



# Our Data Aggregation Advantage

*Simplistic account aggregation can only "cut and paste" information by collecting it and reformatting it to be presented in a consolidated format, sometimes referred to as "Screen Scraping".*

*ByAllAccounts' unique database building process delivers the specific information that is required for investment analysis via a four-step database building and learning process:*

- Step 1: Data Gathering*
- Step 2: Data Normalization*
- Step 3: Semantic Integration*
- Step 4: Business Logic*

*This process allows us to understand transactions and cash flows and create asset and sub-asset classes.*

## Step 1: Data Gathering

ByAllAccounts automatically collects online account information in daily updates, from captive and non-captive accounts at thousands of custodians. The challenge in collecting this information is that it is not available in consistent file formats across institutions.

Once gathered, the data is put through a normalization process to make it as consistent as possible. It is then run through our Semantic Integrator, which interprets the data to “make sense” out of disparate forms of information. Data gathering alone is not sufficient to provide the information required for reconciliation, analysis, and consolidated reporting. In fact, after the initial data gathering process we typically have less than half of the information in the format we require for our analytic purposes.

## Step 2: Normalization

The second phase in our database building process involves taking the raw data that has been gathered using multiple techniques, and normalizing it. This normalization process maps the raw data into ByAllAccounts’ data model.

This data model is based on the model implicit in the OFX data exchange standard, with refinements and extensions to capture and represent the information in such a way that it can be used to perform calculations.

For example, as an exchange standard, OFX defines how the “total” for a buy or sell transaction will be identified, but it imposes no limits on the amount itself: is that total positive or negative? Similar situations exist with balances (OFX defines a means to report balances, but does not require that the actual cash in an account be identified or reported in any specific way) and with transaction identification (OFX defines a way for transactions to be given unique identifiers, but does not require that the financial institution use the same identifier for the transaction each time the transaction is retrieved). OFX scope is also limited to the retail banking and

Raw Data Quality

FORMAT	STD Parsing	STD Value Label	STD Values	STD Activity Type	STD Equity ID	STD Balance ID	STD Trans. ID	Add'l Info	OVERALL RATING
OFX, IFX, OFC	█	█	█	█	█	█	█	█	56%
QIF	█	█	█	▨	█	█	█	█	34%
CSV, TAB, XML	█	█	█	█	█	█	█	▨	19%
HTML	█	█	█	█	█	█	█	▨	6%

Quality After Normalization

FORMAT	STD Parsing	STD Value Label	STD Values	STD Activity Type	STD Equity ID	STD Balance ID	STD Trans. ID	Add'l Info	OVERALL RATING
OFX, IFX, OFC	█	█	█	█	█	█	█	█	63%
QIF	█	█	█	▨	█	█	█	█	41%
CSV, TAB, XML	█	█	█	█	█	█	█	▨	44%
HTML	█	█	█	█	█	█	█	▨	44%

Quality After Semantic Integration

FORMAT	STD Parsing	STD Value Label	STD Values	STD Activity Type	STD Equity ID	STD Balance ID	STD Trans. ID	Add'l Info	OVERALL RATING
OFX, IFX, OFC	█	█	█	█	█	█	█	█	100%
QIF	█	█	█	▨	█	█	█	█	100%
CSV, TAB, XML	█	█	█	█	█	█	█	▨	100%
HTML	█	█	█	█	█	█	█	▨	100%

retail asset management arenas. It makes no provision for more sophisticated financial issues such as position cost basis, tax lot identification, or the division between principal and income portions of a trust.

Normalization takes elements of the raw data and assigns them to appropriate places in the data model. This process largely “standardizes” the non-standard formats, while also beginning to address some of the issues inherent in the standard formats. When this process is complete, ByAllAccounts has about two-thirds of what it needs if the information came from an OFX source, and less than

half of what it needs from other sources (e.g., QIF, CSV, HTML).

### A Data Challenge Example

How many ways can you say “buy”?

- Automatic Investment
- Automatic Purchase
- BOT
- Bought
- Buy
- Contribution
- Direct Investment
- Direct Purchase
- EECON
- Invested
- Investment
- Money Fund Sweep
- New Account Investment
- Purch Thru Dealer
- Purchase
- Purchased
- Shares Purchased
- You Bought @1

...and that’s just one of 25 transaction classes. The process however, is not simply one of building a dictionary of terms.

