Statistical data dissemination in the Czech National Bank

Petr Vojtisek[1] and Martin Kacer[2]

1. Introduction

Dissemination represents the culmination of the entire process of statistical work, which begins with analysing users' requirements, and includes creating the methodology for different statistics, designing statements, determining the reporting population and collecting the actual information, combined with a compilation process that involves dealing with missing information and conducting grossing-up procedures. The quality of dissemination can either strengthen or weaken the process – as well as the ultimate results – of compiling the various statistics. Thus, the Czech National Bank (CNB) devotes significant effort to this aspect of the statistical process, in which both methodological and technical issues are important. The present paper discusses types of data dissemination at the CNB, and describes the Bank's aggregated time series system.

2. Types of dissemination

At the CNB, we distinguish three basic types of dissemination, which have distinct objectives: internal dissemination, external publication and reporting to international institutions. Access to primary data for CNB's data processing personnel is not considered dissemination in this context.

Firstly, statistics, once compiled, are available to analysts within the CNB, mainly in the form of a time series database.

Secondly, there is external publication, which is done exclusively through the CNB website. The bulk of statistical data are published in this external format, through a time series database. Various publications are also available in pdf format at the statistical site, including the actual figures, as well as graphs and brief comments. Publications include "Monetary and Banking Statistics" and "Balance of Payments Developments," as well as SDDS (Special Data Dissemination Standards) publications, which are provided in a standardised format – currently in the form of a simple (but structured) html table.

Thirdly, reporting to international institutions takes the form of data coded in GESMES/TS format and sent regularly to the ECB, Eurostat and the BIS, at frequencies ranging from daily to annual. Reporting to other international institutions, such as the IMF and the OECD, is mostly in MS Excel format.

All three types of dissemination are integrated at the CNB, within a system known as ARAD (an acronym from the Czech phrase for "aggregated time series"). ARAD provides a data

---

[1]   Czech National Bank, Monetary and Statistics Department, Na Prikope 28, 115 03 Prague, Czech Republic, petr.vojtisek@cnb.cz

[2]   Czech National Bank, Monetary and Statistics Department, Na Prikope 28, 115 03 Prague, Czech Republic, martin.kacer@cnb.cz

storage structure, organises workflow and allocates responsibilities at each step of the processing. Last but not least, it serves as the end-user application for data extraction.

This paper discusses the ways in which statistical data are disseminated, with a special focus on the ARAD system, which represents the core of the entire process.

3. ARAD system

ARAD is a system for storing different statistical data in one central database and presenting them as time series. The first version of this system was created in 1996 in an environment completely different from today's. At that time, data were not available through the internet, and were published only in the form of print bulletins. The most advanced format was MS Excel. We began building the system with some basic, simple goals:

- To provide unified storage of statistical data in the form of time series, using a central database

- To provide interactive access to information, including both data and methodology

Since then, the environment has changed and new requirements have emerged. Two challenges were considered paramount:

1. To establish a unified basis for both publishing and reporting data

2. To create a common data interface for different systems

Simplicity was the dominant idea in the nineties, when the first version of the system was created, and the only way we found of meeting the challenges was to make the process clear and natural, avoiding any special coding or structures. The present system is somewhat more sophisticated, but simplicity remains its basic feature.

3.1 The main concept of ARAD – general principles

The basic idea is to store complete blocks of data (data sources), rather than predefined data series. Blocks of data are organised according to various business areas. Sources are described in the form of metadata, and the relevant methodological sheets are provided. Administrators in each business area are responsible for the data source structure and methodology, and for the actual data. Data sources are merely the first level of the system, and their sole purpose is to serve as a source. They are not normally visible to users. The only exception is access by internal users, who can create their own datasets, and therefore need to have access to the complete structure of the database. Creating datasets is an advanced use of ARAD and enables users to create fully customised views of the statistical data.

The second level, available to all users, consists of the time series of individual indicators, which employ ARAD terminology. Technically, an indicator is the formula which refers to the position of the data in the data source. The formula allows for elementary calculations such as addition or subtraction. Indicators are arranged in groups, ie sets of indicators. Indicators provide a simplified means of viewing the data in the data source.

The ARAD system primarily serves users, and is therefore based on their needs. Since requirements change over time, the database is an open one, and can be added to. The only requirement is that a well-structured order be maintained. A two-level structure allows, to a certain extent, for independently adding data sources and predefined sets of indicators. In particular, new data sources can be prepared without affecting existing data presentations or relationships, and vice versa.

For data input, the system provides a common interface for fully automated uploading of data from different primary systems, external sources and compilation processes. It was found useful as well, however, to support manual data entry not constrained by a strict format. Thus, for manual entry, ARAD is the only place where data are stored in the form of a database. To make manual inputting easier, the application is able to read and process pre-prepared MS Excel sheets.

3.2 User access

All of the tools for the presentation of the time series are based on hierarchical selection – from a general level down to specific details. Navigation is considered the third level of the system. The administrators have prepared sets of indicators for standard ARAD users and have placed them on the respective branches of the hierarchical tree. Such users of ARAD do not, therefore, need detailed knowledge of all of the data sources, and navigation is organised intuitively.

