



” AMRON – a quality-driven database

Karol Kacprzak

Project Manager AMRON III,

Specialist for Analysis and Development of the AMRON System

AMRON CENTRE ANALYSIS

March 5, 2026





As far as databases used for real estate market analysis are concerned, not only the number of records collected is crucial, but above all the reliability and consistency of data counts the most. From the very beginning, AMRON System has adopted the principle that the value of a database is built on data quality rather than data volume. For this reason, the development of the system and the processes related to data collection are designed in such a way as to minimise the risk of errors and inconsistencies to the greatest possible extent.

From the very beginning of the System's existence, it was assumed that, as the administrator, we take full responsibility for the quality of the data stored in the database, regardless of how it is sourced. This means that every particular record entered into the System is subject to multi-stage quality control. Data is not entered into the database automatically and without oversight – it is verified both at the moment of entry and at subsequent stages of its functioning within the System.

FIRST: VALIDATIONS

Already at the data entry stage, an extensive validation mechanism is used to eliminate errors. In the new version of the AMRON System, this mechanism has been significantly expanded and divided into three levels.

The first is preliminary validation, performed during the upload of a batch data file into the System. At this stage, the correctness of the file structure and its format are checked. The System verifies, among other things, the consistency of the field layout and the permitted number of records in a single file. The purpose of this validation is to detect technical errors before the actual data processing begins and to ensure the stability and performance of the entire solution. This validation applies only to batch uploads – in the case of individual data entry via the web interface or via API, it is not performed.

The second level consists of first-level validations, which cover all records entered into the System – whether entered individually by a user in a browser, via API or in batches. At this stage, the System checks the basic correctness of the data. It verifies, among others, whether fields are mandatory, whether data types are correct, the permitted number of characters and the formats of individual pieces of information (e.g. masks for specific attributes). In the case of records entered in batches or via API, the correctness of dictionary values is additionally checked – the System verifies, whether a given value belongs to the set of permitted values defined in the System dictionaries.

The most advanced level of control consists of second-level validations. Their purpose is to identify situations, in which the data formally meets all the basic requirements but may indicate a potential substantive error. This applies, for example, to cases, when information in different fields is mutually contradictory or when values significantly deviate from typical parameters for a given type of property. In such situations, the user entering the data must confirm its correctness. Some of these validations may also result in the record being referred for additional approval by a privileged user (i.e. the Central System Administrator). This level also includes mechanisms comparing a new record with records already existing in the database, which makes it possible to detect and eliminate potential duplicates.

SECOND: QUALITY REVIEWS

However, data quality control does not end at the stage of System validations. Each record may be subject to a quality review process and, if any doubts arise, a so-called verification request is created. This is a formal process aimed at checking the correctness of the data and – if necessary – correcting it. In addition, the banks en-



tering data into the System also carry out periodic quality verifications of the data they have entered. As a result, the database is continuously monitored and improved, and potential inaccuracies are systematically eliminated.

THIRD: STANDARDISATION

The scope of mandatory data in the database has been designed in accordance with Recommendation J, which ensures information consistency and its usefulness for real estate market analyses. In the new version of the System – AMRON III, a number of solutions supporting users in the data entry process have also been introduced. One of them is integration with external registers and databases. For example, the “cadastral district” attribute is a dictionary consistent with GUGiK data and after an address is selected, the list of available cadastral districts is automatically narrowed to the selected municipality. The building number entered is verified on the basis of the Address Points Database and postal codes are matched to the indicated location. If a land and mortgage register number is provided, the System automatically indicates the competent land and mortgage register court. Another feature is that records may be automatically supplemented with information (i.e. regarding year of construction, number of storeys in the building, building structure, building density, transport accessibility, surroundings, information on energy efficiency and others) collected in the Buildings Database – a proprietary database of information on buildings in Poland maintained by the AMRON Centre. The automatic retrieval of data from the Buildings Database reduces human error and improves the reliability of records; moreover, users do not have to manually search for and complete many building parameters. The System performs this automatically, which significantly shortens the time needed to enter a record. Additionally, exchange rates may be retrieved automatically depending on the selected currency and transaction date.

