



The Art of Omniture: Some Tips and Tricks

October, 2006

Gary Angel, Joel Hadary, and Paul Legutko

7430 Redwood Blvd.
Novato, CA 94945
(800) 763-2821

By its very nature, web measurement and analysis for large sites is complex. SiteCatalyst, one of the most analytically flexible and comprehensive web analytics tool on the market today, makes web measurement and analysis much easier – easier but not simple. Like most mature enterprise software solutions, it provides a rich user interface with numerous tools and functions. That very richness can work against users, however. As software grows, it tends to accrete layers of functionality – often quite unknown to the vast majority of the user-base. And, of course, having so much functionality can make it hard to figure out just where the thing you need really is.

Here are some great techniques for getting the most from SiteCatalyst – ranging from new ways to use tools you’re probably familiar with to finding some new tools you may not even have known existed. These Tips and Tricks cover the entire gambit of SiteCatalyst tools, SiteCatalyst Report Suites, Dashboards, Omniture Discover™, Data Warehouse, Advanced Segmentation, and the Excel Plug-In.

ASI’s, Discover, and Data Warehouse

Visitor and Visit segmentation is one of the most powerful tools of Omniture and indeed of any web analytics solution. SiteCatalyst has a number of ways to do visitor segmentation, each of which differ in terms of their flexibility and time to process. ASI (Advanced Segmentation Insight), Discover, and Data Warehouse, in our experience, have the following advantages and are best used according to the following criteria:

- **ASI.** ASI is the most comprehensive of the segmentation tools available in SiteCatalyst. By using ASI, your traffic segment is provided with every report also available in the SiteCatalyst Report Suite. This includes – uniquely to ASI – pathing analysis, such as fall-out and pathfinder, summary reports, dashboards, custom metrics, and the ability to conduct data-extracts and excel spreadsheet datablocks. ASI’s provide the nicest reports, since they are identical to those available in the Report Suite. Because ASI processing is slower than other options, and only a limited number of ASI’s are usually provided in standard agreements, ASI should ideally be used for very specific, narrow segments, preferably for page-view containers or commerce variables, representing, say, up to 10% of your overall traffic. Segments which will split your website into larger units should be processed through Discover or Data Warehouse, until a sufficiently detailed and refined segment can be identified for ASI.



- *Data Warehouse.* Data Warehouse is the most flexible of Omniture's tools when it comes to getting data. The segment constructor is identical to that of ASI, and most reports and metrics are available for download. Most analyses which you could possibly want to do are possible – given the right implementation – through Data Warehouse. While its segment-creator might take some time to get used to, an almost limitless range of segments are possible. Since Data Warehouse takes a short while to process (usually 24-72 hours), and since the data comes back as straight excel files, data warehouse is ideal for analysts who can manipulate the data into more user-friendly reports on a regular predictable basis. Monthly reporting is very often carried out through Data Warehouse requests that are regularly scheduled. Since Data Warehouse is done on the complete data set, you know that all numbers coming back are the most reliable possible in SiteCatalyst.
- *Omniture Discover.* Omniture Discover is the analysts' play-tool and the last-minute report savior. Although it does not provide the same number of reports as the report suite and contains fewer possible items for segmentation than the segment builder in Data Warehouse, results appear instantly. An analyst can chase any whimsical idea about visitor segments without worrying about system resources or limited ASI or Data Warehouse requests. Similarly, if a manager approaches you and wants reporting on a specific segment for a meeting later that day (as frequently might happen), Discover's instantaneous results can be a life-saver. For larger websites, Discover data might be sampled, which inherently means that the data is not "real", but overall percentages, rankings, and ratios should be accurate enough. Discover is often the first step of any analysis, so that useful or interesting segments can be identified and then plugged into ASI or Data Warehouse. Exporting out of Discover, while not as straightforward as ASI, is still faster and easier than Data Warehouse, and the downloaded reports contain the Discover charts which offer visual appeal.

