European Loan Level Data: Smart usage beyond Asset-Backed Securities

October 2016

EUROPEAN DATAWAREHOUSE

White paper

European DataWarehouse

Following the European Central Bank ABS loan-level initiative, which became operational in early 2013, European DataWarehouse as the centralised data repository has collected loan-level data for more than 50 million loans across Europe. While the primary purpose is to enhance transparency for ABS market participants, there are also important insights into loan markets more generally given the high granularity of data across issuers, asset classes, jurisdictions and time series. Moreover, the experience gathered so far can provide interesting lessons on data quality management.

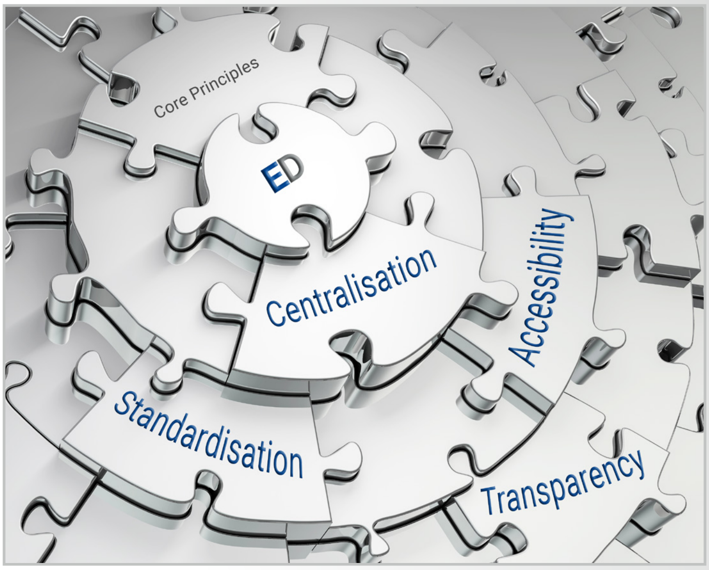
European DataWarehouse (ED) is the first central data repository in Europe for collecting, validating and disseminating detailed, standardised and asset class specific loan level data (LLD) for Asset-Backed Securities (ABS). Developed, owned and operated by the market, ED facilitates risk assessment and improves transparency standards for European ABS deals.

More specifically, ED collects ABS deal, bond and loan level data according to the ECB ABS reporting templates. Simultaneously, ED also acts as a distributor of loan level data and documentation to subscribing entities such as investors, rating agencies, data vendors and analytic firms, investment and commercial banks, accounting firms, trustees and consultants.

A substantial amount of time and effort is constantly spent on improving the quality of the submitted data, along with its accessibility and usability.

Recently ED launched the ED Cloud Pro, a business intelligence solution that enables easy access to the entire universe of ED loan and bond level data.

This paper looks at three typical questions in relation to data projects:



1. What data is available and what are typical use cases - within and beyond the ABS market ?
2. How to overcome the challenge of non-standardised data and other data quality issues ?
3. How to make large quantities of data usable and easy to analyse ?

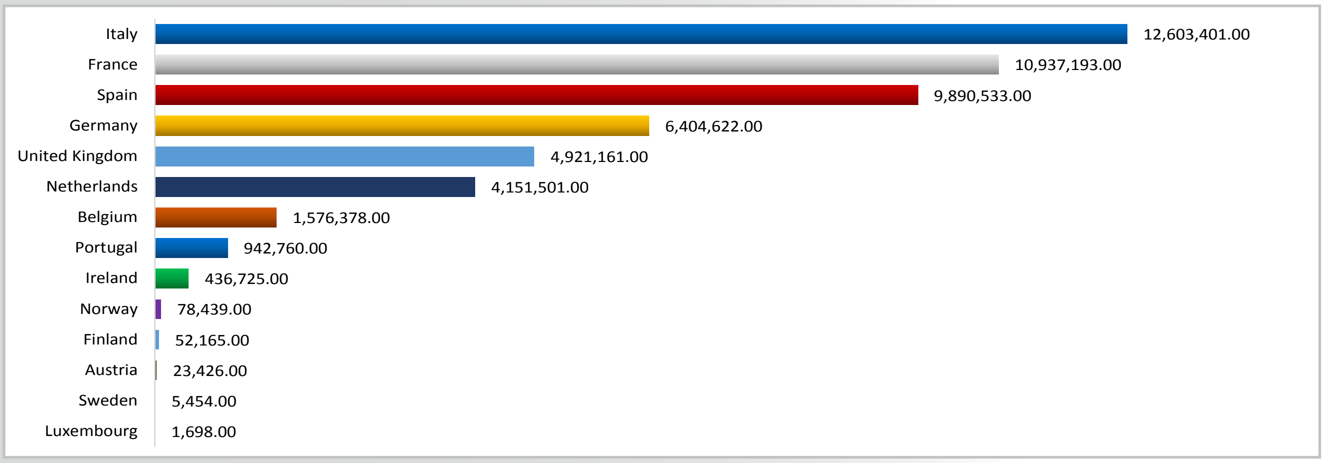
What data is available and what are typical use cases