### Step 3: Semantic Integration

The data-building process for most account aggregation services ends after collection and little or no normalization. The data available at the end of this process is consolidated and fed into a simple user interface or delivered “as is” to users. This data is insufficient for planning by advisors and fueling investment tools and services. ByAllAccounts’ Semantic Integration process uses proprietary, patent-pending technology to make sense of the data retrieved and to fill in any gaps in the information. At the center is a “smart engine” that has been collecting, analyzing and interpreting data for over ten years.

The Semantic Integration engine uses a series of heuristics and algorithms to evaluate the data, respond to inconsistencies and gaps, and combine that data with other sources of information (third-party data sources, previously retrieved information, and user-supplied/manual input) to produce a complete data picture. This is a process that “learns” over time and “is adapted” to the idiosyncrasies of new financial institution data sources as they are brought on board. It is not a process that can be quickly or easily duplicated.

Our Semantic Integration process is essential if data is to be truly understood and analyzed. Retrieving data in richer, more consistent formats addresses part of the challenge, but is not a complete answer and is far from universally available. Normalization improves on the less-standard formats, but falls well short of the mark. Only Semantic Integration closes the gap.

Semantic integration addresses:

- Interpretation of free-form data
  - To what instrument does the data refer?
  - What kind of activity does it represent?
  - What are the relevant values?
- Reconciliation
  - Is this “real” account information or an error?
  - Have we already seen this same information?
  - How does this relate to everything else?
- Adjustments
  - What are the implications for other positions?
  - What are the implications for cash flows?
- Instrumented to detect incorrect decisions

Contribution Type	Transaction Type	Activity Type	Amount	Price	# of Units or Shares
Esop Leveraged Shares	Dividends/Capital Gains	Cash Receipts	\$17.80	\$0.000	0.000
Esop Leveraged Shares	Dividends/Capital Gains	Purchase	(\$17.80)	\$36.063	0.494
Employee Deferral	Contributions	Cash Disbursement	(\$332.06)	\$1.000	0.000
Employee Deferral	Contributions	Sale	\$332.06	\$40.300	(8.240)
Employee Deferral	Contributions	Cash Receipts	\$50.00	\$1.000	0.000
Employee Deferral	Contributions	Purchase	(\$50.00)	\$39.280	1.273

Consider the example to the left of a data page from the 401K site at a top U.S. brokerage.

Can you identify the column whose values accurately describe what happened in this account?

The previous question was a trick one, none of the columns is an accurate description, at least not if you want to use the data for something like cash flow analysis or a dollar-weighted rate of return calculation.

Relying on the “Activity Type” column and naively using a simple dictionary approach irretrievably loses the vital distinction between the cash receipt for \$17.80 (a dividend generated by the account) and the cash receipt for \$50 (a deposit into the account.) Only an aggregator designed to meet the stringent requirements for analytics would know to look for cases like this and avoid losing critical information.

Only after Semantic Integration, does ByAllAccounts have the clean, consistent data set needed for reconciliation, analysis, and consolidated reporting.

#### Step 4: Business Logic

The Semantic Integration process puts together a personalized financial view: the portfolios, their accounts, the holdings in those accounts, and the transactions against those holdings. The final step in the ByAllAccounts database building process is to apply business logic, interpreting and categorizing financial information, which enables us to provide this high quality clean data directly into various portfolio management and trust accounting platforms.

Portfolio, accounting systems, and back office solutions supported include:

- Advent Axys® and APX®
- Schwab Centerpiece® & PortfolioCenter®
- Morningstar Advisor Workstation Office Edition®
- SmartStream TLM®
- AccuTech AccuTrust®
- InData IMS®
- Interactive Advisory Software IAS®
- Rockefeller RockIT™
- WealthTouch
- Orion Advisors
- Black Diamond Performance Reporting

Examples of the business logic are:

- Asset class categorization
- Sub-asset class categorization
- Comparison of actual to target allocation
- Computation of allocation adjustment
- Net worth history
- E-mail alerts (daily security activity, holding “issues”)

#### Summary

There are now just two vendors actively competing in the advisor space – ByAllAccounts and CashEdge. ByAllAccounts has focused on the segment of the market that wants quality, complete, reconciled data suitable for performance measurement. CashEdge, through their partner Advisor Exchange, has focused on the segment of the market that wants consolidated holdings, without a need for reconciliation and performance management. The higher end advisors (in terms of the wealth of clients they serve) go to the BAA solution.

Account aggregation works well for all advisors that are focused on providing the complete holistic picture of their client’s worth. However, unless you can produce clean data, nobody will ever trust the technology. This means that the vendors that provide solutions based on reconciled data will ultimately dominate the landscape.