The System also uses the official Statistics Poland (GUS) TERYT register. During data entry, the TERYT code is completed automatically on the basis of the TERC, SIMC and ULIC registers, so the user does not have to manually enter the full address path.

ACQUISITION OF DATA FROM RCN AND THEIR QUALITY

Recently, the issue of acquiring data from Real Estate Price Registers (RCN) has also gained particular importance. Following the introduction of new forms of access to this data, it has become possible to download large information packages in GML format, covering transactions from many counties. In one of many tests of this solution, we downloaded data initially covering more than 170 counties. At first glance, such packages might appear to be an ideal data source enabling a rapid increase in the scale of the database. In practice, however, the analysis of the quality of this information proved to be crucial. After converting sample GML files into tabular form (accepted by the AMRON System), it turned out that many records did not contain basic information – for example, the address number of a unit. In addition, the GML files also lacked information on land designation, which is available in data obtained directly from RCN through the traditional route.

Even greater differences became apparent during a detailed analysis of data quality. In one county, the package contained more than 28 thousand transactions. After applying basic quality filters, such as completeness of the ownership share in the property, minimum land area, realistic price per square metre or exclusion of non-standard sources of information, the number of records that could be used dropped to around 5 thousand. This means that nearly 80% of the data was rejected due to insufficient quality or incompleteness.

This example clearly shows why AMRON consistently applies the principle of “quality over quantity”. Although it



would be technically possible to quickly increase the number of records in the database through the automatic import of large data packages, in practice this would mean introducing a significant amount of incomplete or questionable information. Instead, data selection procedures are applied that make it possible to retain in the database only those records that meet specific quality standards.

Therefore, in the case of RCN data, in many situations the traditional route of obtaining information is still used, even though it is more time-consuming and often subject to additional limitations, for example regarding the number of records that may be downloaded one-off. However, it makes it possible to obtain more complete data, including full address information and additional attributes important from the point of view of market analysis.

SUMMARY

All the mechanisms described above – from multi-level validations, through integration with public registers, to verification processes – have one common objective: to ensure the highest possible quality of data in the AMRON database. As a result, users of the System can base their analyses on reliable, consistent and thoroughly verified information. Our goal at AMRON is to create a database based on reliable data – a larger number of records matters only when they are of high quality.



KAROL KACPRZAK

Project Manager AMRON III, Specialist for Analysis and Development of the AMRON System

e-mail: karol.kacprzak@amron.pl

AMRON Centre

AMRON Centre is a professional analytical and research center, specializing in issues relating to the real estate market analysis and monitoring. Since over 20 years we provide our clients and partners with reliable information on real estate market, we explain phenomena recorded on that market and we assure the access to dependable economic information. Our expertise and experience covers assessing the influence of transaction prices changes on Polish housing market perspectives – as the only market researches provider, we analyze changes on housing loans market, due to access to the data gathered in System for Analysis of Real Estate Financing Market (SARFIN System). Unlike publications of other housing market consulting companies, our reports express the actual picture of the market, based on transaction prices. We guarantee comprehensive approach to the real estate market issues, including the complex analyses of micro- and macroeconomic factors and long-term perspective grounded on historic data, current market prices and prognoses.

LEGAL NOTE

AMRON Centre states that the Report was prepared with due diligence in purpose to assure the accuracy of the presented information. Nevertheless, AMRON Centre reserves that the Report shall be of a general nature and may not be concerned as advisory service or any kind of service. AMRON Centre shall not bear responsibility for any consequences of using the information included in the Report, in particular any consequences of decisions or actions undertaken or abandoned on the basis of included information.

Contents of the Reports is legally protected due to the regulations of Act on copyrights and other related rights and use of it entirely or in part requires a disclosure on presented data source.

AMRON Centre acts in the name of and on behalf of Polish Banks Association, owner of the AMRON System.