ED collects loan level data according to eight asset specific ECB ABS templates. The templates cover the following asset classes: Residential (RMBS), loans to small and medium enterprises (SME), auto ABS, consumer finance ABS, leasing ABS, credit card ABS and commercial (CMBS) mortgage as well as public sector loans. These templates include mandatory fields, e.g. “Account Status”, “Current Balance”, “Payment Frequency” and “Property Type”, for which the data must be populated, together with a number of optional fields that are often also reported. Examples of optional fields include “Origination Channel / Arranging Bank or Division”, “Occupancy Type” and “Number of Debtors”.

An important aspect to highlight is the anonymisation of data. To comply with the various data protection laws, any personal data such as name and address is omitted. Identifier fields like Loan ID and Borrower ID are also encrypted to ensure anonymisation.

As of September 2016, ED stores more than 50m loans across Europe, allowing in-depth loan analysis and key insights into the main drivers of credit performance as well as the ability to “slice and dice” the data and compare it across deals, issuers and countries.

Exhibit 1: Number of loans per country as of September 2016



These optional fields are usually either specific to a country or type of deal.

The submission of loan level data occurs on a monthly or at least on a quarterly basis, no later than one month following the due date for interest payments. Given this frequency of loan data submissions and the consistency of loan identifiers across submissions, specific time series analysis and cohort analysis can be performed.

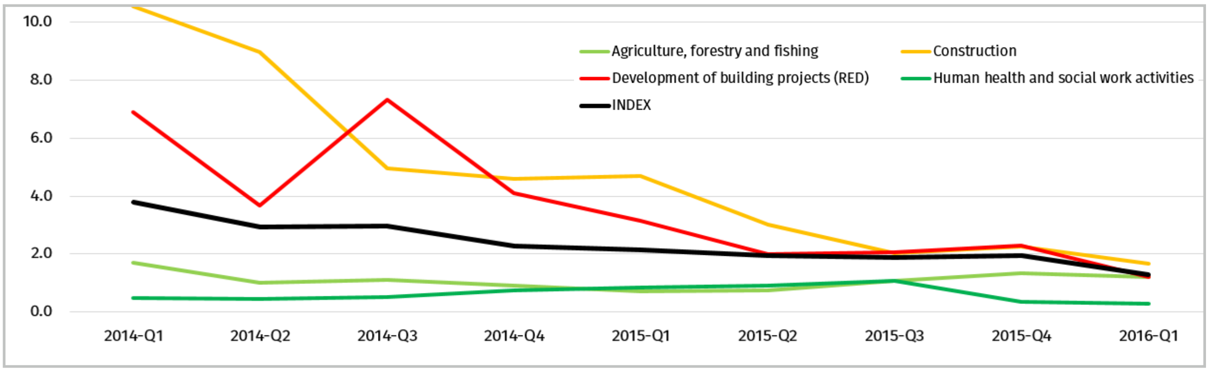
ED has also started to enrich the available data with calculated fields in order to enhance standardisation and improve user experience. As an example, ED will show the actual region name instead of only the postcode or NUTS code. It will also over time provide adjusted property values and performance fields.

Analysts can also use ED’s loan by loan data as input in cash flow and credit scoring models. They can for example estimate future cash flows, default, recovery, volatility and prepayment characteristics of a portfolio.

Individual loans can be selected and classified depending on their individual characteristics such as loan origination date, borrower industry or region in order to ascertain the determinants of performance parameters.

