IFC-National Bank of Belgium Workshop on "*Data needs and Statistics compilation for macroprudential analysis*"

Brussels, Belgium, 18-19 May 2017

Use of AnaCredit granular data for macroprudential analysis[1]

Orestes Collazo Brananova and Gibran Watfe,
European Central Bank

---

[1] This paper was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Use of AnaCredit Granular Data for Macroprudential Analysis

Orestes Collazo, Gibran Watfe

Abstract

AnaCredit will provide a rich set of granular data on credit and credit risk for macroprudential analysis allowing the calibration and assessment of macroprudential instruments. Adapting financial indicators to changing conditions over time is facilitated by the extensive breadth of counterparty and credit data attributes collected in AnaCredit. Compared to the compilation of aggregate statistics, individual data (loan-by-loan and counterparty-by-counterparty) allows the analysis of the distribution of risk across the whole population.

This paper explores macroprudential instruments that can be defined using AnaCredit data, and how they can be used to drill down in the main sources of risk concentration. The paper concludes exploring the need to integrate AnaCredit with other datasets in order to fully exploit the granular data available.

Keywords: Analytical Credit Dataset, loan-by-loan data, credit risk, macroprudential analysis

JEL classification: C81, E44, E51, E58, G28

Contents

1. Introduction

AnaCredit will provide a rich set of granular data on credit and credit risk for macroprudential analysis and various other purposes. Regulation ECB/2016/13 adopted in May 2016 is the legal basis for data collection starting in November 2018 with September 2018 as a first reference date. The breadth of the 88 AnaCredit data attributes in conjunction with its granularity and the fact that data is collected on an individual basis will allow for adapting financial indicators to changing conditions over time. The AnaCredit framework will support the calibration and assessment of macroprudential instruments and thus contribute to financial stability in line with the ESCB's task specified in Art. 127(5) TFEU.

Economic and financial indicators attempt to capture the behaviour of economic agents and describe complex economic phenomena using a small number of metrics. Defining indicators is usually a long process starting from real economic events that become important and which reveal a specific data gap. On this basis indicators are defined either specifically for a particular purpose or as a one-size-fits-all solution for several purposes. The former allows a detailed analysis on a specific issue but often does not cover the data needs for different analyses; while the latter may only be of limited interest for a larger number of analysts. The detailed granular data in AnaCredit will allow the definition of very specific indicators for a large number of different analyses.

The approach in the past, when no granular data were available, was to define indicators and calculate a single metric representing the aggregate behaviour of very heterogeneous agents. Describing the complex behaviour of agents by means of a small set of indicators led to a substantial loss of information. The dispersion of the indicator across the population was condensed into a simple or weighted average. In this process the complex relationships between agents were ignored. Moreover, the definition of the indicator was fixed ex ante and could thus hardly be adapted to changing circumstances.

Individual data by (borrowing or guaranteeing firms) and, in particular, granular data on credit can alleviate the information loss considerably by using a different approach: measuring economic phenomena at the level at which they occur. The AnaCredit data model was conceived from this perspective. AnaCredit is a granular database modelling credit intermediation on an individual instrument-by-instrument, counterparty-by-counterparty and protection-by-protection basis as well as the relationship among these three building blocks. Therefore, it offers three distinct advantages for defining financial indicators relative to the broad approach in the past.

Firstly, AnaCredit provides an unprecedented level of detail for analysing credit intermediation and defining related indicators for the euro area. The database contains 88 attributes that were defined based on the needs of a variety of business areas. This high level of detail in itself provides flexibility to define indicators, including the possibility for back-casting, as well as allowing calibration and estimation of models, e.g. in the context of stress testing. The dimensions and measures included in AnaCredit can be combined in multiple ways to measure

economic phenomena at different levels. The attributes and values provide a starting point for improving existing indicators or defining new ones in a flexible manner. This approach, in particular, stands in contrast to the current methodology to establish indicators to track economic events that in the *past* have been found of relevance. While these stable indicators are undoubtedly useful, a more dynamic way to track current economic phenomena as they occur is needed.

Secondly, AnaCredit will allow users to both identify and construct networks between agents. This stems from the unique identification of counterparties in AnaCredit based on the reporting of unique and exclusive identifiers. This will help, among others, to analyse concentration of risk at different levels, e.g. according to banking and corporate groups, or according to sectors of economic activity.

Thirdly, given that AnaCredit facilitates the measurement of indicators at an individual level, analysis of credit can go beyond the averages. Within the scope of AnaCredit, the distribution of an indicator across the whole population can be identified. Hence, the user can focus on specific parts of the distribution. Most importantly, the tail(s) of the distribution reveal important insights about particularly risky counterparties or sectors for financial stability purposes.

Finally, AnaCredit will be complemented by a full register of financial, governmental and non-financial entities. Data on counterparts involved in transactions will be available in the Register of Institutions and Affiliates Database (RIAD) both at individual (legal) entity level and on group structures. This will also allow for identification, classification by size of firms, by industry activity, but also to consolidate exposures on the lending side and indebtedness on the borrowing side.

Aggregated data and derived indicators currently do not cover the new tasks and needs of financial stability. More granular information is needed to analyse credit intermediation and other parts of the economy to better inform decision-makers. AnaCredit is needed in order to fully understand the financial stability implications of credit intermediation. This paper describes how AnaCredit can contribute to macroprudential analysis.

The paper is structured as follows. The next section describes the main elements of the AnaCredit data model and points out important attributes for macroprudential analysis. Section three looks into the definition of financial stability indicators with AnaCredit data. It also includes a case study on real estate. Section four describes the use of AnaCredit data for exploring financial indicators from the top down and thus provides a different perspective. Possible future enhancements of AnaCredit are described in Section five on the basis of identified data gaps.

2. AnaCredit data model

The AnaCredit conceptual data model describes the broad structure of the AnaCredit database. As AnaCredit is an instrument-by-instrument database, the data model is structured on the basis of modelling credit. In a more detailed manner, the logical data model specifies each part of the structure of AnaCredit

including all attributes contained therein. It also specifies all required primary and foreign keys which allow an identification of relationships between entities. Annex 1 provides a graphical representation of AnaCredit conceptual data model[1].

Main elements

The conceptual data model aims to replicate any financial intermediation on the basis of the three main distinct but interconnected arising in each credit transaction: instrument, counterparty and protection. The instrument entity is at the heart of AnaCredit. It contains the information about the individual credit, i.e. individual commitments with unique terms under a credit agreement such as product type, maturity date, interest rate, etc. Instruments arise under a contract. The AnaCredit model allows registering multiple instruments under a single contract to identify instruments which raise common risk for the counterparties involved in these instruments.

Whenever an instrument is secured by a protection, the latter is captured in the separate protection entity. The protection entity contains information about all protection items (i.e. valuable assets or rights) that are committed to the fulfilment of the terms of an instrument (i.e. that secure the payments under a credit transaction), as specified in the (credit) contract that gives rise to the instrument. All protection items in the protection entity contain an instrument identifier such that the protection item can be mapped to the instrument it secures. Detailed information about protection (e.g. value, valuation date and type) is potentially important for estimating recoveries in macroprudential stress test exercises.