Data Warehouse

The Omniture Data Warehouse is probably the most powerful of Omniture's products. It relies on a segmentation builder which, because of its sophistication, can be tricky to use. Here are some tricks and guides to help you along the way:



- Whether you're including or excluding in the Segmentation Wizard, remember that the default logic is "or" for non-nested containers, and "and" for nested containers. Putting a series of Page View containers under the "Exclude" canvass will not exclude PV's that match all of those criteria, but PV's that match any of them. For example:

<u>Example A</u>	<u>Example B</u>
(Include)	(Include)
PV = x	Visit where
PV = y	PV = x
PV = z	PV = y
	PV = z

- In Example A, this will give you all pages that are either x, y, or z, while Example B will be much more narrow, only giving visits that include x, y, *and* z.
- Including a visitor container but excluding PV's will not get rid of visitors. Make sure that include and exclude canvasses have the same level of granularity.
- Many "whole number" variables, such as path length, visit number, or clicks to page, are – unbeknownst to you – actually stored in SiteCatalyst with two decimals, so asking the segment wizard to include where a variable "equals" 1 will not return anything, because it's looking for "1.00". One way to avoid this is to use "greater than" or "less than" instead of "equals".
- Delete your segments – when loading the Data Warehouse interface, it has to load all your segments, which can take a long time. The "library" function allows you to store useful segment components and won't clog the system as heavily.
- For any segment, it is useful to include a "totals" report no matter what other kind of metrics you're interested in, because the Data Warehouse is not consistent in providing blank or empty data-points. For example, if you run a "search engines" report, you might get a "none"



value with PV, Visits, and Visitors, but you won't get a "none" value if you run a "Clicks to Page" report. So for percentages and ratios, you'll need a totals report that provides total visits, visitors, and page views for any segment.

- De-Duping of visitors and visits is very important in any analysis, and the numbers one gets in Data Warehouse are not always de-duped. This manifests itself in several ways:
 - There is no "daily, monthly, yearly" unique visitors option in metrics in Data Warehouse. This is because the date range you enter sets the parameters for uniqueness in Data Warehouse. If you pick a date range of 68 days, your visitors will be unique for that 68-day period. So your numbers from the report suite will not necessarily match up with the numbers you get from Data Warehouse.
 - Any report which contains multiple pages (pages, previous pages, next pages) will provide page views which add up to the total, but visits and visitors will not be de-duped. Entry pages, Search Engines, Exit Pages, Referrals, or keywords will not be de-duped for visitors, but will be de-duped for visits. Certain kinds of commerce reports and E-vars will de-dupe visitors as well, depending on how these have been set up.
 - Remember that for certain reports (keywords, referrals, for example), the report suite records "instances" or "occurrences". Data Warehouse, however, only gives you Page Views, Visits, or Visitors (and your commerce variables). "Instances" do not always correspond to "Visits" or "Page Views", and you should take care when comparing these metrics.

Omniture Discover

Discover is an attractive Omniture product because it allows you to process segments immediately, rather than waiting for data warehouse or ASI. Pay attention to the following, however, when using Discover:



- The segmentation wizard in Discover, although it looks the same as the segmentation wizard in Data Warehouse, actually gives you a different set of variables. Certain variables are not available in Discover, such as original entry page or clicks to page, while certain are available in Discover but not in Data Warehouse, such as Entry and Exit reports on custom variables.
 - One corollary to this is that it is not possible to download certain reports. In Discover, you can “download” or “export” most reports. Download simply means putting what you see into an excel spreadsheet (or word, pdf, etc). Export, on the other hand, means transferring the current configuration – segment and report – into Data Warehouse, so that you get real, instead of sampled, data. But if you’re using an item in the segment builder which is incompatible with Data Warehouse, this won’t work.
 - This also means that certain Data Warehouse segments will not work in Discover. If you load a segment into Discover and see a phrase such as “where Unknown equals...” this means that Discover can’t recognize the item on which you’re segmenting. This “unknown” also appears on items which do cross over between Data Warehouse and Discover, so pay attention -- you may need to enter the segment and re-adjust.
- In large websites, Discover samples the data by as much as 9 to 1. This can distort the calculations by as much as 10 to 15%. When working extensively with Discover, it is useful to run a few reports concurrently in Data Warehouse or the Report Suite, in order to establish a general estimate of the error rate between sampled and un-sampled data.
- Discover is run on Javascript, so make sure you have a recent version installed. Remember that you can’t cut and paste numbers in Omniture Discover, so have a convenient scratch-pad ready if doing calculations.

