

## Data Aggregation for Asset Reconciliation

### Introduction

This paper focuses on data consolidation for asset reconciliation within hedge funds and long only asset managers. The process requires aggregation of data from a number of external sources such as administrators, prime brokers, custodians and counterparties as well as from internal accounting or order management systems (e.g. Beauchamp FundManager, Simcorp Dimension etc.), and posting the resultant data to the reconciliation system.

It is common even for small hedge funds to have multiple prime brokers, and, since the collapse of Lehman Brothers, many hedge funds are seeking to spread their risk across a greater number of trading partners. Similarly asset managers often have to reconcile data from a number of different custodians which are not always chosen by themselves but by the clients whose funds they are managing.

The problem of dealing with multiple sources of data is often tackled by the implementation of SWIFT or by outsourcing the problem to a data aggregation service. Both are viable approaches but come at a cost, require additional IT infrastructure or may not support the required set of providers. With the increasing availability of timely data provided via secure ftp, or the use of custodian web portals, in-house data aggregation is an increasingly attractive and cost effective option as long as the right tools are available.

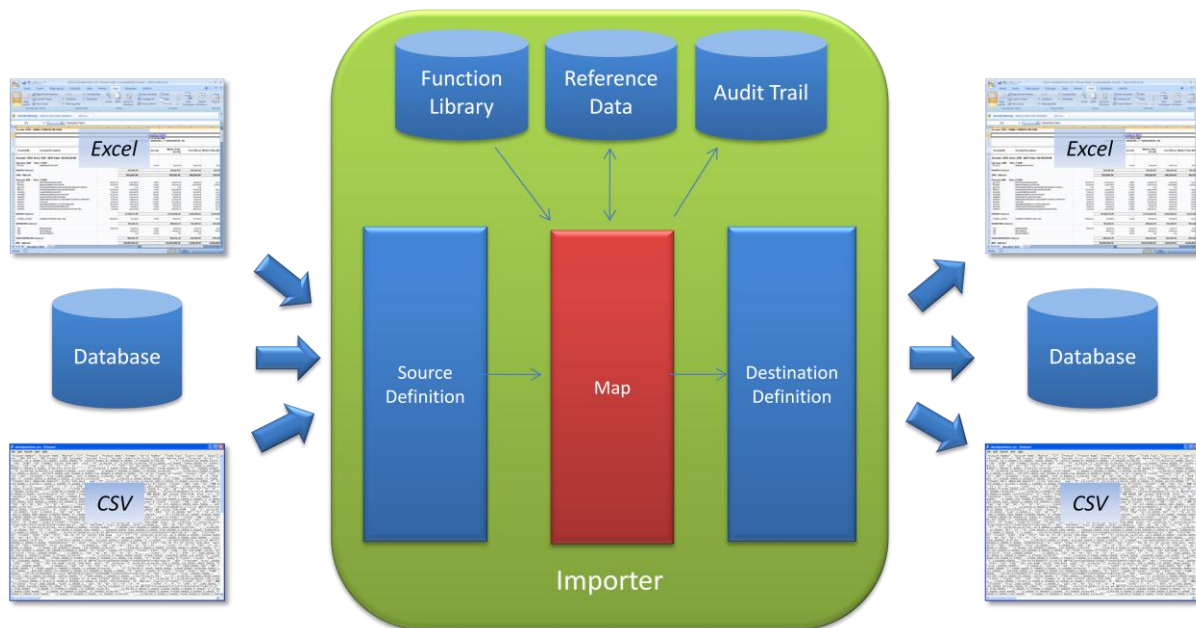
This paper discusses this alternative route and draws on experience of delivering internal data aggregation tools and services for a number of firms over several years.

For simplicity the reports or data provided by prime brokers, custodians, administrators and internal portfolio accounting systems are referred to in this document as 'data suppliers'.

### Key Areas in Data Consolidation

The following topics are discussed: -

1. Dealing with Multiple Data Sources (Delivery and Automation)
2. Differing Data Structures
3. Adjusting Quickly to Changes in File Structure
4. Data Transformation Functions
5. Security Identifier Recognition
6. Data Destination
7. Auditing and Accountability



**Figure 1: Example of Data Import and Export Process**

### 1. Dealing with Multiple Data Sources (Delivery and Automation)

Data delivery, like the structure of the files, varies greatly between different data suppliers. At the most basic end of the spectrum files are emailed, unencrypted, to users on a daily basis. However the most usual methods of delivery are either directly via secure FTP or through a web gateway such as BNY Inform or JPM Access. The main requirement in all methods is that the data arrives in the same format each time. Even a seemingly insignificant intervention, such as resaving a data file in Excel, can lead to distortion of the data structure and therefore possible failure of the parsing process.