For instance, Exhibit 2 shows Spanish SME delinquency rates per industry. In particular, ED’s LLD make it possible to identify loans to Real Estate Developers (R.E.D.). These loans are typically more volatile than others but are usually not shown separately in investor reports. Loans related to the real estate/construction sector clearly performed worse than others, while non-cyclical sectors such as healthcare and agriculture performed best.

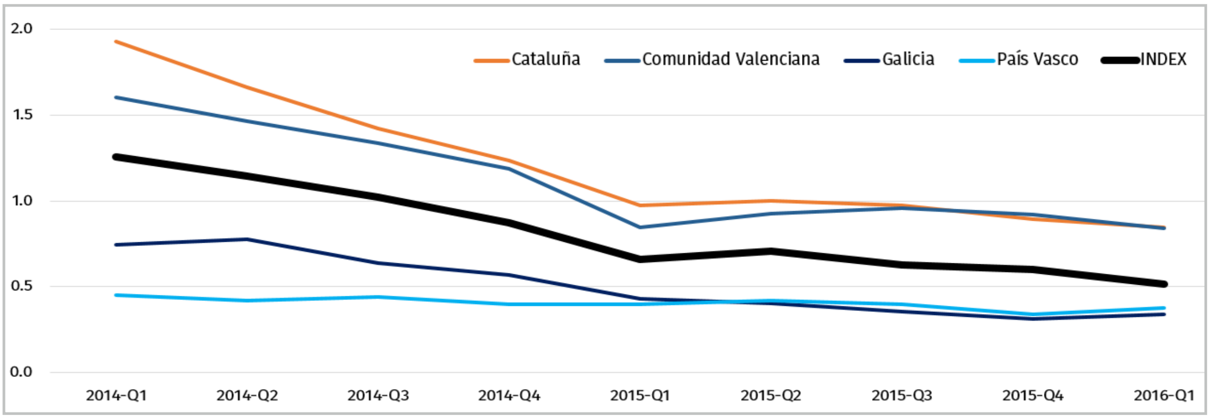
Exhibit 2: Delinquency 90-360 days in Spanish SMEs, per industry (as % of outstanding loans)



Source: European DataWarehouse, Index Commentary

Exhibit 3 uses Spanish residential mortgage data to show mortgage delinquency rates at the regional level. Performance is better in the Basque Country (País Vasco) and Galicia than in regions like Catalonia or the region of Valencia. Because of the generally greater number of observations the lines in Exhibit 3 are smoother than those in Exhibit 2, where e.g. the default of a large borrower caused the spike visible in 2013-Q3.

Exhibit 3: Delinquency 90-360 days in Spanish RMBS per region (as % of outstanding loans)



Source: European DataWarehouse, Upcoming Index Commentary on Spanish RMBS

In addition to insight into the ABS market, ED's data has broader relevance for portfolio management and credit policies given that the underlying portfolios provide data points which are otherwise not available.

- For instance, it is possible to observe loan origination practices for various markets across periods. Given that loan performance is updated on a monthly or quarterly basis, it is possible to observe performance trends and look at the characteristics of delinquent or defaulted loans.

- Researchers interested in the loan market can also find out how loans to a particular category of borrowers are originated and priced, given that the database contains actual interest rates and interest margins. Hence, it becomes possible to find out what interest rates were paid by a certain category of borrowers in a certain market, depending on the characteristics of these borrowers.