All protections securing an instrument are registered in AnaCredit. Moreover, a single protection may secure multiple instruments; however, the protection is only uniquely identified locally, i.e. at the level of the credit institution reporting the information to AnaCredit. If the same protection is securing two loans held by the same credit institution, then the protection is uniquely identified. However, if the same protection, e.g. commercial real estate, is securing two loans held by different credit institutions, two different protections appear in AnaCredit. To overcome this limitation, AnaCredit includes the attribute 'third party priority claims against the protection' which registers the maximum amount of any existing higher ranked liens against the protection with respect to third parties. This allows a quantification of the share of the protection which is committed with third parties.

The counterparty entity contains information about counterparties related to the instrument entity and, if relevant, to the protection entity. The counterparty entity also contains information about certain counterparties affiliated with debtors and

---

[1]   See AnaCredit reporting manual Part I – General Methodology for a detailed description of the AnaCredit conceptual model.

protection providers of an instrument in the instrument entity. In a similar way to the modelling of the instrument and protection entities, the modelling of the counterparty entity takes into account the fact that the same counterparty may relate to several instruments and protection items. For the purpose of AnaCredit, all counterparties which take any of the following roles should be reflected in the counterparty entity:

1. the creditor of the instrument;
2. the originator of the instrument, if the instrument is a securitisation transaction;
3. the servicer of the instrument;
4. the debtor of the instrument;
5. the protection provider that provides protection to the instrument (if any protection item is provided);
6. the head office undertaking of (any foreign branch or a special fund that is) a debtor of the instrument or a protection provider that provides protection to the instrument;
7. the immediate parent undertaking of any debtor of the instrument, or of any protection provider that provides protection to the instrument;
8. the ultimate parent undertaking of any debtor of the instrument, or of any protection provider that provides protection to the instrument.

The individual connections between the instrument and the counterparties participating in the instrument allow identifying all relevant risks that each counterparty is holding or raising. Concretely, both joint liabilities where multiple debtors are jointly liable for a credit, and instruments held by multiple creditors can be identified. In addition, AnaCredit allows connecting the multiple shares of a syndicated loan held by different credit institutions.

In contrast to the local identification of the protection, AnaCredit provides universal identification of counterparties. The individual identification of all counterparties participating in the instrument, regardless of their residency or type, provides a comprehensive view of the level of indebtedness and credit that each debtor and creditor hold.

On the basis of the conceptual data model, the logical data model of AnaCredit comprises six entity tables which in turn include one or more actual datasets.

1. the instrument entity table includes three datasets: instrument data, financial data, accounting data;
2. the protection entity table corresponds to protection received data;
3. the instrument-protection received entity table corresponds to instrument-protection received data;
4. the counterparty reference data entity table corresponds to counterparty reference data;
5. the counterparty risk/default entity table includes two datasets: counterparty risk data and counterparty default data; and

6. the counterparty-instrument entity table includes two datasets: counterparty-instrument data, joint liabilities data.

This structure enables the user to identify relationships among instruments, counterparties and protections. Of these connections, the relationships among counterparties may prove of significant relevance for macroprudential analysis as they allow identifying concentration of risk in areas of activity and possible contagion effects.

Attributes

AnaCredit contains 88 data attributes collected with a monthly or quarterly frequency depending on the dataset and, possibly, on derogations for certain small reporting agents. The large number of attributes enables the user to conduct extensive analyses with AnaCredit data. Next to macroprudential analysis, users from various business areas in the ECB and NCBs will make use of the data. From the perspective of macroprudential analysis multiple categories of attributes are of particular interest. Moreover, the use of standards and common definitions across countries in AnaCredit will enhance the comparability of results in the Euro Area.

A significant set of attributes inform about the financial aspects of a particular instrument. Examples include interest rate, outstanding nominal amount, off-balance sheet amount, transferred amount, default status, arrears and date of past due. The combination of outstanding nominal amount, transferred amount and the unique identification of counterparties in RIAD allow, for example, to track securitisation transactions in which the credit institution continues to service an instrument after the transfer.[2] The attributes default status, arrears and date of past due are important for the analysis of NPLs in particular given that these attributes are reported at a monthly frequency.

An important aspect for macroprudential analysis is that AnaCredit contains both backward-looking (loss measures) and forward-looking (risk measures) attributes. For example, attributes such as transferred amount, accumulated impairments and cumulative recoveries since default contain cumulative information from the past. In addition, the history of data provided in AnaCredit will enable the precise calculation of the payment history of debtors, e.g. debtor days past due and largest days past due within the last 3 months (across all instruments). Other attributes in AnaCredit are forward looking, enabling the user to estimate the future cash flows of an instrument (e.g. amortisation type) or the subjective probability of a negative credit event using the probability of default. This attribute can serve as an ex ante expected default rate and for the calculation of expected credit losses. While expected credit losses are based on the creditor probability of default, the probabilities of default provided for a single debtor by multiple credit institutions

---

[2] For a more detailed description of RIAD, see Neudorfer (2016) and Thijs & Corvoisier (2017).

may allow the calculation of a more objective measure of default which may then be used for industry-wise stress test purposes. In addition, related derived data may be defined to include various measures of industry probabilities of default allowing for sectoral risk analysis.

Counterparties are described in detail by way of reference data attributes that are collected under AnaCredit and stored in RIAD. Attributes inform among others about the name, the address, the legal form, institutional sector, economic activity as well as the size of a counterparty. Moreover, the counterparty risk/default table contains attributes on the probability of default and the default status. As described above the counterparties in AnaCredit take different roles and can thus be connected based on the instruments in which they have one or more of the roles.

In the context of analysing concentration of risk, for example, the analyst could make use of the identifiers relating counterparties with their parent companies and using outstanding nominal amounts in order to gauge the risk at group level. Alternatively, risk concentration based on location could be analysed with the help of address attributes, e.g. Address: country. Finally, the analyst could exploit the attribute on economic activity to identify risk concentrating in specific sectors of the economy.

Another example relevant for financial stability purposes is the analysis of exposure levels. Absolute exposure might be based, for instance, on the sum of outstanding nominal amount and off-balance sheet amount. It could potentially be widened by taking into account transferred amount, joint liability amount and carrying amount. In one specific use case, exposures to non-financial corporations of different size could be derived based on the attribute enterprise size. The aspect of risk might be introduced by including the attribute probability of default in the analysis.
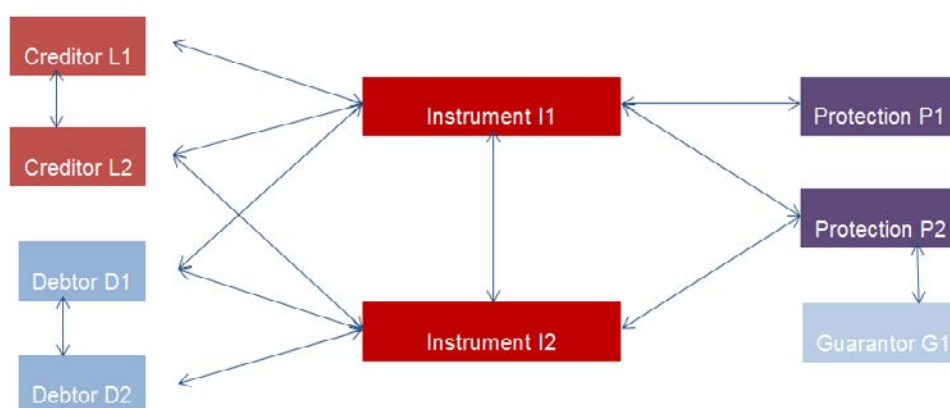
Connections among counterparties

Figure 1: An illustration of n-to-m relationships between instruments, counterparties and protections that can be traced with AnaCredit data.

