HBX and SiteCatalyst: Comparative Analysis

June, 2006

Gary Angel, Joel Hadary, and Paul Legutko

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

Table of Contents

HBX and SiteCatalyst Comparison: Introduction

There are many third-party web analytical tools available, among which WebSideStory's HBX and Omniture's SiteCatalyst are among the most prominent. With SiteCatalyst aiming for an IPO later this year and grabbing a larger market share, and WebSideStory doing everything they can to prevent that from happening, it is important for businesses considering a third-party web analytics tool to know what the major strengths and weaknesses of each tool are, and how they compare to each other.

This short report looks at some specific aspects of web analytics, and how HBX and SiteCatalyst approach the issue, how they compare, and other problems we have encountered while using them. The areas to be examined include:

- HBX and SiteCatalyst Comparative Analysis
- Implementation Process
- Overview Reporting
- Conversion Analysis
- Path Analysis
- Visitor Segmentation
- Report Distribution
- General Usability

At the end of the day, SiteCatalyst and HBX are remarkably similar in scope and function, with SiteCatalyst providing more analytical capabilities, but generally more difficult to use, and with software generally less reliable than HBX. The strengths of HBX, we feel, lie in its MS Excel report building integration and its easy-to-use active segmentation, while SiteCatalyst is superior in its analytical flexibility and comprehensiveness.

Implementation and Tagging

Three years ago, one of the major decisions facing businesses looking for a web analytics solution was whether to employ log files or to use javascript tagging. Now, the industry standard has sided with javascript tagging, largely because of tagging's greater flexibility and analytical potential. It should be emphasized, however, that there are still many challenges to effective tagging, particularly at the organizational level. HBX and SiteCatalyst vary in the degree to which these difficulties can be overcome, even if the basic system of tagging pages is similar.

When evaluating SiteCatalyst and Hitbox we considered three aspects of their tagging strategy: how easy it is to tag pages; how much information needs to be embedded in the tags to make measurement work; and how much flexibility does the software have in making changes once the initial tagging is completed. In terms of tagging pages, we see no significant difference between the two vendors. Both use almost identical strategies for tagging pages and we do not believe that the up-front implementation cost would vary between the two. In terms of how much information needs to be in the tag, we also see little difference between the two vendors. Post-tagging analytical capabilities are also similar between both HBX and SiteCatalyst.

Both HBX and SiteCatalyst rely on tagging to define significant hierarchies of pages. In both cases, we see this as a significant analytic and reporting limitation – probably the biggest drawback to each product. Grouping pages by type and studying their usage and impact as a unit is a central technique of good web measurement. Since pages are often published in physical directories that bear little resemblance to the desired analysis structures, it is essential to be able to group content logically. There is no reason why a tagging system need rely on the tags to group content – but unfortunately, they tend to do this. In our view, this is a significant weakness. Hierarchy is fluid and much more likely to change in response to analysis goals than things like conversion funnels. When adjusting a hierarchy means re-tagging a page, it is a pretty safe assumption that this type of analysis will never be conducted.

Failures of hierarchy are evident everywhere in both SiteCatalyst and HBX. In many cases, reports are nearly unreadable because

like content isn't grouped, aggregated at levels and properly named. SiteCatalyst's "like-with-like" limitation, in which pages can only be compared to pages, or site sections with site sections, severely limits the kinds of analysis one can do on sites with large page populations. Although one of the advantages to tagging is that each page can be appropriately and clearly named, in practice this is rarely done, and it is often the case that page names in the SiteCatalyst or HBX reports read like encoded gibberish, compounding the problem resulting from lack of hierarchy. To expect editors to comb through listings of hundreds of pages which are not transparently named, in order to identify interesting patterns of usage, is simply not practical. Nor is it reasonable to ask managers to extract out all pages of a specific kind from a list of 100 pages when compiling volume. When people have to do stuff like this, web measurement simply gets ignored or misused.