Currently, there are six branches at the first level of navigation in the internet version, reflecting various statistical areas: monetary and banking statistics, balance of payments, financial markets, macroeconomic statistics, other indicators and SDDS. One additional branch – government financial statistics – will be ready soon.

The money and banking statistics branch provides a wide range of indicators concerning monetary financial institutions, from balance sheets (eg monetary and credit aggregates) to interest rates for clients. It also covers information from selected financial sub-sectors (eg collective investment funds and financial corporations engaged in lending).

The balance of payments branch includes standard balance of payments and international investment position items, as well as indicators of foreign indebtedness. Figures on international reserves are also available in the IMF template format.

Financial market statistics provide information on the money, capital and foreign exchange markets, including PRIBOR, long-term interest rates, foreign exchange rates (including nominal and real effective exchange rate indicators) and Prague Stock Exchange indices.

All of the above-mentioned statistics are compiled by the CNB. By mutual agreement, ARAD also contains the information produced by the Czech Statistical Office, consisting of four types: national accounts (GDP), production statistics, prices (both consumer and producer) and labour and wage statistics.

More advanced internal users can create their own sets of indicators (in the internal version only) and store them in the central database. The system's ability to handle methodological changes and data updates is seen as the great advantage of the database format, as compared to maintaining one's own dataset indicators in MS Excel sheets.

3.3 Data access tools

There are currently three versions of user access tools: local, intranet and internet versions. The initial local version is still available, primarily as an application for manual data entry. Even in this case, data are stored only in the central database. The same approach applies to other versions in use.

The CNB intranet version contains, of course, more detailed and confidential data than the public internet version. The intranet version has tools to enable access to confidential or temporarily restricted data. The internet version is accessible at the CNB website via standard internet technologies.

All versions provide standard tools for data gathering, such as filtering by time range, full-text searching and explicit selection of indicators. There are various output formats designed to

meet different user needs. In addition to an html table providing a simple display of the results of a selection in a web browser, MS Excel is available – this being the most popular format for further calculation and analysis – while CSV (comma-separated values), which is also available, serves as the all-purpose format for further data processing. In addition, time series can be displayed in charts for a more complex view of the data history.

As regards metadata and methodology, the application interactively provides additional information on the relevant data, in the form of methodological sheets, news and frequently asked questions. Current data updates are provided directly on the main page. Concurrently, information is distributed through the RSS (Really Simple Syndication) channel, which provides users with convenient monitoring of new publications.

4. Data dissemination through ARAD

As mentioned above, ARAD is the core of the CNB's dissemination system. However, there is another component, namely, the timetable that controls the entire dissemination process. In very simple terms, the timetable contains what must be sent or published – and when – on any given day, plus additional information such as which procedure should be performed and to whom the notification should be sent.

The system maintains two types of time schedule systems. The first is used for managing the time series updates based on information from the data sources; the second is used to orchestrate publication and reporting – with these two levels being independent. While the data update schedule deals with data sources, the publication and reporting schedule concerns the presentation layer (predefined sets of indicators) and must be addressed separately.

A description of the publication day, as it appears on the CNB website, may be the clearest explanation of the publication layer of the system.

Before publication can occur, the timetable must be met, and data must be prepared in the source database. The first step of the process consists of loading data from the source system to the ARAD database. This ordinarily takes place one day before the publication date. The publication process itself usually begins at 8:30 a.m., when the automatic procedure begins to create the html table based on information from the timetable. Fifteen minutes later, the file is transmitted to the Internet Publication System used for managing our website content. At 9:00 a.m., the file is transferred to a protected folder, where the data are accessible only to registered users such as press agencies and journalists. The information, of course, is under embargo until 10:00 a.m. Registered users are concurrently notified of publication by e-mail. At 10:00 a.m., the document is moved to the target folder and becomes available to the public.

The entire process explained here is automated and is controlled by the information from the timetables, with one subsystem passing data to another according to the instructions outlined. At any stage, it is, of course, possible to enter the process and perform the procedures manually. This data publication structure has been successfully used for almost five years.

5. Conclusions

The advantages of our solution have been proved by nine years of failure-free ARAD operation. Naturally, every system has its weak points, but we believe that the advantages of this system outweigh its disadvantages. The fact that all data are reported and published from one main source database is considered a fundamental feature of the system. A further

advantage is that the same data – along with metadata and methodology – are interactively accessible to users.

When we began building the user interface in 1997, the main task was to introduce as simple an application as possible, so that users would not be discouraged from using it. The simplicity began, at one point, to be a limitation, but fortunately users have increased their skills, while technology developments have allowed us to add more sophisticated functions and procedures. When we began doing this, an important decision was taken to instruct and train users to be independent. Instead of preparing personalised outputs for one or several users, we offered users standardised datasets (based on their requirements) and instructed (and sometimes compelled) them to obtain the required data in the most effective way – an accomplishment which, with almost ten years of experience behind us – we now see as one of the greatest successes of the work we have carried out.

Reference

Link to ARAD: http://www.cnb.cz/docs/ARADY/HTML_new/index_en.htm