ED sample size makes it useful for macro-economic research

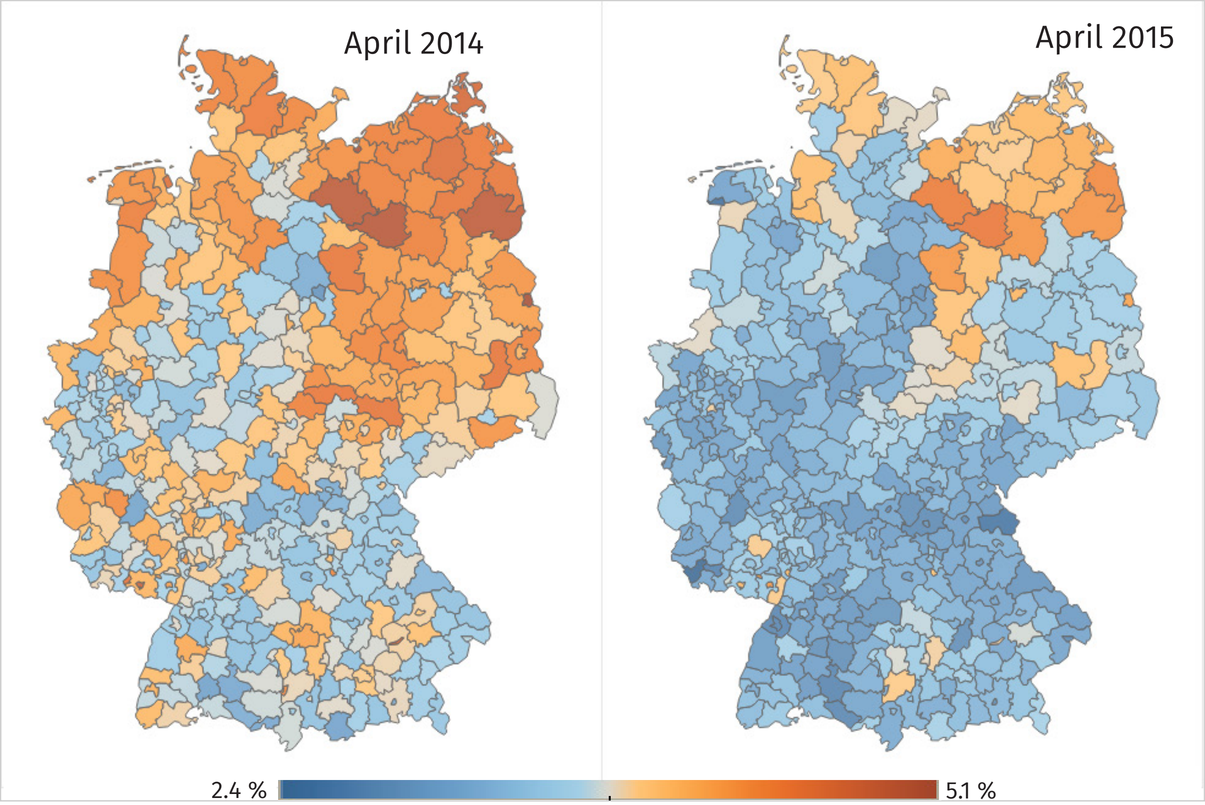
- ED found that the regional distribution of the loans tends to reflect the economic activity in its largest markets such as France, Italy and Spain and that the geographic breakdown of the loans tends to be stable over time in spite of changes in the composition of the loan originators.

- The amount and diversity of the data could also permit research in consumer behaviour. Amongst others, ED loan data features borrower information such as region of residence, date of birth, primary income and property value. This makes it possible to study the main characteristics of borrower buying behavior (e.g. a certain type of property or vehicle) and use this data for consumer credit intelligence purposes.

- Exhibit 4 compares the average fixed interest rates paid on 81,048 auto loans originated in April 2014 to 76,692 loans originated in April 2015 in Germany. The average fixed interest rates paid on auto loans decreased overall, but strong regional disparities are clearly visible. The interest rates paid on auto loans in Southern Germany are lower than those paid in the North East. The overall size of the sample makes it very representative. For instance, the database includes almost one million German auto loans issued in 2015. The availability of NUTS codes for all the German regions allows the precise identification of loan origin down to the city or district level.