A central feature of AnaCredit is that it allows users to infer information from the relationships among counterparties. Connections can take different forms: connections that identify an entities' position within a group, connections from the

perspectives of contagion and concentration as well as aggregation of lenders with similar characteristics (e.g. institutional replicas).

First of all, AnaCredit collects information on group structures which are enhanced via RIAD combining other sources of information. Specific identifiers for head offices, immediate parent and ultimate parent undertakings are reported. Hence the position of a particular counterparty within a group can be identified. This information may allow the users to prepare pseudo-consolidated data for a group on the basis of the individual data provided. This applies to both banking and corporate groups as reference data is collected for creditors and debtors. Moreover, as reference data will be updated on a monthly basis, mergers can be tracked in a timely manner.

Secondly, AnaCredit will be important for macroprudential analysis of contagion effects both ex ante and ex post. The coverage of the database enables the user to analyse positions within the MFI sector and thus allows for inference on possible contagion through interbank markets. Possible contagion effects from the financial to the real sector or vice versa can also be analysed with AnaCredit data. A caveat is that AnaCredit does not contain information about natural persons and will thus not allow for possible contagion effects stemming from and/or affecting household exposures. Importantly, contagion effects can not only be analysed on the basis of direct exposures (i.e. loans or securities), but also based on collateral exposures. AnaCredit will mainly inform about positions in the form of loans while the Securities Holdings Statistics Database (SHSDB) contains the relevant information about securities. Efforts are currently ongoing within the ESCB to facilitate an integration of the two databases.

Thirdly, and related to the aspect of group structures, AnaCredit will facilitate the analysis of risk concentration. For instance, the exposures of creditors within the same banking group can be aggregated to analyse the degree of concentration within the group. Similarly, the degree of dependence of debtors within the same corporate group on particular lenders can be analysed with AnaCredit data. In addition to that, the attribute on economic activity of a counterparty enables the user to identify sectoral concentration and dependencies of particular lenders or of a banking group.

Another type of analysis that is facilitated by being able to connect counterparties in AnaCredit is the construction of portfolio replicas. The portfolio of a group of small banks taken together might replicate the portfolio of big, possible systemically important, lenders. Alternatively, banks with similar business models might be grouped together. AnaCredit attributes help the user to identify concentration of exposures within sub-groups of lenders and thus to trace another form of macroprudential risk over time. Finally, the data attributes enable the user to analyse instruments with one or more lenders (e.g. syndicated loans) or with multiple debtors as well as to exploit the connections between multiple instruments connected to one or more protection items.

3. Definition of financial stability indicators

AnaCredit offers the possibility to construct a large amount of financial indicators for various purposes including macroprudential analysis. AnaCredit was developed as a multi-purpose dataset. The granularity of the data in AnaCredit enables the user to obtain information about individual instruments and individual counterparties. Financial stability indicators can therefore be constructed and tracked over time on instrument and counterparty level. Moreover, both the quantity and the variety of the attributes make it a useful tool to develop new indicators. The unique identification of counterparties as well as the use of harmonised concepts and definitions facilitates a connection of AnaCredit with other granular datasets such as the SHSDB.

Flexibility

A challenge often encountered in the past was that data needs became evident only after crisis events unfolded. This is frequently exacerbated by long time lags between formulating data requirements and the provision of reliable data, let alone histories sufficient for meaningful inference. To some extent, granular data will rectify this challenge as it can be aggregated and used in various contexts. Granular data can help to understand, for example, underlying forces at work during an imminent crisis in a particular sector which cannot be foreseen ex ante. Changing conditions can hence be addressed with current information without the need to establish new and lengthy reporting requirements. The monthly frequency and relatively tight time frames for reporting agents will make AnaCredit useful to understand unfolding events almost in real time compared to existing macro datasets on credit.

Distributions

Apart from tracking financial indicators over time, the individual information in AnaCredit allows the user to look at full distributions of financial indicators for nearly the whole population. For example, macroprudential analysis may use various aggregates of probabilities of default and their development over time as important measures for financial stability. Apart from that, probabilities of default assigned by all lenders of a particular debtor can be monitored. On a more aggregated basis the probabilities of default assigned by lenders to particular sectors might be checked.

More generally, distributions derived from AnaCredit data may be used to enrich aggregated macro datasets and thereby provide macroprudential policy with more precise information. Possible aspects that enrich the commonly used average figures are measures of central tendency, dispersion and concentrations. This will reveal a new sort of information previously hidden behind averages (see Figure 2). This allows analysts, for example, to focus on the tail of the distributions in order to identify particularly risky (groups of) counterparties. Finally, the information value of AnaCredit can best be harnessed when indicators are combined in order to obtain a more holistic view. A practical example for macroprudential analysis would be the

combination of loan-to-value, NPL ratio and bank size. Another option would be to introduce correlations between indicators to increase the breadth of the analysis.
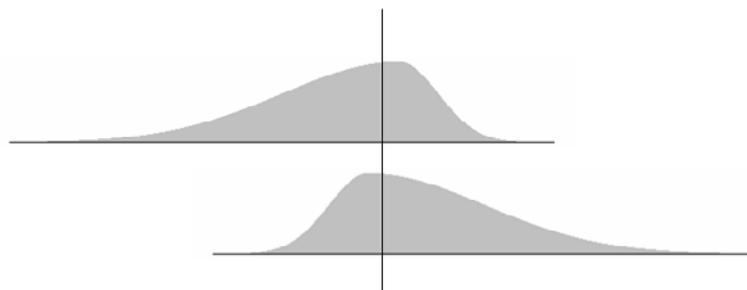

Figure 2. Illustration of two different distributions with the same mean.

Case study: real estate data gaps

In October 2016, the ESRB published a Recommendation on closing real estate data gaps (ESRB Recommendation)[3]. It lists various key indicators for monitoring developments in the residential and commercial segments of the real estate market. National macroprudential authorities and the European Supervisory Authorities (ESAs) are recommended to collect data for these indicators from existing data sources. The real estate data gaps identified in the Recommendation serve as a case study for the use of AnaCredit data.

The Recommendation explicitly refers to AnaCredit as a possible source for the proposed indicators. However, it points out that AnaCredit cannot be relied alone for meeting the information needs identified in the Recommendation due to some of its features:

- The definitions for residential real estate (RRE) and commercial real estate (CRE) differ;

- Natural persons are beyond the scope of AnaCredit;

- Some key indicators and market segments are not covered;

- AnaCredit focuses on Euro Area members;

- Loans from non-credit institutions are only covered if they are serviced by a credit institution;

- Small banks might be excluded by way of derogations.

---

[3]   Recommendation of the European Systemic Risk Board of 31 October 2016 on closing real estate data gaps (ESRB/2016/14) (European Systemic Risk Board, 2016)

Despite these caveats, our analysis will show that AnaCredit could be used to cover some of the indicators required in the Recommendation. Moreover, an extension of AnaCredit to cover loans granted to natural persons would significantly increase the coverage of the indicators proposed in the Recommendation.

A detailed analysis of the information in AnaCredit reveals that while the definition of CRE and RRE in the Recommendation and in AnaCredit differ, the use of additional information included in AnaCredit may reduce the gap between the two definitions.

