

Choosing the Best Account Aggregation Solution

by Robert Huebscher

Once you embark on a plan to implement a data aggregation solution, the initial questions you must answer are how the data will be gathered, what technology will be employed, and how much are you willing to pay for better quality data. A variety of data aggregation techniques can be used, and the right choice will depend on your requirements along dimensions such as cost, scalability, and coverage. This paper examines the different approaches to account aggregation, how they differ along these and other dimensions, and how you can measure data quality to insure that a given solution will meet your needs. The focus of this paper is on wealth management applications and meeting the needs of high and ultra-high net worth investors.

For more background on account aggregation, its definition and use in various applications, please refer to the companion paper “What is Account Aggregation?”

Different Approaches to Account Aggregation

In its simplest form, account aggregation consists of collecting position, balance, and transaction data from paper statements. In fact, until about ten years ago, this was the accepted method for creating a consolidated statement for investors. The drawbacks to this approach are obvious: it is costly in terms of manual labor, it is error prone, and it does not scale (it becomes challenging and unwieldy as the volume of data increases). A major disadvantage is the time delay involved in gathering the data. In a typical cycle, the investor gets the paper statement at the end of a month or a quarter, mails the statement to their advisor, who in turn does the manual entry, creates a consolidated statement, and provides the result back to the investor. Often, data is 30 days old before the investor sees it, and this can stretch to 90 days if the statements are only quarterly.

Despite these disadvantages, manual aggregation from statements is still a common practice in the marketplace. Several firms offer such services on an outsourced basis, charging as much as \$600/account/year for the data entry. A key advantage to paper statements is that 100% coverage is possible – every financial institution can and typically does provide paper statements to their clients.

Beginning in the late 1990s, it became common for institutions to make position, balance, and transaction data available on-line, via web sites, fostering an immediate improvement in the manual account aggregation approach. Financial advisors were now able to access their clients' information on-line and re-enter the data to create consolidated statements. It was no longer necessary to wait for clients to send paper statements, provided that clients were willing to set up on-line access to their accounts and provide their advisors with the credentials (account number, PIN or password) necessary to access the data on-line. Advisors were able to surmount the timeliness issues inherent in manual aggregation via paper statements, but the solution was error-prone, and neither scalable nor cost efficient.

Direct feeds have been the traditional mechanism for aggregating assets from established institutions (global custodians, money center and regional banks, and larger brokerage firms). These institutions provide a nightly feed, on behalf of investors, to RIAs and family offices that manage those investments. The investor must authorize the institution to provide the feed to the RIA, and security of the transmission is typically achieved through password-protected access. The institution sends the RIA a file for those accounts that are authorized. Direct feeds are supported by major portfolio accounting vendors, such as Advent and Schwab/Performance Technology. Direct feeds offer enormous advantages over manual (paper or web site) approaches, in that they are secure, scalable, timely and cost-efficient. Advent supports over 400 interfaces to various institutions through its Advent Custodial Data (ACD) service. Direct feeds are also the norm for institutional investors, who use vendors such as Evare and Electra as centralized hubs for data collection and redistribution. Where a high number of accounts are concentrated within a small group of institutions, direct feeds are usually the best solution.

The disadvantage to direct feeds is the very limited universe of institutions that are supported by vendors, and the time required to establish new interfaces. Although Advent's coverage of 400 institutions is impressive, it is only a small fraction of the 14,000+ commercial banks, 6,000+ broker/dealers, and 5,000+ wealth managers that currently serve the U.S. market (not to mention those serving non-US markets). Adding coverage for a new institution can be a 6-12 month process for many vendors. For an RIA that must be responsive to the needs of their clients, striving to achieve 100% coverage of their clients' assets, direct feeds alone are inadequate. Direct feeds typically capture about 60% of the assets for high net worth and ultra-high net worth investors.

Electronic aggregation via web sites has emerged as the preferred solution for capturing the 40% of assets that are not available via direct feeds. Virtually every institution makes investor information (positions, balances, and transactions) available on-line. ByAllAccounts emerged as the leader in using on-line account aggregation in wealth management applications, specializing in delivering 'reconciliation-ready' data to be used within a portfolio management platform. Advent and Schwab users are able to combine data obtained via direct feeds with ByAllAccounts' feeds to achieve nearly 100% coverage for their clients' assets. Electronic web-based access is scalable, secure, and cost-

efficient for HNW and UHNW investors, particularly in situations where account volumes are spread across a relatively large number of institutions.

