

MonFiDS: An Integrated Financial Database

Rasika Amarasiri¹, Philip Gharghori²

¹Department of Accounting and Finance, Monash University, Australia, Rasika.Amarasiri@monash.edu

²Department of Accounting and Finance, Monash University, Australia, Philip.Gharghori@monash.edu

INTRODUCTION

A vast amount of research in accounting and finance requires access to stock market and accounting data of listed companies. These data are obtained by subscribing to various commercial databases, which provide access to different bits of data. Researchers often have to merge these data by matching the company tickers or other identifiers, which in some cases requires a significant amount of effort and time. In certain cases, the matching process is significantly hindered by the unavailability of a common reference and requires a significant amount of manual matching.

As there are a number of researchers working on the same datasets, this increases duplication of these efforts and accumulates a significant waste of valuable resources and time, which could have been spent on more productive work. The Monash Financial Database System (MonFiDS) was developed with the objective of minimising these efforts and to allow researchers to concentrate on the actual research by taking off the pressure of integrating the different datasets and checking the integrity of the merger.

MonFiDS currently integrates financial, accounting, market index and earnings estimate data for Australian listed companies from five different data sources. The database allows researchers to download this data selectively or as a whole via a web portal. Common manipulations in the merging process are included in the extraction process that minimise the requirement for post-processing of the data.

DATA SOURCES

MonFiDS currently merges data from five different data sources. The main company database that is used as the backbone of the database is sourced from the Share Price and Price Relatives Database (SPPR) [1]. This database also provides the monthly average stock price data and other valuable company information. The daily stock prices and price-relatives are sourced from the SIRCA Core Research Data (CRD) dataset [2]. Accounting and other company financial data is sourced from the Morningstar Equity Data Licensing (formerly known as Aspect Huntley Data Feed) product [3]. Earnings estimates and recommendation data are sourced from the Thomson Reuters Institutional Brokers' Estimate System (I/B/E/S) dataset [4]. In addition to these, the daily market index data are sourced from the Iress terminal [5].

The SPPR database's company information was selected as the back-bone company table on MonFiDS as it is the only known database that can be used to track company name changes, mergers and acquisitions through the recorded history of the Australian Securities Exchange (ASX). It has built its own group company code that allows one to track all incarnations of a company through its listing on the ASX. The database also allows the tracking of the previous and next instance of the company via a linked list type of pointers.

The hardest dataset to be matched is the I/B/E/S dataset, as the dataset uses its own code for companies, which is a bit hard to match directly. In most cases the matching needs to be done via the company name, which is again heavily abbreviated to fit in its limited field length. Currently, the company table of I/B/E/S is only matched to about 92% with the SPPR company table. The remaining companies (around 32 companies) cannot even be matched manually, as there are multiple matches in the SPPR to the abbreviated company names in I/B/E/S.

Due to the timing of the availability of the different datasets, the update of MonFiDS is done annually. The update is usually carried out around April/May each year as the SPPR dataset is updated around this time.

TECHNOLOGIES USED

MonFiDS has been designed as an online database to allow wider access to the researchers. Using the popular LAMP framework of Linux, Apache, MySQL and PHP, the entire database is built upon free/open source resources. The scripts used for updating the database from the source data use a combination of Java and Perl scripts running on both Windows and Linux platforms.

One of the challenges faced during the production of the database was the long running time of some extractions. Due to the complex nature of some of the extractions, the run time can vary from a couple of minutes to over an hour. The default PHP installations do not allow such long running scripts, as they time out after about 30-60 seconds. It is also awkward for the user to be staring at a blank web page, which will keep on loading for more than a minute. In order to improve the user experience, the database produces a submit confirmation web page and releases the browser while sending the actual data generation process to the background. Once the data is ready, the user is notified via an email that

the data is ready for download. The links provided on the email are valid for 48 hours from the time of completion. This time limit was introduced to reduce the compilation of large extraction copies on the server.

Due to license restrictions on the data sources, the access to the database has to be regulated to valid users and accounted. This was achieved by providing a user login using the university's central authentication system via LDAP. This login process also allowed the capture of the user's basic information such as their name and email address automatically without manually adding each user to the system.