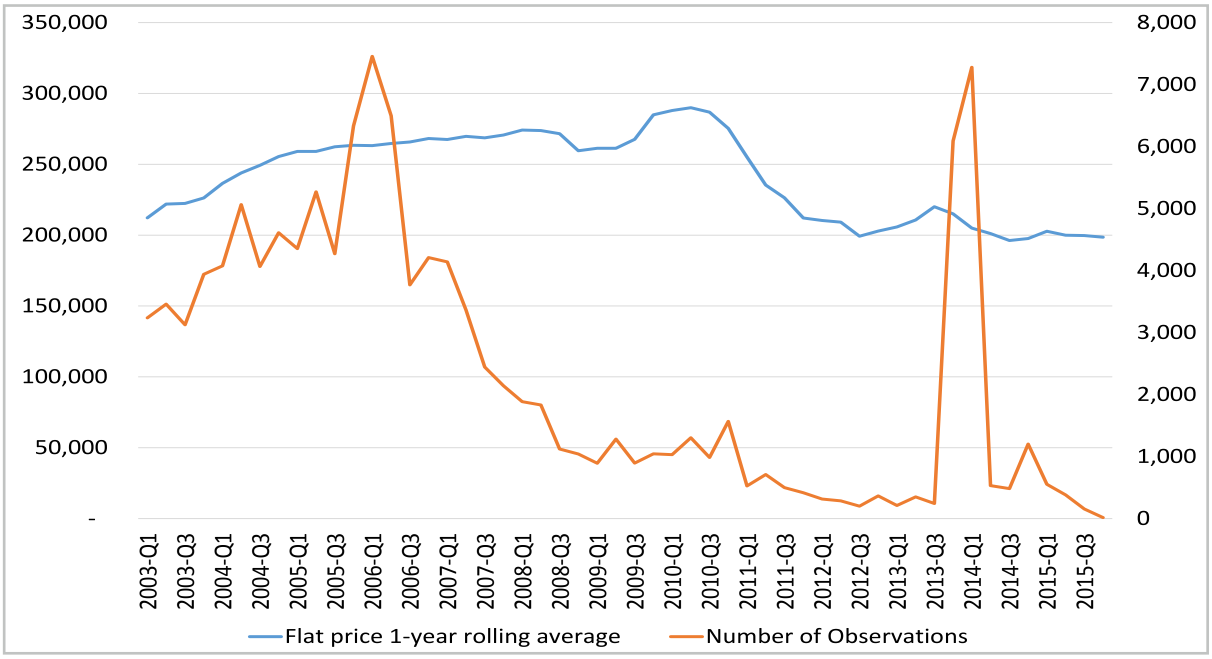
- ED also used static data for Spanish mortgage loans originated since 2003 to calculate a Spanish property price index per region (Exhibit 5). The index is based on the field "original property value". This field is typically filled in at loan origination or prior to securitisation of a portfolio. For the region of Madrid, the first spike in 2006-Q1 reflects the buoyant real estate market at the time whereas the second spike in 2014-Q1 reflects the inclusion of a large deal including properties valued at that time. Although the number of observations fluctuates, this seems to have fairly limited impact in itself on the actual observed price. The last index values for 2015 include a few loans only (154 as for 2015-Q3). The number of observations available for this data point will increase when new Spanish RMBS deals including loans valued in this quarter will be added to the database. This shows how in addition to official statistics a market assessment can be carried out with ED data.

Exhibit 4: Average fixed interest rates paid on auto loans in Germany (loans originated in April 2014 versus April 2015)



Source: European DataWarehouse, Tableau

Exhibit 5: Average flat price in Madrid based on property values at loan origination



Legend: average property value (left hand scale), number of observations per quarter (right hand scale)
Source: European DataWarehouse

Overcoming data quality challenges

The centralised collection of granular credit data and credit risk information at the individual loan level often presents challenges for European institutions in terms of achieving and maintaining good and consistent data quality which adheres to the standardised templates.

As part of its mandate, ED checks the collected loan level data for completeness and correctness. Given the high level of resources, systems and processes deployed for data quality management, ED gathered significant practical knowledge on loan level data quality issues.

While there are many different reasons for data quality issues, they can in general be allocated to one of the following three broad categories:

1. Insufficient clarity of definitions of data fields
2. Erroneous data entries
3. Inconsistencies of the data field content

Each of these issues requires separate solutions to overcome data quality challenges.

1. Data quality problems due to insufficient clarity of definitions of data fields

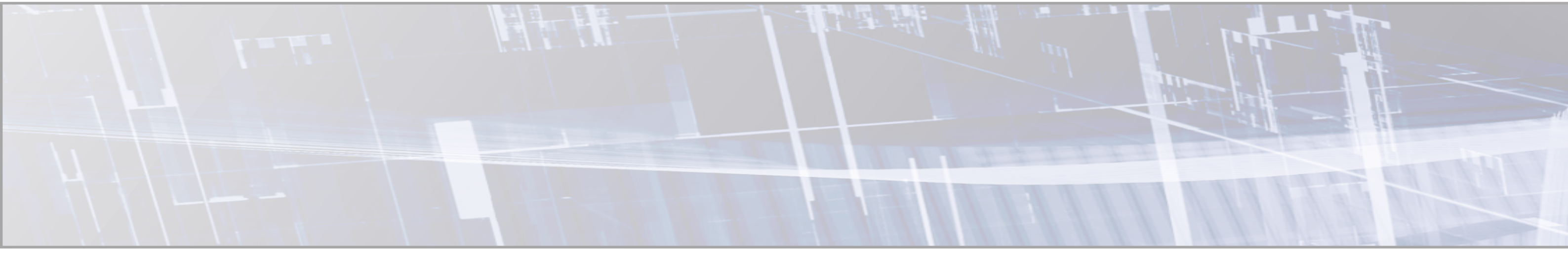
Together with its templates, the ECB also provides dedicated compilation manuals (taxonomies).