In the context of AnaCredit CRE and RRE are not defined in isolation but used as part of two attributes, namely 'purpose' and 'type of protection'. The attribute 'purpose' classifies AnaCredit instruments according to their purpose. Two values are of relevance for the identification of CRE loans, namely 'residential real estate purchase' which applies to instruments financing residential property, and 'commercial real estate purchase' which applies to instruments financing real estate property other than residential property.

On the other hand, the attribute 'type of protection' allows the identification of the type of protection securing an instrument reported to AnaCredit. Three values are of relevance in this context:

a) Residential real estate collateral which is defined as residential property in accordance with Article 4(1)(75) of CRR;

b) Offices and commercial premises which is defined as real estate other than residential real estate that qualifies as "offices or other commercial premises" for the purposes of Article 126(1) of the CRR; and

c) Commercial real estate collateral which is defined as any real estate property other than residential real estate collateral and offices and commercial premises.

The distinction between offices and commercial premises and commercial real estate collateral is of no relevance for the identification of CRE loans as it is based on the relationship between the collateral and the creditworthiness of the debtor. Thus we will refer to the combination of both as commercial immovable property.

The values of these two attributes rely on the definition of residential property within the meaning of Article 4(1)(75) of the CRR. Residential property is defined as residence which is occupied by the owner or the lessee of the residence. Consequently, any real estate property that is not residential real estate is considered to be commercial immovable property. On the other hand, the ESRB Recommendation defines CRE as any income-producing real estate, either existing or under development, and excludes: (a) social housing; (b) property owned by end-users; and (c) buy-to-let housing. The relevance of this definition lies on its use to identify CRE loans, which are defined as loans aimed at acquiring a CRE property (or set of CRE properties) or secured by a CRE property (or set of CRE properties). The indicators suggested focus on this loan portfolio.

A direct approach to identify CRE loans in AnaCredit would be to focus on loans whose purpose is to finance commercial real estate or secured by commercial

immovable property. As indicated in the ESRB Recommendation, the definitions used in both frameworks are not aligned, and the differences may be significant. However, this mapping can be further refined. The loans registered in AnaCredit are granted to legal entities[4]. As a result, the loans granted to legal entities for the purpose of financing residential real estate purchase or secured by residential property should satisfy the CRE loan definition as this real estate is expected to be income-producing real estate. Moreover, these loans do not include buy-to-let housing[5] as those are loans granted to natural persons.

It is important to note that this portfolio may include loans financing social housing or protected by this type of real estate. However, once more the portfolio can be refined to exclude these loans. Social housing is not defined in the ESRB Recommendation, however, it is commonly understood as real estate owned by the state or non-profit organisations which is rented to provide affordable housing. As a result, these loans can be identified as i) loans granted to the general government or to a non-profit institutions serving households (NPISH), or as ii) loans secured by residential real estate property provided by the general government or by a NPISH. These categories can be identified in AnaCredit by means of the institutional sector of the debtor and protection provider of the loans.

The portfolio of loans defined so far may still diverge from the CRE loans portfolio defined in the ESRB recommendation as it may include loans aimed at acquiring property owned by end-users which is not income-producing real estate or secured by such property, and which are excluded from the definition of CRE loans. The identification of such loans can only be approximate in AnaCredit, and a new property would be needed to identify this category. A proposed approach to identify these loans under the current AnaCredit specification would be to exclude loans granted to debtors in economic sectors of activity which are not expected to acquire income-producing commercial real estate and which are secured by commercial immovable property.

While a perfect matching of CRE loans defined in the ESRB Recommendation is not possible with the current definitions, a better approximation of CRE loans can be achieved by combining additional AnaCredit data. It remains to be assessed, once data is available, to which extent the proposed mapping would cover the CRE loan market defined in the ESRB Recommendation, also taking into account the differences in scope indicated before. Similarly, residential real estate loans will be covered in AnaCredit when and if the scope of borrowers is extended to natural persons, but currently does not provide information for this segment.

---

[4]   In case natural persons are included in a future stage, natural persons and legal entities will be distinguishable in AnaCredit on the basis of the attribute 'legal form' of the counterparty.

[5]   'Buy-to-let housing' means any RRE directly owned by a private household primarily for letting to tenants.

The indicators defined in the ESRB Recommendation can be broken down into a set of variables needed for their calculation. We propose a mapping of AnaCredit attributes to calculate the variables used in the indicators proposed in the ESRB Recommendation:

- Current lending: this variable is accurately captured by the attribute 'outstanding nominal amount' at each reference date.

- Lending at origination: the mapping requires distinguishing loans which may have a loan commitment associated to them (e.g. credit line) from those which do not. For the latter, this variable is correctly captured by the 'commitment amount at inception'. When the instrument allows for a loan commitment, the variable can only be estimated by the attributes 'commitment amount at inception' and 'off-balance sheet amount'. These values can be further refined for loans granted after AnaCredit goes live in September 2018 by using the 'outstanding nominal amount' at the first reference date after the loan was originated.

- Current property value: an accurate mapping using the attributes 'protection value' and 'type of protection' to distinguish real estate collateral.

- Property value at origination: this variable can be accurately mapped by the attributes 'original protection value' and 'type of protection'.

- Loan service at origination: an estimate can be calculated on the basis of the 'interest rate', 'amortisation type' and the history of 'outstanding nominal amount' and 'accrued interest' reported.

- Total debt: out of scope

- Debt service: out of scope

- Disposable income at origination: out of scope (this variable is related to natural persons which are not in the scope of AnaCredit).

- Maturity at origination: this value can be accurately mapped by the attributes 'settlement date' and 'legal final maturity date'.

- Non-performing loans: this value can be accurately mapped by the attribute 'performing status of the instrument'.

- Loan-loss provisions: this value can be accurately mapped by the attribute 'accumulated impairment amount'.

- Loans in foreign currency: mapped by the attribute currency

The ESRB Recommendation also proposes a series of breakdowns which are partially available in AnaCredit.

- Buy-to-let: Out of scope (this variable is related to natural persons which are not in the scope of AnaCredit).

- Owner-occupied: Out of scope (this variable is related to natural persons which are not in the scope of AnaCredit).

- First time buyers: Out of scope (this variable is related to natural persons which are not in the scope of AnaCredit).

- Fully amortising / partially amortising: the attribute 'amortisation type' allows identifying the combination of these two categories. However, the separation of the two categories can only be estimated.

- Non-amortising: this value can be accurately mapped by the attribute 'end date of interest-only period'.

- Initial interest rate fixation period: this value can be accurately mapped by 'initial interest rate fixation period'.

- Renegotiation: this value can be accurately mapped by 'status of forbearance and renegotiation'.

- Property type: this value can be accurately mapped by 'type of protection'. However, one-to-one mapping is not available. Coarser categories have to be used (ESRB: office, retail, industrial, and other are mapped to CRE in AnaCredit and residential to RRE).

- Property location: this value can be accurately mapped by 'real estate collateral location'.

- Lender / investor type: this value can be accurately mapped by 'economic activity'.

- Lender / investor nationality: this value can be accurately mapped by 'address: country'.

- Property under development: this value can be estimated using the attribute 'purpose' with the value 'construction investment'.

The ESRB proposes the presentation of the distribution of some indicators on the basis of a fixed set of categories (e.g. 8 categories used for the maturity at origination of RRE loans). The use of AnaCredit data enables the calculation of the complete univariate distribution of any indicator calculated across the whole population or a specific subset of the population. Once the complete distribution is calculated, the user may choose to define different categories (e.g. percentiles) to describe its different areas or define statistics (e.g. standard deviation).