On-line aggregation is not the appropriate solution for all applications. Web sites change, both in their navigation and their content, and in some cases this requires engineering in order for on-line aggregation to continue to work properly. Direct feeds are more stable in this respect. Therefore, in cases where reliability is an absolute necessity, direct feeds offer advantages.

The primary methods for account aggregation are compared in Figure 1.

	Coverage (number of institutions supported)	Scalability (ability to handle large account volumes in a cost-efficient manner)	Timeliness (ability to provide data soon after it is reported by the institution)	Security (how the confidentiality of the information and the privacy of the investor is safeguarded)	Reliability/Accuracy (the degree to which results that match the custodian can be produced on an ongoing basis)	Cost per Account	Best suited for situations where
Manual aggregation from paper statements	High	Low	Very poor	Relies on mail or Fed-X	Very poor	Very High	Paper statements are the only way to access account data
Manual aggregation from web sites	High	Low	Poor	Relies on credentialed access with human intervention	Poor	Very High	Advisors do not have access to electronic aggregation
Direct feeds	Low	High	Good	Relies on credentialed access to read-only information	High	Low	High volumes of accounts in a limited number of institutions, or requirements that demand high reliability
Electronic aggregation from web sites	High	High	Good	Relies on electronic credentialed access to web sites	Very Good	Low	Direct feeds are not available, or where accounts are distributed across a large number of institutions

Measuring the Quality of the Data

The table above illustrates many of the key differentiators among account aggregation methods, but there is one criterion that deserves a much closer examination: data quality. Data quality means two things: data accuracy – how well the positions, balances, and transactions match what was provided by the custodian; and whether the data is ‘reconciliation-ready’ - so as to encounter as few errors as possible when imported into a portfolio management platform. We will look at these two measures of data quality – accuracy and reconciliation-readiness, see how this is measured, and what factors affect data quality.

Accuracy is not absolute, in the sense that no aggregation solution can offer data that exactly matches the activity in an account. Aggregation technologies can only achieve the standard of replicating what is provided by the custodian. Custodians make errors, and it is the responsibility of the reconciliation process within the portfolio management application to identify these errors and allow them to be resolved. Examples of errors coming from a custodian include failing to report a corporate action, updating positions but not transactions, incorrectly identifying a security, providing incorrect amounts, or mis-posting transactions.

Reconciliation-readiness goes beyond accuracy. To be reconciliation-ready, the data must be presented in a uniform and consistent manner, so that it can be processed by the portfolio management application in a general way. In other words, the portfolio management system should not have any engineering for 'special cases' to deal with the idiosyncrasies of how different custodians might report their data. To achieve reconciliation-readiness, aggregation technology should offer the following:

- **Translation.** Consistent translation of transactions and securities from multiple institutions to a single accounting system format: All transaction types should be identified and presented in a consistent manner, with enough granularity to permit processing by the portfolio accounting system. For example, 'buy,' 'purchase,' 'invest,' and 'sweep' should be classified as the same transaction type, while preserving the original classification, in case it is needed. Securities must be properly and consistently identified by industry-standard symbols (in the US, either CUSIP or ticker). When a custodian reports only a CUSIP or a ticker, the aggregator should provide the missing identifier. More importantly, when the institution reports neither CUSIP nor ticker, the aggregator must, if possible, use the name provided to determine the appropriate CUSIP and ticker. To provide full functionality an aggregator must maintain a security master file with a complete universe of CUSIPs, tickers, and security names.
- **Problem Resolution.** Altering users to problems in the data. When a transaction or security is identified that has not been previously defined, or if the transaction or security is undefined, it must be presented to the user in a convenient format, that facilitates the research and resolution process. This process should extend to allow the data to be corrected and re-submitted through reconciliation engine.
- **Customization.** Enabling users to customize how data is translated for the accounting system to comply with that firm's designated guidelines. For example, transaction types must be mapped, certain transactions (e.g., unmanaged assets and corporate actions) must be filtered out, fees must be properly identified, and missing data (such as a price per unit) not supplied by the custodian must be computed.

Manual aggregation technologies, whether via paper statements or from web sites, can achieve high levels of accuracy and reconciliation-readiness, but at extremely high per-account costs. Data must be checked and double-checked manually, by different individuals, to insure accuracy. Care must be taken to insure daily consistency in data entry and customization, especially when different individuals are involved. Computer software behaves consistently; people, unfortunately, do not.