For certain fields, the information is reported according to the transaction definition included in the prospectus or offering documentation. Such definitions are not generally standardised and particularly not for performance related fields.

The taxonomies are meant to help issuers in the process of mapping bank-internal data to the required data fields and to avoid potential misunderstandings. However, despite the fact that the taxonomies contain generally comprehensive explanations at field level, there can still be open questions for certain situations which are not clearly addressed in the taxonomy and which require further guidance. Since a taxonomy – as detailed as it might be – cannot cover all possible eventualities, the only way to solve this problem is to prepare a continuously growing collection of case-by-case questions and answers.

This is particularly relevant for areas where national reporting specifications play an important role such as for performance related data fields (arrears, defaults, recoveries etc.). Given the high reporting diversity across Europe, a pan-European

> ***Given the high level of resources, systems and processes deployed for data quality management, ED gathered significant practical knowledge on loan level data quality issues.***



standardisation is difficult to achieve on an ad-hoc basis.

Given the large number of questions it receives, the ECB maintains a dedicated web-page providing a list of ‘frequently asked questions (FAQs)’ concerning the data templates and other reporting issues. In addition, the ECB maintains a helpline which is available for individual clarification requests.

2. *Data quality problems due to erroneous data entries*

European ABS issuers submit the loan level data via the populated data templates to ED. The actual upload into the database follows the XML syntax and, during the upload process, a series of data formatting checks (so called syntax / schema checks) are performed. Should the data not meet these formal criteria the data upload will be rejected. Examples of these formal errors are:

- Entries as text instead of numeric values
- Incorrect date formats (e.g. 12-2015 instead of 2015-12)
- Values outside predefined ranges, e.g. in list fields (e.g. data entries can range from “1” to “6” but the actual entry is “7”)
- Ignoring the required syntax (e.g. currency codes according to ISO 4217, regional names according to NUTS)

Irrespective of these formal checks the data has to be further checked for data errors. While many of these errors can be easily identified at the individual loan level, the high number of loans in any given transaction (there are more than 50m loans in the ED database as of July 2016) requires a multi-step data quality analysis. Typical examples for these data errors are:

- Missing decimal point: data entry 10000000 instead of 100000.00
- Use of proxies or dummies, e.g. borrower’s primary income “0.00” for 80% of all loans
- Duplication of loans with the same loan identifier

In practice, there are numerous data quality feedback loops regarding unresolved data quality issues between the reporting institutions and the recipients. A high level of data quality requires ongoing communication between reporting agents and information recipients.

3. *Data quality problems due to inconsistencies of the data field content*

A comprehensive and focused data quality analysis does not only check for compliance with criteria at data field level but also questions the content-related consistency of the data. In the case of loan level data, such assessment comprises the reconciliation of individual fields within the data template. In addition, it should be checked if the data entry has been in principle correct (i.e. the data correctly reflects the underlying fact) but the taxonomy requires a different definition. In such cases, a transformation would then be necessary. The check for content-related consistency takes into account that a single loan is described through numerous data fields / attributes. In order to assess the reliability of the individual information, data must be organised in a logical context and the content must then be verified. For example, in the case of RMBS, once the amortisation of a loan is made through one single payment at the end of the loan term (field AR72 Payment Type, data entry: “6” Bullet), there should be a corresponding data entry in the field for the repayment method (field AR69 data entry should be “1” for Interest only).

The data quality checks described above have to entirely rely on the submitted loan level data as additional data outside the data templates is typically not available. While many data quality issues can be eliminated through these checks, the data analysis process of ED nevertheless needs to incorporate further analysis based on discussion and feedback with reporting agents. This step is crucial as data reporting in Europe continues to be fragmented given the influence of national accounting or regulatory regimes combined with reporting agent specific issues such as IT related constraints or mixed data standards as a result of merger and acquisition activities.