Know when to use ECommerce

Commerce variables and analysis make up one-third of the SiteCatalyst report suite interface, yet many companies do not take advantage of all the possibilities these reports offer. Conversion metrics are clearly important, but SiteCatalyst goes far beyond mere conversion, allowing for



all levels of detail in terms of revenue, traffic, and visitor segment. Some items to think about:

- SiteCatalyst can actually assign revenue values to particular pages or events on your site. Rather than just collecting page-views of pages you think are “conversions”, consider putting monetary values to pages which might correspond to advertising revenue, sales amounts, lead estimates, and so on. You might find sources of revenue-producing traffic that have better and more lucrative optimizing potential than straight conversion pages.
- Data Warehouse segments use commerce variables to create segments. This opens up many analytical possibilities independent of page views, visits, and visitors, which are the usual segmentation variables. You can segment not just on visitors, for example, but visitors who spend a certain amount of money on your site. Or you can look at high-revenue sessions to identify traffic patterns.
- Evars and Custom Events are very similar to Custom Variables. Know when to use each kind of variable. Evars are great for scalar, numerical values, or alphanumeric variables with which you would like to correlate commerce events, such as revenue, purchases, etc. Events are great for counting particular activities on a site (and are considered a metric, like page views and visits, for reporting purposes). Custom Variables are great for both scalar and string variables, but are not as easy to link to commerce metrics.

The Value of Correlations

Correlations in SiteCatalyst can be very powerful, if you know how to use them and set them up. This is not, strictly speaking, a correlation in a statistical sense, but rather a cross-tabulation component. For example, I might have 3 business units defined in a custom variable called “Business Unit”: video games, DVD’s, and Audio CD’s. I also have a custom variable which captures the keywords a visitor types in on internal search pages (called “Internal Search Keyword”). Correlation will allow me to filter and cross-tab one variable by the other, so that if I want to see a report on every search term used *only for* DVD’s, I can run a correlation report that will link the two variables. Some things to be aware of when setting up correlations:

- The larger variable should be the correlation filter. To put it in SiteCatalyst diction, “Internal Search Keyword” should be correlated to “Business Unit”, not the other way around. This is



- because correlation reports can only be run on one filter at a time. I can see a report showing me every search keyword for DVD's, but not a single report showing me three columns, one for DVD's, Videos, and Audio CD's. Clearly I would not want to do it the other way around – showing every business unit for one keyword at a time.
- Remember that most kinds of correlations can also be retrieved through a data warehouse pull. In the above example, if I don't have a correlation set up, I could still ask Data Warehouse for a segment of every page view where the business unit was "DVD's", and then ask for an "Internal Search Keyword" report. So don't feel that you have to have a correlation for everything – just items on which you expect to do regular reporting or want to incorporate in dashboards.

VISTA Rules

VISTA (Visitor Identification, Segmentation and Transformation Architecture) is one of SiteCatalyst's solutions for post-tagging revisions and modifications. Basically, I might have my website nicely set up, with each page properly tagged, following my initial implementation effort. Then something happens – a new reporting requirement, a re-design of the site hierarchy, new campaign landing pages, or a new content management system. I might need to create new variables or re-adjust the definitions of the existing variables, but re-tagging the entire site seems daunting. Enter VISTA. VISTA allows you to set up, in effect, new variables or modifications to existing variables by applying query logic to existing sets of variables. So if I need to assign the label "2006 Offerings" to 100 pages on the site, VISTA rules will allow me to do so easily, and without having to re-tag the site. VISTA also allows for visitor segmentation, much like ASI (see below). Although VISTA implementation can take some time and is dependent on contractual stipulations, it has the advantage of avoiding back and forth between site designers, programmers, and production implementation, and is thus an attractive option.