Moreover, the ESRB Recommendation also proposes joint distributions of indicators. In this area AnaCredit proves to be of great value as it allows the combination of as many indicators or variables as the user considers relevant for the analysis without pre-specifying any of those. Due to the potential exponential growth of indicators by combining different variables, the ESRB restricts the combination of indicators to two indicators (see for instance the joint distribution of loan-to-value vs. loan-service-to-income proposed in the ESRB Recommendation).

The ESRB Recommendation proposes two different data types to register their indicators, flows and stocks depending on the indicator. Stocks are directly captured in AnaCredit. As regards flows, the ESRB definition can be accurately matched in AnaCredit by using the attribute status of forbearance and renegotiation. AnaCredit makes no distinction and all indicators created can be defined in terms of stocks and flows without the need to make a distinction on the basis of the indicator. Moreover, the ESRB Recommendation contains three different indicator metrics, namely amount in national currency, number of contracts and average. The use of granular data to define metrics also proves to be more than sufficient as not only these three different metrics can be computed, but many others depending on the focus of the analysis.

The main purpose of this comparison is not to merely identify the loans in scope of for the ESRB Recommendation, but to stress the value of granular data in order to capture specific financial phenomena. Different users of credit data may require specific definitions which are costly to implement. This analysis shows that it is worth considering capturing the main elements of the financial intermediation and combining them in different ways to match the economic phenomena that are needed for analysis. On the other hand, this shows that an adequate exploitation of AnaCredit data requires a significant level of knowledge of the AnaCredit data model and the attributes and values therein.

4. Exploring indicators from the top down

The value of AnaCredit for macroprudential analysis does not only stem from the enhancement to track existing indicators or define new indicators, but also from exploring the economic phenomena highlighted by these indicators in more detail.

In practice, this inductive approach involves several steps. First, indicators are selected at highly aggregated levels. After identifying indicators of interest, the analyst may decide to delve into a particular indicator for one specific country for which the indicator shows particularly high potential risk. Focusing on one country, the analyst can look further at, for example, the distribution of the indicator across the population of credit institutions of the particular country at both group and individual level. The definition of the distribution allows the analyst to focus on the counterparties which concentrate the highest risk revealed by the indicator. The analyst may also abstract from the less risky counterparties and aim attention at the counterparties that pose the highest systemic risk.

A single indicator may only be a part of the puzzle and the user may need to quantify multiple indicators in combination. AnaCredit individual and granular information enables the user to calculate the matrix of correlations among different indicators which may allow the identification of relationships among different sources of risk that may otherwise go unobserved.

The identification of intra-group links on the basis of ownership relationships and inter-group connections due to inter-MFI transactions registered in AnaCredit allows the identification of networks. The identification and measurement of these

connections enables the users to quantify how the identified risks may spread in the financial sector. This information may be of significant relevance for stress test exercises which can be enriched with contagion effects.

This represents only one way of how indicators can be explored from the top down with AnaCredit data.
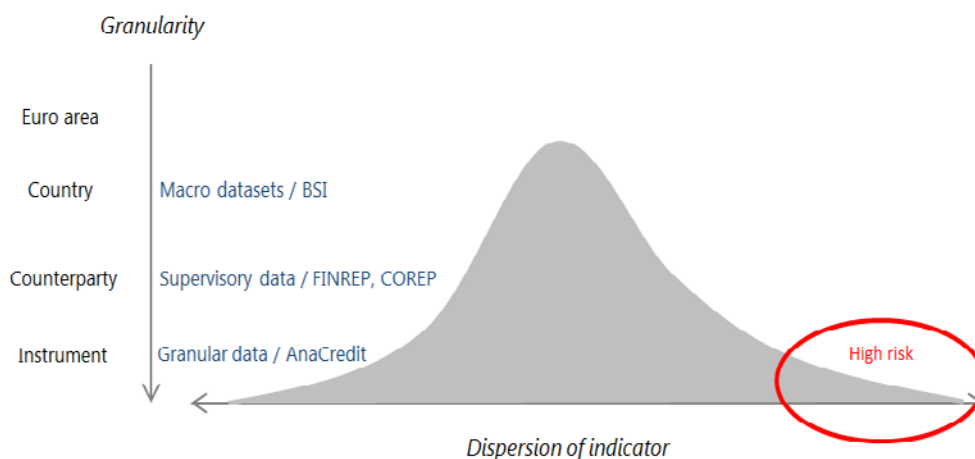

Figure 3. Granular datasets such as AnaCredit provide a zoom lens that can reveal specific instruments that generate the highest risk for the system as a whole. These sources of risk are often invisible when data is only available on an aggregate level.

More concretely, the analysis could focus, for example, on the indicator loan-to-value ratio. Depending on the purposes of the analysis, the definitions for the concepts of loan and value might be different. Using AnaCredit data, the amount of the loan may be calculated with the attribute 'outstanding nominal amount' when the focus lies on financial intermediation, or the 'carrying amount' when the interest lies on the credit risk for which the credit institution is currently exposed to. Similarly, the underlying value of the protection may be calculated in different ways depending on the target of the analysis, whether the assessment of the credit institution represented in the 'protection allocated value', or the maximum value of the protection that could be used to secure the instrument, which can be derived on the basis of the 'protection value' and the set of instruments secured by the protection.

Furthermore, the loan-to-value may be restricted to specific types of protections, e.g. real estate collateral, or take a more holistic view. Another relevant dimension that matters for the definition of loan-to-value ratios is the moment in time.

Also, if the analysis lies on the decision process of economic agents, it is of relevance to compute these indicators at the time when the decision was taken, i.e. at origination. However, if the analysis focuses on the status and evolution of the financial intermediation and the risks involved, then the current and historical values will be needed. AnaCredit allows both types of analysis as, in addition for the values to be reported for each reference date, AnaCredit also includes the main metrics at origination.

This breadth of options demonstrates the flexibility of using AnaCredit to conduct multiple types of analyses as it is evident from the use of different attributes to properly define the indicators that best represent the object of analysis. AnaCredit empowers the user providing the necessary data for her to take the final decision on what should be the adequate metric to be calculated.

Given a specified definition of the loan-to-value ratio, the AnaCredit user can decide in a flexible manner on the focus of the analysis. It might be sensible to focus on specific groups of entities, e.g. banking/corporate groups or SMEs. Alternatively, analysis may focus on the location of the creditor, the debtor or the real estate collateral securing a loan. Another perspective within the realm of AnaCredit would be to focus on the performance of instruments, e.g. to look at loan-to-value ratios specifically for non-performing instruments or counterparties. The benefit of the flexibility of the granular database is precisely that the user does not necessarily need pre-defined assumptions, but can instead let the data speak for itself given the various available dimensions and possibilities for combinations.

In the particular example of the loan-to-value ratio that the analyst might wish to explore, investigating the distribution of the indicator would be highly informative for financial stability purposes. AnaCredit allows the analyst to look at the distribution of the loan-to-value ratio across nearly the entire population. This reveals the degree of dispersion as well as those entities or portfolios with the highest loan-to-value ratios. More advanced analysis could involve the inclusion of more indicators and looking at the joint distributions. Valuable information may also be contained in correlations of loan-to-value ratios with, for example, non-performing loans.

Overall, drilling down from aggregate indicators at Euro Area or country level down to the level of the instrument has many benefits. It is the inductive, data-driven approach that will allow users to make maximum use of the information value of AnaCredit. This perspective underlines the potential of AnaCredit not only to provide macroprudential analysis with more detailed information, but also to allow analysts to see risks that would otherwise remain hidden.