In most small firms we have encountered there is no automation of the delivery process and initiation of the data translation stage is purely user-driven; clearly it is desirable to automate this process so when files arrive by whatever means they automatically trigger the translation process. Many data providers send their data in the early hours of the morning, and with an automated process, data translation will have occurred and when users log on to the destination system the consolidated data is available and ready to use.

We believe a variety of methods are useful to users depending on the level of control required. This ranges from automatic processing once the data is dropped onto an FTP site, to scheduled processing of data on various timescales or simply leaving the choice of when to trigger the process up to the user. Decryption of data should also be factored into the import process.

Any type of process must also take into account issues such as late delivery of data files, which although rare can cause a number of knock-on problems. Bank holidays and other non-trading days

can also cause issues if the automated process is not setup correctly to deal with the various holidays in different parts of the world.

In addition to the considerations required for accepting multiple data sources, addressing the volume of data is also extremely important. Many suppliers provide multiple files (for example positions, trade, cash adjustments, NAVs, prices etc) on a daily basis and these may all need to be imported for the reconciliation process. In addition to this we see many suppliers providing individual files per portfolio which will multiply the number of files received on a daily basis. The large volume of data demonstrates the need for an extremely fast and highly automated data consolidation solution.

## 2. Differing Data Structures

We have found that the quality of data from suppliers varies from well organised 200+ column trade report files to unformatted text files containing very basic data and omitting some required fields. Generally hedge funds and small asset managers rely on the standard offerings by their data suppliers which in most cases contain all the relevant data required, however a minority of suppliers are able to modify the data on request (for example by providing additional security identifiers) however this usually involves a long lead-time.

We find that data suppliers typically use Microsoft Excel format, .csv files or fixed format .txt files when delivering reconciliation data. There are some exceptions such as XML and other delivery types which appear to be an amalgamation of any of the above! In the case of internal systems (such as a portfolio accounting system) the data might not be available in report format without the use of additional API software, in which case the import mechanism needs to be able to draw data in directly from tables in the underlying database.

While the data provided from different custodians/prime brokers is usually different in structure there are a number of suppliers using the same systems to produce this data. SunGard GMI is used by a number of suppliers, and while the reports sometimes look different on initial inspection the logic used to report trades, balances etc is the same. This enables portions of import logic to be used across a number of different suppliers and can reduce the time spent writing import procedures for new data suppliers' connections.

When attempting to interpret supplied data it is preferable that a file specification is provided, as some aspects of the file interpretation such as cancellations and corrections or data continuity can be difficult to understand quickly. In our experience these specifications can be extremely helpful in understanding the files and speed up the entire interpretation process, although only a minority of custodians or prime brokers seem able to provide them at present.

Any system or process needs to be able to handle a number of different file types and structures in a consistent and transparent manner. Much of the data provided is a 'standard offering' from the suppliers but sometimes organisations such as hedge funds are able to negotiate a degree of control over the files provided by their chosen prime brokers.

### 3. Adjusting Quickly to Changes in File Structure

It is inevitable that the data offering by custodians and prime brokers for example will change over time, usually because the supplier wishes to enhance the quality of information provided or they are responding to other clients' requirements. Also, some suppliers use a degree of manual intervention with both the production of the files and their content which can make obtaining consistent data more difficult.

Even a small change in structure such as the addition of a new data column or using a signed net consideration on trade cash flows could render the use of a macro or hard coded data translation method useless. To make matters worse the majority of data suppliers don't inform the users of the data before making minor changes as they generally assume the data analysis (or reconciliation in this case) is performed manually by spreadsheet.

The solution is either an easy to use data translation tool that can be adjusted to cope with issues that may crop up due to data structure changes or to have a responsive IT Support function to fix any complicated code. We believe a translation tool should be flexible enough to handle small changes in the data supplied without the need to alter code or book development time with IT support professionals. Hard coded solutions provided by third parties can work out to be extremely costly in these situations.

Another method is to use an external partner such as a data aggregation company; however they can only be as responsive as their systems allow and this may cost the client per change.