Custom Metrics

Custom metrics are an under-utilized feature of SiteCatalyst, but one which can provide more insight than just page views, visits, and visitors. Custom Metrics can be defined according to a wide set of different available metrics using standard arithmetic, and are then available for a wide set of traffic and pathing reports, as well as trending. Some Custom Metrics we have found useful in the past include:



- Page views per visit (measures page-view consumption on a site)
- Page Views per Visitor (measures visitor engagement with the site)
- Entry Propensity (entries per visits – a measure of independence)
- Exit Propensity (exits per visits – the probability that a visit to a page or group of pages will end there).
- Session Engagement ((visits – exits)/visits – the probability that a visit to this page will go on to do something else on the site)
- Exit Rate (exits/total site exits) – the probability that this page is the exit page for the site.

Excel Plug-In

Excel was never designed to handle anywhere near the millions of records that are generated by a large web site. However, it is an excellent reporting tool and is extremely useful for completing many web analyses and creating management reports. Even though Omniture has done a good job integrating these two products (SiteCatalyst and Excel), they are designed for very different functions. We have found there are few tricks to getting the integration even better:

- Overall we recommend structuring your SiteCatalyst-Excel spreadsheets with input worksheets which only receive data-block data and reporting/analysis worksheets that are linked to the input worksheets. All formatting, calculations, charting, and reporting should be done in the reporting/analysis worksheets – not in the input worksheets.
 - Ideally, it is best to have one data-block per Excel worksheet. If that is not possible, have as few as possible and make sure they are separated by several rows or columns.
 - Minimize Excel formatting of the data-blocks, especially when dates are included.
- Data-blocks should be arranged horizontally, not vertically, across the worksheet. That way, if one data-block gets confused, there's less of a chance that it will overlap another.
- In the last screen of the construction of a datablock, Omniture asks you for the datablock name (this is right above the box where you locate a cell to place it in). Although many people ignore this box



and just stick with the default, naming the datablock something simple, with no special characters, dashes, or spaces, improves performance

- Worksheets which have data-blocks in them should have no spaces, dashes, or special characters in their names.
- If a spreadsheet doesn't refresh correctly after hitting "refresh all", try doing each worksheet individually. We've found this to work well, even if it takes longer.
- You should not – or even cannot – have another excel spreadsheet open when refreshing a SiteCatalyst Excel spreadsheet. It's also best not to have other MS Office applications running either – Excel spreadsheets embedded within PowerPoint presentations have been known to foul up the refresh command.
- It is better to create a spreadsheet all at once, rather than incrementally, and do not logout or login again while creating it. Under no circumstances should you login as someone else while constructing the spreadsheet. Use a single login until the spreadsheet is done, then "refresh all" before logging out and sharing with others.

Regarding the actual construction of datablocks, here are some tricks we've found. Note that the same interface is used when conducting a Data Extract, so many of these observations apply to both.

- In recent versions (up to June 2006 or so), the "start date" and "end date" values for datablocks which look to cell references for their date ranges were reversed in SiteCatalyst. What we've done, therefore, is simply switch the values within the spreadsheet – make A1 the end date, and A2 the start date. Then it works. Monitor your results, so that, if with a future release, this is fixed and you're all of a sudden not getting valid results, simply switch the start and end dates.
- Say you want a report – trending, in particular – where one needs date, variable value (a value of a Custom Variable), and metrics in the same datablock, as follows:

<u>Date</u>	<u>Search Engine</u>	<u>PageViews</u>
Jan 1	Google	98769
Jan 1	Yahoo	65747
Jan 1	MSN	54325
Jan 2	Google	98876



- To get this, in the datablock-builder, put the date on top, the variable on the left, and the metric on the right.
- Suppose you want the following configuration, where you want, say, “last month” as the data range:

<u>Date</u>	<u>Search Engine</u>	<u>Page Views</u>
January	Google	98798
January	Yahoo	87654

- To get this, select granularity – in this case, “month” – in the date selection interface. When you do this, even if you only have one month, it will allow you to show all three in the same datablock.
- Remember that if you aren’t selecting granularity in your metrics, and are asking for totals only, there isn’t any de-duping of results. For example, if you are running a pages report and ask SiteCatalyst to list the pages, you would get something like:

<u>Page Name</u>	<u>Page Views</u>	<u>Visits</u>	<u>Visitors</u>
HomePage 18,000	30,000		20,000
Page A 13,000	28,000		15,000
Page B 10,000	25,000		13,000

However, if you just ask for totals *in a pages report* (by putting “pages” in the upper-left container in the data-structure window), this is what you’ll get:

<u>Page Views</u>	<u>Visits</u>	<u>Visitors</u>
73,000	48,000	41,000

- Page Views are accurate, but Visitors and Visits are inflated (remember Visitors and Visits are not additive). If you do want totals in an Excel plug-in, use the Report Suite Totals report, not a summary of a different report. If you’re doing a complex search within the datablock, then you’ll have to go to DataWarehouse or Discover to get the de-duped totals.



SiteCatalyst Dashboards

The Dashboards in SiteCatalyst are relatively straightforward to use and construct. Based entirely on the report suites, they tend to be reliable and accurate. Still, there are some discrepancies and tricks we've noticed which can be useful.

- We've found it very useful to attach a Contents and Glossary at the top of each dashboard. This can be done by inserting a text-based reportlet and then writing the glossary and contents in simplified html. Note, however, that underlining, table borders, frames, special spacing, and other html code will not work in the reportlet, though bold , italics <i>, certain table commands <table>, <tr>, <td>, and breaks
 will. Note, however, that the maximum size for any table is 15 rows, so if you have more, use multiple tables.
- The "summary" reports, such as page summary, can sometimes be misleading, particularly if you're reporting on a custom variable which aggregates many pages into one (like "conversion" processes). This is because even when multiple pages are rolled up into a single report, the report still uses page-views as its atomic level of calculation. For example, if I've grouped 30 pages of my site into a single variable and then ask for a summary, SiteCatalyst calculates exits on the page-level, not on the level of the 30-pages aggregated. There is no hard-and-fast guideline to provide here, since the configuration of custom variables varies from site to site, but just pay attention to the figures that come out of the summary reports, particularly entries and exits.
- Remember that calculated metrics do not cross between users, and that each calculated metric has to be separately set up by each individual user. Name your calculated metric something short (ca. 10 characters), since the name usually gets truncated in reports.
- The Omniture Dashboard Viewer is meant for management to be able to view rapidly the contents of a dashboard without having to go into the SiteCatalyst interface. While convenient, remember that only charts, and not tables, are included in the Viewer. The second line of the reportlet header is often a good place to put verbiage explaining the reportlet contents.

Summary

With software as feature rich and powerful as SiteCatalyst and its associated tools, it's hopeless to think of a truly comprehensive set of Tips and Tricks. A book, even many books, probably wouldn't suffice. So in this white paper we've tried to cover techniques that seem to us either



particularly powerful or especially common in use. In some cases (like ASI/Discover/DataWarehouse), we've seen a great deal of user confusion about when to use each tool. That isn't much of an issue with Dashboards or Excel Plug-Ins. In the former case, it seems to use that the most common pitfall is failing to take full advantage of the customization opportunities. While with the Excel Plug-in, the most common stumbling blocks seem to be with operational "how-to-dos."

No matter your experience level with web analytics and/or SiteCatalyst, we hope this White Paper will give you at least few good ideas on how to work better, faster and smarter!

About SEMphonic

SEMphonic is a leading tool-independent web and search engine marketing analytics consultancy. Founded in 1997 as a web-analytics consultancy and based in the San Francisco area, SEMphonic's clients have included some of world's largest, most complicated and sophisticated web sites. Supporting companies like Charles Schwab, AOL, Intuit, Cybertrader, Morgan Stanley and American Express, our mission has been to deliver analytic consulting that drives web channel success.

Our customers run a wide variety of web analytic solutions and PPC tools - tools like WebSideStory's HBX Analytics, Omniture's SiteCatalyst, Google Analytics, Webtrends, Clicktracks and more. We believe that practitioners with the deepest knowledge of web analytics are most likely to get the maximum value from a tool. Our experience across tools gives us a deeper and better appreciation of how web analytics can be done.
www.SEMphonic.com