5. Future enhancements

The AnaCredit Regulation sets up the first stage of AnaCredit and contains reflections on possible future enhancements in the recitals. Although each new reporting requirement would be subject to a dedicated merits and costs exercise according to standard practice within the ESCB, the Regulation explicitly mentions the following envisaged extensions:

- a reassessment of national discretion with respect to granting derogations for small reporting agents (Recital 11)

- an extension of the reporting population to non-deposit-taking institutions and other financial corporations in later stages to deposit-taking

corporations other than credit institutions, asset management vehicles and other financial corporations (Recital 12)

- an extension of the instruments to be reported to derivatives, other accounts receivable, off-balance-sheet items (such as financial guarantees) and credit extended to persons other than legal persons, including to sole proprietors (Recital 12)

- the requirement to report on a consolidated basis (Recital 12).

For our purposes, we will focus on four particular possible extensions. Firstly, the interoperability between AnaCredit and other databases could be enhanced. The connection of AnaCredit with other granular databases such as the Securities Holdings Statistics Database (SHSDB) and the Centralised Securities Database (CSDB) would significantly enhance the information value for macroprudential analysis. Moreover, connecting AnaCredit to other databases with data on the individual level would allow for an even broader perspective. This could include the iBSI, as well as FINREP and COREP data.

Secondly, AnaCredit could be extended by adding attributes on - or facilitating a connection to – complete balance sheet data of legal entities. This would allow inter alia a comprehensive view of the risks taken by banks. Analysis of credit intermediation would be enhanced by data on the sources of finance and leverage of the lenders. Assessing the total indebtedness of borrowers would furthermore require including data on off-balance sheet items such as loan commitments and guarantees as well as derivatives.

Thirdly, Regulation (EC) 2533/98 requires that the ESCB uses confidential statistical information, including AnaCredit, exclusively for the exercise of the tasks of the ESCB. The granular data will therefore not be accessible for the general public except for scientific research bodies which may be granted access provided the identification of counterparties is not possible. However, market participants may adopt the AnaCredit data model to collect granular credit and credit risk data and establish market-led initiatives for the sharing of granular data for other purposes.

Finally, enlarging the scope of AnaCredit would further enhance the flexibility of the database for financial stability purposes. In particular, there is a need for macroprudential policy to focus increasingly on shadow banking, i.e. banking activity conducted by non-banks (Constâncio, 2017). As demonstrated in the case study, other relevant counterparties include, for example, natural persons. This group represents a significant part of banks' portfolios. Additionally, the analysis of banks' exposures to households via deposits and mortgages is an important source of systemic risk that could be analysed in case natural persons are included in the scope of AnaCredit in the future.

6. Conclusion

Granular data captures the behaviour of economic agents closer to reality than traditional aggregate datasets. By modelling individual elements of the economy and explicitly taking into account the interaction between agents these novel databases provide new insights that can be used to enhance current analysis and to define new indicators for financial stability purposes and beyond.
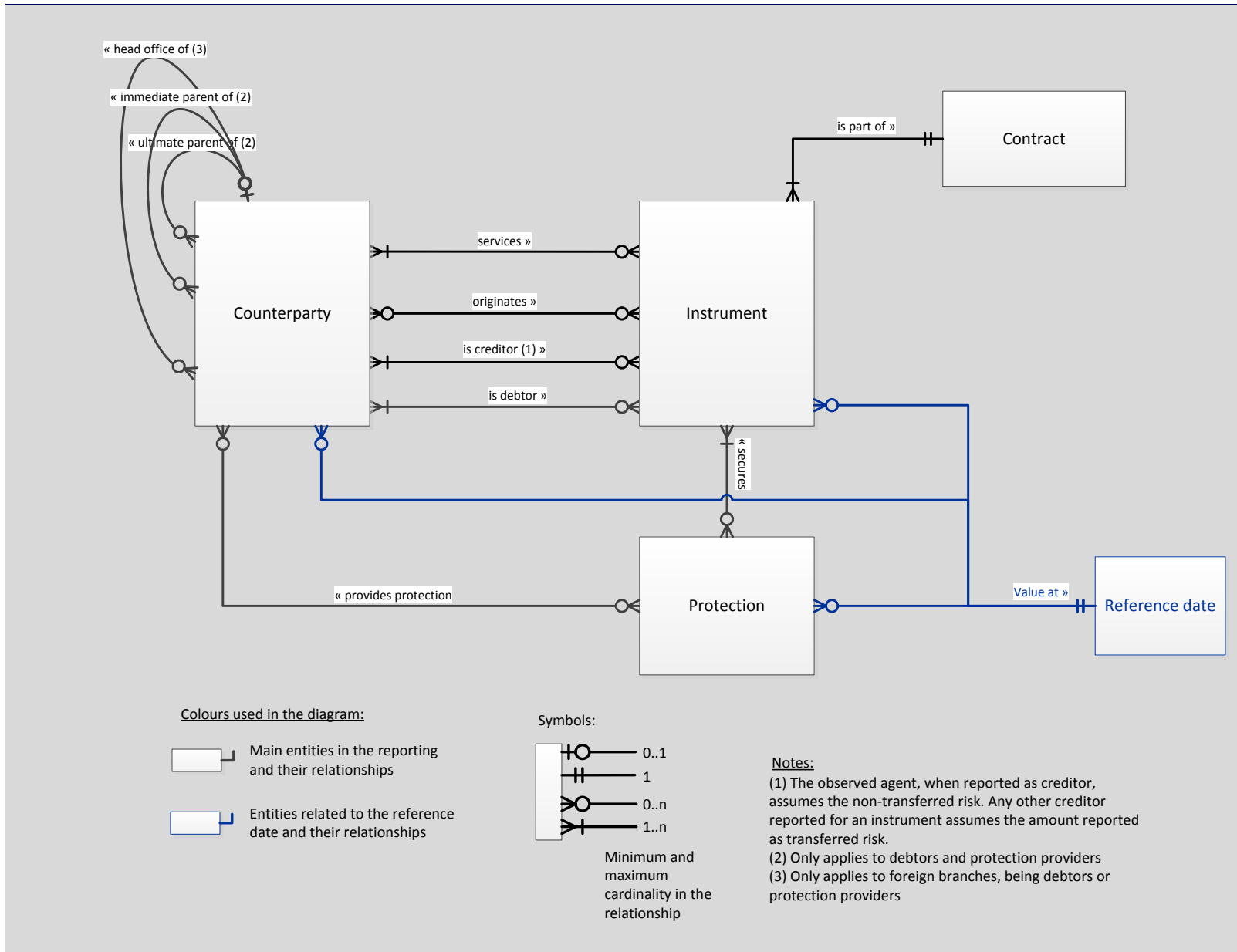
AnaCredit is an example of a granular database for credit intermediation, modelling credit on the basis of its main elements: instruments, counterparties and protection. The user of AnaCredit data can identify connections between these elements and conduct analyses at various levels, from the instrument and counterparty levels up to the aggregate for the Euro Area as a whole (bottom-up). In addition, high-level indicators previously available only as average figures can be analysed from the top-down without pre-defined assumptions.

Hence, macroprudential analysis can benefit from AnaCredit data in three ways. Firstly, it provides a higher level of detail to analyse credit intermediation at various levels. Secondly, networks of counterparties can be traced and constructed based on various dimensions. Thirdly, full distributions of indicators across nearly the whole population can be viewed and analysed.