As a result, the transformation of data required might pose additional challenges as data needed for the transformation might not be available in the reporting agent’s systems. In practice, ED will highlight these problems in individual deal commentaries so that data users fully understand any potential limitations. In addition, where practicable, ED will adjust specific data fields and publish these in addition to the original data in order to make the data sets more comparable.

While many errors can be easily identified at the individual loan level, the high number of loans in a given transaction requires a multi-step data quality analysis.

ED Cloud Pro: bringing large data sets to life

Large data sets, be it financial data or otherwise, are analysed to understand the underlying business developments, trends and risks. Traditionally, such analysis has been left to IT departments and big data specialists, but advancements in data storage and retrieval, combined with the evolution of modern business intelligence tools have enabled the functional experts to self-delve into large data sets. These tools, although powerful and effective, are designed as one-fits-all, creating a steep learning curve and customisation to cater to specific use cases. To understand, analyse and consume Loan Level Data (LLD), with a specific set of use cases, data users typically look for an off-the-shelf solution that is easy to use, that can answer key questions effectively and that just works "out of the box".

ED built from scratch a LLD intelligence solution, solving traditional time consuming problems associated with large data sets like computation of aggregates, stratifications, filters, time-series and comparisons across data sets. The ED Cloud Pro brings a wealth of information straight to the desktop as it is built on Microsoft Excel, providing a user interface that is familiar and easy to navigate.

The Excel based interface enables high-speed data processing of the millions of loans in the database and a user friendly visualisation of the LLD across submissions. The design, driven primarily by functional experts, makes ED Cloud Pro a powerful tool bringing the most important use cases and functions to the forefront whilst abstracting the technology underneath. The design also makes ED Cloud Pro a compelling solution for users across the entire spectrum including risk management professionals, researchers, rating agencies, investors, issuers and academics alike.

The ABS loan level database architecture

The ED Cloud Pro loan data intelligence solution is built on a powerful SQL Server Database, leveraging both standard database storage methods and the modern column store technology. This technology enables users to look into the loan level data where the use case is specific to certain data fields (columns) , for example loan-to-values (LTVs), interest rates, etc.

This dual storage designs enables both (1) securitisation use cases like looking at bond and loan records for a particular deal and (2) analytical / research oriented use cases like interest rate analysis, LTV analysis, property information etc.

> ***The ED Cloud Pro brings a wealth of information straight to the desktop as it is built on Microsoft Excel, providing a user interface that is familiar and easy to navigate.***

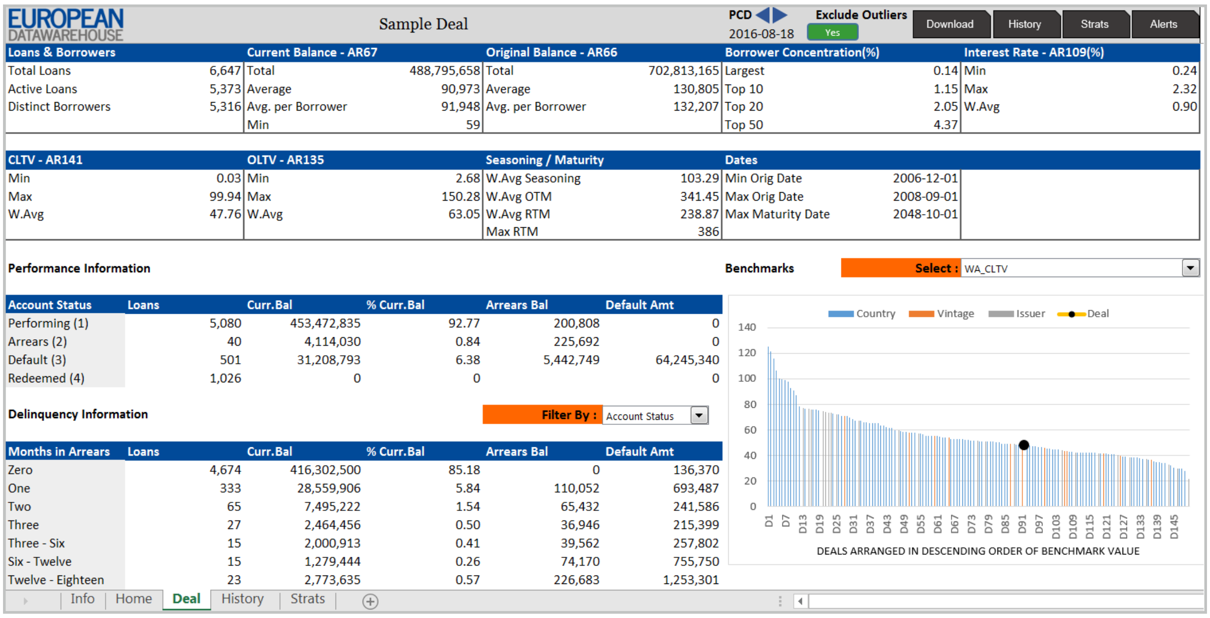
An Excel interface to the loan level database