For hierarchy analysis, the best solution would be to be able to create a hierarchy of pages after tagging is finished, so that it could be adjusted and re-adjusted, but this feature is not yet available in either HBX or SiteCatalyst. While both vendors do provide hierarchy reports, these continue to be based on tagging implemented ahead of time. We thus give the edge to neither vendor here.

Overview Reporting

Website reporting has a relatively small set of core data items: visitors, visits, page views, referring domains and browser/os types. Basic web site reporting provides OLAP analysis against various combinations of these variables over time.

This is the simplest part of web site measurement and most any competent software package is going to provide complete coverage of the core information. However, web measurement does provide some interesting twists on standard OLAP and it is useful to consider how well HBX and SiteCatalyst handle these.

Good overview reporting is a combination of an easily understood and navigable user interface, readable presentation of the information and the ability to easily cut across and combine the data in sensible fashion. Both HBX and SiteCatalyst have state-of-the-art web-based GUIs. It is largely a matter of taste which one prefers in terms of navigation. Indeed, they are strikingly similar in terms of the basic layout of application elements. It is hard to imagine a preference here driving a decision. It has been our experience that online tools like HBX and SiteCatalyst (and also their Report Builders) are for fairly sophisticated users. It's not that you need to be a professional statistician to use them, but neither are they the domain of VP's of Marketing. The ideal user for software of this complexity range is a business analyst – and any competent business analyst will quickly adjust to either of these interfaces.

There is a bit more to be said about the presentation of information. Both HBX and Catalyst generally present reports with a graphical element on top and a tabular report below. HBX has a nicer graphical tool – however, in both cases the chart presentation is almost always a useless appendage. Most web variables (especially content) have too many segments to be presented graphically. This is especially true given the limited hierarchy capabilities of the two products. We can't recall the last time we actually looked at a chart of this type when doing real analysis.

On the other hand, both HBX and SiteCatalyst have developed some web specific presentations that work quite well. The graphical presentation Site Catalyst uses for Page Summaries is quite nice. One only wishes that SiteCatalyst would allow more flexibility in its comparative trending – the most interesting charts (in so far as charts are interesting) are those which compare today/this

week/this month data to yesterday/last week/ last month data, but these occur sporadically and unpredictably within the SiteCatalyst interface. HBX does not present this kind of data except as an "average", the calculation of which is unclear. Similarly, SiteCatalyst's path presentation is perhaps the nicest we have seen. HBX takes a different approach to pathing that we find quite a bit less readable but that does have some analytic advantages.

Both products rely on Search to help limit the returns on a report. In our view, this is a poor substitute for good hierarchy management. Search is a very unreliable tool for finding pages of common functionality. We think this is error prone and probably leads to a lot of incorrect analysis. However, the approach is very similar (even in tools like Advanced Search) for both products. An advanced filtering system with a more complete set of available logical expressions would be much more useful, for both products. WebTrends has introduced this kind of querying function, though generally the range of variables available in WebTrends is much more limited.

The third element of overview reporting is the ease of crossing variables – this is a fundamental part of OLAP analysis. Surprisingly, both systems are weaker in terms of cross-tabulation analysis than one would expect. There are numerous cross-tabulation possibilities that you can't do and you can't always get the statistics you need from the cross-tabulation you want. SiteCatalyst offers a "Correlation" option (for a price), which is really not correlation in a statistical sense, but rather a cross-tabulation function allowing you to filter reports through different variables. While this can be useful, the interface is very limited, essentially presenting the user with a case-by-case summary (the user must specify the variable value), rather than a true cross-tabulation of the data which would present all cases at once. A similar filtering mechanism is available in Omniture Discover™, but this tool presents a far more limited range of variables for analysis and is otherwise error-prone. HBX does not provide such a function, but would rely on active segmentation to achieve the same result.