As the case study of real estate data gaps has shown, granular data alleviates to a large extent the difficulty of having different definitions in different datasets. Indicators on loans, for example, might use varying definitions of a loan itself. The granularity of AnaCredit helps to overcome different definitions as they can be reconstructed by combining certain data attributes.

Ultimately, many indicators for financial stability purposes can be constructed with AnaCredit data. Not only will decision-makers have a more sophisticated view of credit intermediation, but also the effectiveness of policy tools might be enhanced as they can be used in a more targeted manner.

Annex 1: AnaCredit conceptual data model

Colours used in the diagram:

- Main entities in the reporting and their relationships
- Entities related to the reference date and their relationships

Symbols:

- 0..1
- 1
- 0..n
- 1..n

Minimum and maximum cardinality in the relationship

Notes:
(1) The observed agent, when reported as creditor, assumes the non-transferred risk. Any other creditor reported for an instrument assumes the amount reported as transferred risk.
(2) Only applies to debtors and protection providers
(3) Only applies to foreign branches, being debtors or protection providers

Bibliography

Constâncio, V. (2017, May 11). Macroprudential policy in a changing financial system (speech). Frankfurt am Main, Germany.

European Central Bank. (2016). Regulation (EU) 2016/ 867 of the European Central Bank - of 18 May 2016 - on the collection of granular credit and credit risk data (ECB/ 2016/ 13). *Official Journal of the European Union*.

European Central Bank. (2017). *AnaCredit Reporting Manual. Part I - General Methodology*. Frankfurt: European Central Bank.

European Central Bank. (2017). *AnaCredit Reporting Manual. Part II - Datasets and data attributes.* Frankfurt: European Central Bank.

European Systemic Risk Board. (2016). Recommendation of the European Systemic Risk Board of 31 October 2016 on closing real estate data gaps (ESRB/2016/14). *Official Journal of the European Union*.

Israël, J.-M., Damia, V., Bonci, R., & Watfe, G. (2017). The Analytical Credit Dataset - A magnifying glass for analysing credit. *ECB Occasional Paper Series*(187), 1-41.

Neudorfer, P. (2016). Multiple usage of the ESCB 'Register of Institutions and Affiliates Database' (RIAD) - features and challenges. *25th Meeting of the Wiesbaden Group on Business Registers*, (pp. 1-10). Tokyo.

Thijs, S., & Corvoisier, S. (2017). Integrating reference data for monetary policy and supervisory purposes - The European System of Central Banks (ESCB) experience. *IFC Bulletin. 43*, pp. 1-12. Basel: Bank for International Settlements.

IFC-National Bank of Belgium Workshop on "*Data needs and Statistics compilation for macroprudential analysis*"

Brussels, Belgium, 18-19 May 2017

AnaCredit: From broad to flexible macroprudential analysis [1]

Orestes Collazo Brananova and Gibran Watfe,

European Central Bank

---

[1] This presentation was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

EUROPEAN CENTRAL BANK

EUROSYSTEM

Orestes Collazo
European Central Bank

Gibran Watfe
European Central Bank

AnaCredit:
From broad to flexible
macroprudential analysis

IFC-National Bank of Belgium Workshop

Brussels, 19 May 2017

Disclaimer: The views expressed are those of the authors and do not necessarily reflect those of the ECB.

Contents

Past

Broad approach

- Aggregate indicators

- Constrained by *static definitions* and *complex relationships* between agents

Future

Top-down and bottom-up approaches with three advantages:

1. **Detail**: flexible definition of *indicators* (with back-casting) and calibration and *estimation of models*

2. **Networks**: *concentration of risk* and *contagion effects*

3. **Distributions**: rich analysis of the complete population with simulations

Financial intermediation is complex

- **Instruments**   Multipurpose contracts
- **Counterparties**   Creditor, debtor, protection provider, etc.
- **Protection**   Collaterals and guarantees
- **Relationships**   Banking groups / corporate groups

    Syndicated loans / joint debtors

    Umbrella contracts

    n-to-m instrument-collateral relationships

AnaCredit Regulation (EU) 2016/867 adopted *18 May 2016*

AnaCredit go-live: *September 2018*

- Reporting population: credit institutions

    - Resident in the euro area

    - Includes all foreign branches

- Counterparties:

    - Creditors: credit institutions & other sectors (for loans serviced by CI)

    - Debtors: legal entities (including Government)

- Instruments (assets): loans (including inter MFI positions)

    - Serviced or held

- Threshold (creditor-debtor): EUR 25,000

88 attributes to assess credit intermediation

- Counterparties: ➡️ *Who is who?*
  - Identification of creditors and debtors
  - Characterisation: e.g. size, sector of economic activity

- Balance sheet status ➡️ *What?*
  - Classify exposures by type (e.g. type of product) and use (e.g. securitisation)
  - Needed for internal consistency: avoid double counting (e.g. joint liabilities)

- Exposure features ➡️ *What (in greater detail)?*
  - Classify the exposures for analytical purposes (e.g. maturity, interest rate)

- Risk measure ➡️ *Will it happen?*
  - Provide a forward-looking view (e.g. probability of default)

- Loss measure ➡️ *What would happen?*
  - Provide a backward-looking view (e.g. accumulated impairments)

- Valuation ➡️ *How much?*
  - Book values, nominal values, market values

➢ Four indicative use cases

+ Examples for AnaCredit attributes that may be relevant

Concentration of risk

- Head office/immediate/ultimate parent identifiers
- Address: country
- Economic activity
- Outstanding nominal amount

Contagion

- Outstanding nominal amount
- Type of securitisation
- Type of protection
- Protection value

Level of exposure

- Enterprise size
- Outstanding nominal amount
- Off-balance sheet amount
- Probability of default

Losses

- Legal final maturity date
- Settlement date
- Arrears for the instrument
- Accumulated impairment amount

- **Banking groups**
  - Pseudo-consolidation
  - Different levels of banking groups (e.g. national level, EU level)
  - Syndicated loans

- **Corporate groups**

- **Location**

- **Sectors of activity**
  - Institutional sector
  - NACE level 3

- **Contagion**
  - Inter MFI positions
  - Unique identification of counterparties
  - Banks with similar business models

- **Extensibility via RIAD***

* RIAD: Register of Institutions and Affiliates Database
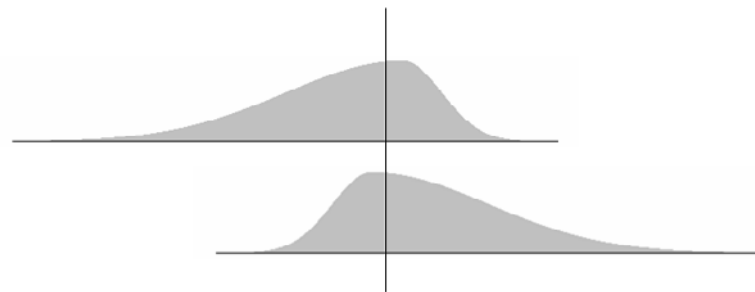
AnaCredit distributions

- Individual information

- Near complete population

➡ Building the full distribution for all entities

- Describing the complete population

  - Average (traditional)

  - Additional descriptive statistics for the population

    + Central tendency

    + Dispersion

    + Concentrations

➡ e.g. focus on the tail
(entities that accumulate risk)