Wealth managers using direct feeds typically encounter errors in roughly 1-5% of the positions to be reconciled. Variations are due to the level of transaction activity, the types of securities and transactions encountered, and the number of institutions utilized. Electronic web site aggregation produces a slightly higher rate of errors, and variability is due to the same set of factors. For electronic aggregation technologies, whether via direct feed or via web sites, accuracy and reconciliation-readiness are dependent on the data source and quality of the technology. Ultimately, with any electronic solution, manual input is still required to resolve problems due to inaccuracies in reporting or to problems in the aggregation solution.

The Format Factor

Direct feeds come in formats that are unique to the institution providing the feed. No standardization exists, and institutions define formats using fixed column, XML, CSV, and other techniques. No correlation exists between the format and the accuracy or reconciliation-readiness of the data; that is determined by the institution producing the data and by the sophistication of the aggregation technology.

Electronic aggregation via web sites relies on two general approaches: data gathering via HTML, often referred to as 'screen scraping,' and data gathering via fixed format. Both approaches require credentialed access (the use of an account number plus a PIN or password) to the web site, and depend on navigating through the site to obtain the desired information. With HTML, aggregators are tapping into the programmatic representation of the data that is being viewed by the user. Less sophisticated aggregation technology looks for data in certain positions on the screen; better and more modern approaches rely on column headings and labels, and dynamically adapt to format changes on the part of the institution. However, data gathering via HTML is becoming less common, with aggregators such as ByAllAccounts now relying on pure HTML for less than 30% of the sites they support. Supplanting HTML access has been the growth of fixed format feeds provided by institutions, specifically to support the needs of account aggregators and downloads to other applications. Fixed formats resemble direct feeds, but actually adhere to more standardization. Most banks provide data in an OFX format (to support personal finance applications). Other institutions use fixed column, XML, CSV, and other formats.

As with direct feeds, there is no correlation between data format and accuracy or reconciliation-readiness (see "Myth versus Reality" number two). Although intuition suggests that a "better" format (such as OFX) should mean "better" data

quality, this is not the case at all. Any level of quality, from poor to excellent, can be - and is - delivered via any format, and data from two sources using an identical format can - and do - differ significantly in quality. Why is this the case? Formatting is simply a technique for physically presenting data elements, and the essence of those elements does not change due to that presentation. An amount and price can be placed in a text file with a comma separating them, in an OFX or XML file with special tags, in a custom feed file in specific column positions, or in an HTML page in table cells. If the financial institution fails to provide an amount or price, they will be missing regardless of format. If the financial institution provides an incorrect amount or price, they will be incorrect regardless of format.

Some financial institutions use different means of assembling data to be delivered in different formats, resulting in disparity of data quality among that institution's data format options. In such cases, there is no way to predict which format will have the best quality data, and an aggregator's reliance on a specific format could limit the aggregator's ability to get the best quality data available.

How to choose the best account aggregation solution

There is no substitute for accurate and reliable data. Regardless of your intended use, and regardless of your intended audience, the data you provide must be accurate. Account information must be updated on a timely basis and it must match what is provided by the institution. Wealth managers catering to a high net worth and ultra-high net worth audience understand this clearly, and leave no room for errors. Financial service providers offering solutions to retail and mass affluent audiences are hearing this message as well, as some providers have seen their support costs skyrocket and their reputations suffer because of inaccurate data. Some have dropped their consolidated reporting offerings, rather than invest in improving the data. It turns out that no data is better than bad data.

Accuracy and reliability are the most important criteria when selecting a solution, but they cannot be considered in isolation. Coverage, scalability, timeliness, and cost-effectiveness all play important roles in the context of a thorough evaluation. At some cost, it is always possible to obtain data that is accurate and reliable, and provides the required coverage, scalability, and timeliness. The question is whether the cost is justified in the greater business model.

Obtaining accurate data that is reconciliation-ready requires a careful evaluation of the technology to be employed, and understanding how this technology offers translation, problem resolution, and customization. Not all wealth managers have the resources to perform a full technical due diligence on their aggregation providers. Instead, it is prudent to rely on the advice of peers and general industry trends. As vendors emerge that purport to offer viable solutions, and these vendors gain acceptance and usage in the wealth management marketplace, then it is usually a safe bet to rely on these vendors.

About the Author

Mr. Huebscher is a consultant serving the wealth management industry. With over 25 years of experience in the financial services market, he advises companies on topics such as new business development, strategic partnerships, market size analysis, competitive issues, and marketing tactics. On a functional level, his expertise extends to account aggregation, investment management and accounting, client reporting, reference and market data, corporate actions, and enterprise data management. He can be reached at: rhuebscher@mba1982.hbs.edu.