation”) has been prepared by European DataWarehouse GmbH (the “Company”) and is being made available for information purposes only. The Presentation is strictly confidential and any disclosure, use, copying and circulation of this Presentation is prohibited without the consent from the Company. Information in this Presentation, including forecast financial information, should not be considered as advice or a recommendation to investors or potential investors in relation to holding, purchasing or selling securities or other financial products or instruments and does not take into account your particular investment objectives, financial situation or needs. No representation, warranty or undertaking, express or implied, is made as to the accuracy, completeness or appropriateness of the information and opinions contained in this Presentation. Under no circumstances shall the Company have any liability for any loss or damage that may arise from the use of this Presentation or the information or opinions contained herein. Certain of the information contained herein May include forward-looking statements relating to the business, financial performance and results of the Company and/or the industry in which it operates. Forward-looking statements concern future circumstances and results and other statements that are not historical facts, sometimes identified by the words “believes”, “expects”, “predicts”, “intends”, “projects”, “plans”, “estimates”, “aims”, “foresees”, “anticipates”, “targets”, “May”, “will”, “should” and similar expression. The forward-looking looking statements, contained in this Presentation, including assumptions, opinions and views of the Company or cited from third party sources are solely opinions and forecasts which are uncertain and subject to risks.