- Combining indicators

  - Loan-to-value - NPL ratio - bank size

  - Correlations

- Based on ESRB Recommendation (ESRB/2016/14)

- Two main loan portfolios: CRE* and RRE** loans

  - Different definitions of CRE and RRE

  - Combination of granular data allows for multiple definitions

  - Large reconciliation possible with current information

- Coverage:

  - CRE loans are captured (missing loans to natural persons for RRE)

  - Large coverage of main variables required to calculate the indicators

  - Many of the proposed breakdowns

  - All required distributions, indicator metrics and data types (flow/stock)

- AnaCredit allows analysing in detail the portfolio of CRE loans

  - Large set of indicators

  - Correlation with other indicators

  - Ex-post definition of indicators – not constrained to a fixed template

* CRE: Commercial real estate

** RRE: Residential real estate

Case study: ESRB Real Estate Indicators - CRE

ESRB indicator concepts	AnaCredit mapping
Lending – current	Outstanding nominal amount
Lending – at origination	Commitment amount at inception, off-balance sheet amount (requires estimate for credit lines)
Property value – current	Protection value, type of collateral
Property value – at origination	Original protection value, type of collateral
Maturity – at origination	Settlement date, legal final maturity date
Investment in CRE	Out of scope
Non-performing loans	Performing status of the instrument
Loan-loss provisions	Accumulated impairment

ESRB break downs	AnaCredit mapping
Property type	Type of protection (less detail in AnaCredit)
Property location	Real estate collateral location (Analyst must identify prime locations)
Lender type	Economic activity
Lender nationality	Address: country
Investor type	Economic activity
Lender nationality	Address: country
Property under development	Purpose with the value Construction investment

- Loan-to-value

- Different definitions can be specified

  - What is a loan? (outstanding amount, carrying amount?)

  - What is value? (protection value, amount of protection that secures the loan?)

    + By type of protection: Financial / real estate / other

  - What time? (current, origination?)

- Defined at loan level –
  Focus of analysis can be defined ex-post based on available details:

  - Groups of entities

  - Location (creditor, debtor, real estate)

  - Performance

- Distribution

  - Focus on the least secured (highest loan-to-value) entities/portfolios

    + Tail of the distribution

  - Joint distribution

    + Correlation with other relevant indicators, e.g. non-performing loans
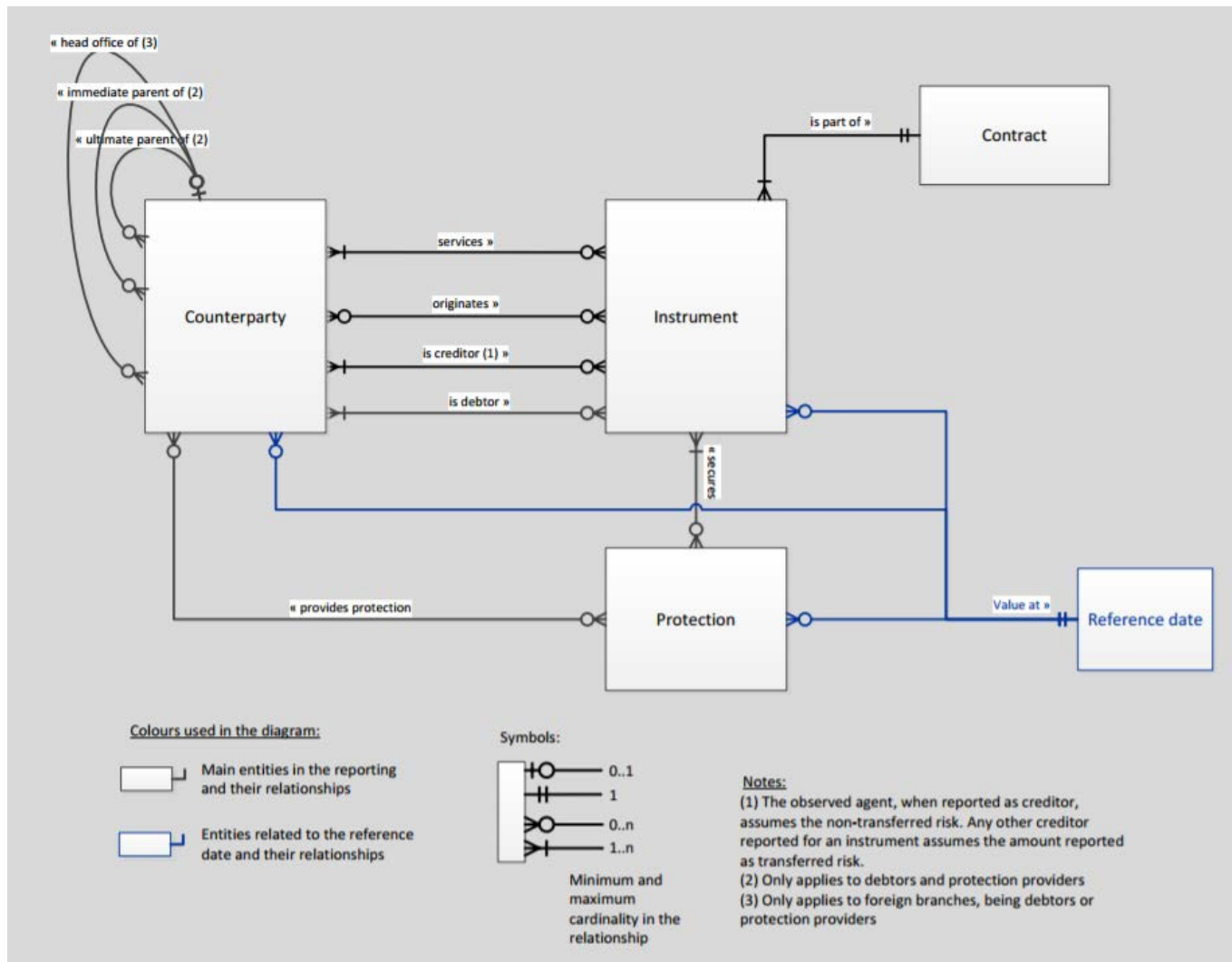
Future extensions

- Combine with information from other datasets
  - Granular data (SHSDB, CSDB)
  - Individual basis (e.g. iBSI, FINREP, COREP)

- Complete balance sheet*
  - Comprehensive view of risks taken by the bank
    + Off-balance sheet items, e.g. loan commitments, guarantees provided
    + Derivatives
  - Sources of finance and leverage

- Enlarge the scope*
  - Shadow banking
  - Other relevant counterparties, e.g. natural persons

* Not planned, based on user needs

- **Granular data** captures the behaviour of economic agents closer to reality

- **Macroprudential analysis** can benefit in particular from

  – The higher level of *detail*

  – The ability to trace and construct *networks of counterparties*

  – The possibility to analyse full *distributions* of indicators

- Granular data alleviates discrepancy between different **definitions**

- Many economic & financial **indicators** can be constructed with AnaCredit

  – **Drill-downs and distributions** provide important *insights for analysis* where averages are insufficient and dispersion measures are needed

Annexes

Conceptual model

- 6 entity tables with one or more datasets each

Instrument entity table
instrument data, financial data, accounting data

Counterparty reference data entity table
counterparty reference data

Protection entity table
protection received data

Counterparty risk/default entity table
counterparty risk data, counterparty default data

Instrument-protection received entity table
instrument-protection received data

Counterparty-instrument entity table
counterparty-instrument data, joint liabilities data