Take, for example, browser reporting. HBX and Site Catalyst include detailed reports showing browser share by vendor and version. A user can also get detailed reports by screen-resolution and host of other occasionally relevant environment variables. This information isn't particularly valuable. However, we do sometimes look at browser in special situations – particularly in terms of abandonment issues with a page, which also could be affected by

screen-resolution. With “Correlation” in SiteCatalyst, a user could select a particular screen-resolution and get a report showing different browsers used in that particular screen resolution, though not for all screen resolutions in the same report, forcing the user to run many reports, each selecting a different screen-resolution. In HBX, one would have to go to datawarehouse or active segmentation, which might be just as easy an alternative in SiteCatalyst. Thus while SiteCatalyst has laudably introduced a new functionality in order to address this weakness in cross-tabulation, its current implementation and interface does not give it much advantage over HBX in the long run.

This is just one example of a common problem in each system. The problem arises because of the high-cardinality of some key measurement variables (particularly pages) that limit the number of permutations you can produce in a cube. This same problem significantly effects visitor segmentation – an issue discussed later.

Conversion Analysis

The analysis of conversion processes is a significant part of many web projects. Conversion analysis typically falls into two categories: indirect conversion and conversion processes. With indirect conversion, the analyst is concerned primarily with the relationship between viewing a page of content (typically not part of an order process) and eventual success. This type of analysis is the single most common technique we use when evaluating web sites. The second type of analysis focuses on understanding what happens in a conversion process. This analysis generally focuses on what happens at each step of a well-defined sequential flow.

For indirect conversion, both tools are lacking some significant capabilities. Neither do a credible job of handling multi-session conversion. Since multi-session conversion is more the rule than the exception this severely limits the value of either tool for good indirect conversion analysis. Multi-session conversion can be handled in two ways currently: by setting up campaigns (in both SiteCatalyst and HBX) or by Active Segmentation. HBX's campaign tool is more easy-to-use than SiteCatalyst's, with greater ability to create campaigns without resorting to pre-defined tracking codes. SiteCatalyst, however, has greater visitor segmenting capability -- HBX can only create a visitor-based segment in retrospect, not an ideal solution for daily, ongoing reporting. SiteCatalyst also has a more impressive list of Commerce variables which, when combined with visitor segmentation, allows for some powerful conversion analysis.

For single influence conversion both vendors have improved on previous releases, allowing for the ad-hoc creation of fairly complex point-to-point reports. SiteCatalyst has some advantages, with greater flexibility in defining custom variables and events, making it possible to perform some fairly complex analyses.

HBX has three ways to accomplish indirect influence analysis. The first is to use conversion funnels. However, conversion funnels are really intended for sequential process analysis. They don't do a particularly good job of influence analysis even when they are defined beforehand (which used to be a requirement in previous releases). The second HBX approach would be to use pathing and specify search criteria to restrict the domain. In theory, this is quite powerful since you see every possible path. In practice, it's actually quite inconvenient for most purposes. It makes it very hard to identify the key intervening steps and come up with counts. In

addition, HBX crops the data so that the analyst has only X number of top paths. This can significantly impair analysis results, since the size of the tail remains generally unknown. The third possibility is to use page-affinity in HBX, which shows the number of visits to one page which also includes a visit to a different page. While this is a useful way of looking at correlations in traffic patterns, the data as presented is somewhat misleading, since percentages do not refer to page-visits but to the percentage of the whole data set.

For conversion analysis both vendors are therefore useful, with the edge perhaps going to SiteCatalyst for its wider range of analytical capability.

Path Analysis

Both HBX and SiteCatalyst provide powerful tools for visualizing the Paths that are being used by visitors to the site. At one time, Path Analysis was the Holy Grail of web measurement and to those who don't do analysis it seems like it is the ideal way to examine web sites. As with so many other web measurement truisms, however, path analysis is often much less powerful than the uninitiated might expect. The reasons for this are several: many significant web behaviors cross sessions and most path analysis is single session; there are so many pages of content on most web sites and the navigation paths are so open that paths become mind-numbingly confusing; and where paths are tight funnels, they generally don't reveal anything very interesting.