### 4. Data Transformation Functions

Ideally the data provided by suppliers would be clean and each required field could be inserted into the destination 'as-is'. However, in practice this is rarely the case and consequently powerful data transformation functions are required to convert disparate source data into a standardised set of records. For example, if signed trade quantity is required and isn't provided in the source file the translation tool needs to be able to combine a debit/credit indicator with unsigned quantity to produce a signed figure. Other common examples of transformation functionality include: -

- Price calculation from quantity, gross consideration and charges
- Mapping transaction types from source to destination
- Account/Fund/Portfolio mapping from the suppliers references to the internally used account numbers
- Security Identifier recognition to enable mapping of Sedol, Bloomberg codes, ISINS etc to internal security ids

Aggregate functions are also required to collapse data from multiple to single records. In some instances this is simple e.g. sum rows to calculate a balance. In others this can be more complex and a range of aggregate functions may be required covering average, first, last, count etc.

The scope of the transformation tool must therefore match and exceed the demands places upon it by the wide range of data shapes that are presented for translation. Over several years of processing asset management data, we now have this in place.

## 5. Security Identifier Recognition

Along with basic trade information (dates, quantity, net consideration etc) there is a need to recognise and assign trades, positions and dividends with unique security id of the target system. Receiving reliable security identifiers in the source data is critical if this is to be achieved. The data translation method also needs to incorporate a lookup table which can be added to easily or automatically as new securities are set up.

We have found that most data suppliers can provide at least one type of security identifier, with some providing a choice of two or more. In the example of long-only reconciliation we find that SEDOL is the most common identifier provided, followed by CUSIP and ISIN. For the OTC market, the counterparty usually assigns its own internal identifiers or delivers the market identifier for the underlying security.

Security recognition must support multiple serial lookups. For example SEDOL might be looked up first and, if this was not found or was not provided, a secondary lookup using the CUSIP might follow and so on until a match were found. This logic must be supported by the translation tool.

With listed futures and options, identifier recognition can be more complex and we find that the preference is for using Bloomberg codes such as the BB Unique identifier or the Options Clearing Corporation (OCC) symbology. Due to the volume of derivative trades in some hedge funds it becomes extremely important to obtain a good 'hit-rate' of finding the correct internal id from the identifier when trying to automate data reconciliation.

## 6. Data Destination

The ultimate destination of the data is considered throughout the process but this also needs to have some flexibility to deal with changes in process and data structure.

We often find that a direct database link from the consolidation tool to the destination system is more preferable than exporting the aggregated data to an output file which has then to be re-imported elsewhere. While relatively rare, a database schema change can pose exactly the same problems for this output as it does for the import side of the process. Again the need for flexibility in the output format and structure is required, it should effectively mirror the import side of the process.

It is considered vital that the destination system should be able to validate the data received; this can be anything from simple checks on the date and time of the data to full validation. Using a reconciliation system as an example we are able to validate daily trade and position data by taking the previous closing position, adding the daily trades and checking if this equals the current closing position. This logic is extensible to validating cash balances and margin balances and these checks can quickly identify problems with the source data or import process.

## 7. Auditing and Accountability

The data consolidation process has a number of complex stages which can prove confusing when trying to trace a specific item of data back through the process to its source.

Archiving of the supplied source data needs to be considered as many suppliers merely provide an undated file on a daily basis. Time-stamping imported data and archiving this in a secure area organises these files, enabling the users to refer back to the correct data if required and it reduces the need for any user interaction at this stage in the process.

Any translation or manipulation of data also needs to be traceable to determine whether data errors are caused by incorrect logic in the consolidation process or are simply the result of unexpected logic or data in the source files. It is preferable to have an easy to use 'step-through' process which logs the changes to each line of data that is being imported and whether any aggregations (such as totalling balances) is occurring.

In addition to being able to audit the consolidated data any changes in import procedure also need to be archived. For example, a change to import logic on a certain date makes it necessary to store a copy of the older logic in order to trace back any issues that may have occurred.

### **Our Experience in this area**

Watson Wheatley Financial Systems uses a proprietary data consolidation tool known as Importer to transform data from various sources and place it in the i-Recs Reconciliation system.

At present Watson Wheatley has connections to a large number of prime brokers, custodians, administrators and counterparties as well as portfolio accounting systems such as Beauchamp FundManager and Simcorp Dimension.

In total we have written nearly 1000 data import and export maps, enabling our clients to consolidate incoming data from multiple sources on a daily basis and then export data to upstream systems after reconciliation.

*Watson Wheatley is a software and consultancy company with over 20 years experience of building and delivering products and custom solutions to the asset management industry. If you wish to comment on this paper or request further information please visit [www.watsonwheatley.com](http://www.watsonwheatley.com) or email [info@watsonwheatley.com](mailto:info@watsonwheatley.com)*