To cater to the needs of non-technical users who are not familiar with SQL, the ED Cloud Pro solution employs a user friendly Excel Workbook that provides key securitisation functionalities for each transaction – including stratification tables, weighted averages, benchmarks and the ability to compare one deal with one or more deals across issuers, vintages, etc. The focus on performance is part of the ED Cloud's design using column-store database technology that enables exponential performance gains in comparison to standard databases.

The Workbook also contains special features such as the ability to perform loan portfolio analysis - wherein a selection of loans (irrespective of deals) can be stratified and filtered.

The Excel Workbook is fed with ED's data through HTML (Web) pages and therefore is technically very similar to browsing a website. This design ensures users can use the Excel Workbook directly out of the box without the need of any software installation / Excel Add-ins, etc.

The primary users of ED's loan level data – the ABS investors – look at loan data sets typically on a deal-by-deal basis. Naturally, ED has been hosting a copy of the loan by loan data as separate Excel / CSV files – with one file containing the loan data for each deal. ED Cloud Pro's design injects a high value into ED's universe of data by empowering users to manage large datasets across submissions. ED Cloud Pro also allows users to perform loan portfolio analysis by looking at a set of loans – irrespective of the deals they belong to -and compare or slice them by country, vintage, etc. including time series evolutions.



Sample Deal

PCD 2016-08-18

Exclude Outliers: Yes

Loans & Borrowers		Current Balance - AR67		Original Balance - AR66		Borrower Concentration(%)		Interest Rate - AR109(%)	
Total Loans	6,647	Total	488,795,658	Total	702,813,165	Largest	0.14	Min	0.24
Active Loans	5,373	Average	90,973	Average	130,805	Top 10	1.15	Max	2.32
Distinct Borrowers	5,316	Avg. per Borrower	91,948	Avg. per Borrower	132,207	Top 20	2.05	W.Avg	0.90
		Min	59			Top 50	4.37		

CLTV - AR141		OLTV - AR135		Seasoning / Maturity		Dates	
Min	0.03	Min	2.68	W.Avg Seasoning	103.29	Min Orig Date	2006-12-01
Max	99.94	Max	150.28	W.Avg OTM	341.45	Max Orig Date	2008-09-01
W.Avg	47.76	W.Avg	63.05	W.Avg RTM	238.87	Max Maturity Date	2048-10-01
				Max RTM	386		

Performance Information

Account Status	Loans	Curr.Bal	% Curr.Bal	Arrears Bal	Default Amt
Performing (1)	5,080	453,472,835	92.77	200,808	0
Arrears (2)	40	4,114,030	0.84	225,692	0
Default (3)	501	31,208,793	6.38	5,442,749	64,245,340
Redeemed (4)	1,026	0	0		0

Delinquency Information

Filter By : Account Status

Months in Arrears	Loans	Curr.Bal	% Curr.Bal	Arrears Bal	Default Amt
Zero	4,674	416,302,500	85.18	0	136,370
One	333	28,559,906	5.84	110,052	693,487
Two	65	7,495,222	1.54	65,432	241,586
Three	27	2,464,456	0.50	36,946	215,399
Three - Six	15	2,000,913	0.41	39,562	257,802
Six - Twelve	15	1,279,444	0.26	74,170	755,750
Twelve - Eighteen	23	2,773,635	0.57	226,683	1,253,301

Benchmarks

Select : WA_CLTV

Please visit www.eurodw.eu for more information

Contact information:

European DataWarehouse
Walther-von-Cronberg Platz 2
60594 Frankfurt am Main
Germany

Tel. +49 (0) 69 8088 4300

Email: enquiries@eurodw.eu

EUROPEAN DATAWAREHOUSE