LICENSING ISSUES

Due to the heavy competition between various data vendors, the default licensing agreements when subscribing to financial and accounting databases are very restrictive. They specifically do not allow the re-creation of other databases utilising the data provided in their products. This was one of the main barriers that was faced during the initial stages of the design of MonFiDS. Lengthy discussions and negotiations along with additional guarantees of usage tracking and security of the data had to be put on the table before getting the approval of the data vendors to put their data together in MonFiDS.

The main difficulty in organizing such special allowances is because the original license agreements are mostly devised towards commercial clients and do not take into consideration academic use. In most cases, the original license agreement goes through several rounds of amendments between the university solicitors and the vendor's solicitors before they get signed. These negotiations lengthen the subscription process by a significant amount of time.

ADVANCED FEATURES

MonFiDS database currently caters for a research community of about 100 users, which includes researchers from both finance and accounting backgrounds. Due to the nature and theory in the different research fields, flexibility in the matching process had to be allowed. These included the lagging of price data by different time periods from the accounting and forecast data. To accommodate this, the database provides the researcher the ability to lag the data by a number of months.

In previous versions of Excel, the number of rows was limited to around 64,000. The database has an output format that allows the creation of a set of CSV files that limit each file to be 64,000 rows maximum. It currently has output formats of CSV (both split and full), tab delimited and HTML. Larger extracts are automatically compressed to reduce the download size.

Certain research projects require the extraction of data for only a specific group of companies. The database allows extraction of data for a list of companies by manually creating this list. Since repeated creation of such a list is tedious, the database allows the saving of these company lists in their profile. Commonly used data field combinations are also allowed to be saved in the profile and standard combinations are provided as global templates.

FUTURE PLANS

MonFiDS has been a valuable asset to researchers in the department working on research projects on Australian listed companies. It has been widely used and has existed for over 10 years now. With emerging needs, several updates to the database are planned in the near future. These include addition of new data sources such as the company announcements (Signal G) data and corporate governance data, a feature to upload company lists instead of manually creating them and the addition of new output formats such as SAS, STATA and XML data formats.

REFERENCES

1. *Share Price and Price Relatives (SPPR) database*. Product description available from: http://data.sirca.org.au/downloads/SPPR_Items_2010.doc, accessed 29 June 2011.
2. *SIRCA Daily Data (CRD) database*. Product description available from: <http://data.sirca.org.au/uni/services/Data/equities/asxdaily.html>, accessed 29 June 2011.
3. *Morningstar Equity Data Licensing*. Product description available from: http://corporate.morningstar.com.au/Equity_Data_Licensing_0901.pdf, accessed 29 June 2011.
4. *Thomson Reuters Institutional Brokers' Estimate System (I/B/E/S)*. Product description available from: http://thomsonreuters.com/products_services/financial/financial_products/a-z/ibes/, accessed 29 June 2011.
5. *Iress Equity Information Terminal*. Product description available from: http://www.iress.com.au/assets/FactSheets/IRESS_FactSheet.pdf, accessed 29 June 2011.

ABOUT THE AUTHORS

Rasika Amarasiri: Rasika Amarasiri obtained his PhD from the Faculty of Information Technology in Monash University. He completed his Bachelor of Science in Engineering (honours) and the Master of Science degrees from the University of Moratuwa, Sri Lanka. Prior to his PhD, Rasika has held various positions in the industry and academia including being the Hostmaster for the .LK domain registry, web master, systems administrator, consultant and lecturer. He has had numerous publications accepted in journals and conferences including a best paper award. Currently, he is the Manager, Research IT Services at the Department of Accounting and Finance in Monash University, Australia.

Philip Gharghori: Philip Gharghori is a senior lecturer in Finance at the Department of Accounting and Finance at Monash University. Prior to this, he was a PhD student and tutor in the department. His PhD was supported by both an Australian Postgraduate Award and a Departmental Scholarship. He is the author of the finance chapters of Principles of Accounting and Finance, a text used in undergraduate business and commerce courses. He has won awards for conference papers presented at the Australasian Finance and Banking Conference and at the AFAANZ Conference. Phil's research is primarily in the area of asset pricing. In particular, he has focused on the performance of multifactor models and stock market anomalies. He also has written papers in the areas of funds management, default risk analysis and the market reaction to corporate actions.