Neither HBX nor SiteCatalyst do anything to solve the multi-session pathing problem. The too-much-content and too-many-paths problem is more interesting. HBX chooses a top path approach that shows color-keyed icons strung out in a row. The paths can then be tuned using a search function. To see what pages are, you roll the mouse over them (there's a legend but it's generally scrolled off-screen). There are real advantages – and disadvantages – to this approach. On the one hand, this provides very detailed multi-step path information sorted in usage order, which is great because paths of any length are included. On the other hand, we find it very difficult to take in. An analyst would have to be pretty tuned in to this method of presentation to use the tool without a lot of sweat equity.

Site Catalyst's approach is different. Site Catalyst provides path information as a series of vertically descending paths with each one clearly labeled. Again, the approach has advantages and disadvantages. The layout is much more readable – but the paths consume a lot of space and it is very difficult to take them all in.

Both tools allow the user to screen-off paths using the search functionality. This is quite useful and makes each tool significantly better than it otherwise would be. We frequently do path analysis using roll-ups in a hierarchy. SiteCatalyst does support this feature when hierarchy levels have been defined, but is a little limiting since the hierarchy paths must be at equal channel levels – something that isn't always the case. But the feature is, nevertheless, a nice plus.

Finally, both tools have a “previous” and “next” pages report (“to” and “from” in HBX’s terminology). For immediate analysis of Click-Through Rate and other navigational analysis, these can be very useful.

On the whole, we give a very mild nod to SiteCatalyst’s style of presentation. As with Overview reporting, however, we don’t think this advantage is particularly significant in a purchase decision and we could easily imagine an analyst with slightly different tastes preferring the HBX approach.

Visitor Segmentation

Most good web site analysis demands visitor segmentation – the ability to look at just the behavior of one slice of visitors. This is not always the industry view of visitor segmentation – which often focuses on the ability to count how many visitors have a certain behavior. These are very different activities. The second is important, but primarily as a prelude to the first. Both SiteCatalyst and HBX include visitor segmentation capabilities – but in somewhat different fashion.

In HBX, a certain number of active segments per month are allowed by contract. The segment is defined by the user using a fairly clear and consistent logic. After some time (perhaps 12 – 24 hours), the segment is active and can be viewed. All reports allowed in the unsegmented suite are available for the segmented one. If even more analytical capabilities are desired, datawarehouse requests can be made through an API using full query logic. One limitation in HBX's segmentation is that the segmentation filters are only available for visits and visitors, while SiteCatalyst allows, in addition, Page Views or commerce variables. This allows greater flexibility in nesting variables one within the other, and allows an analyst to approach the same problem in multiple ways.

Whatever advantage SiteCatalyst has because of its wider analytical capabilities, it loses when it comes to ease and clarity of usage. SiteCatalysts' drag-and-drop segment constructor can be frustrating to use, since the user is limited to the options offered (and does a business analyst or IT professional really need the nursery-school drag-and-drop feature?). This same constructor interface handles Active Segmentation, Data Warehouse, and Discover, so that it is not possible for an analyst at any point to write his or her own query logic, another potential point of frustration. Both Datawarehouse and especially Active Segmentation take a notoriously long time to process. In order to address this, SiteCatalyst introduced Discover, which is based on a sample of the data (usually 3-7 to one, depending on the size of the account). While this makes segmentation much faster, the segment constructor here has even fewer options than the one for Active Segmentation, and the application in general is somewhat buggy. All in all, we prefer the more reliable and easy-to-use, if more simple, segmentation capabilities of HBX over SiteCatalyst.

Report Creation and Distribution

Most consumers of web-measurement information aren't skilled analysts but website and marketing managers. Their primary interface to tools like HBX and SiteCatalyst is (and should be) via reports generated from those systems and distributed out.

The HBX report-builder creates files directly in Excel (which we think is ideal) and appears to include all of the necessary functionality to support most end-user needs. The report-builder is limited by the same kinds of considerations that limit the analyst tool (weaknesses in hierarchy and visitor segmentation) but insofar as it can be considered a standalone product it meets any reasonable need of the client.

SiteCatalyst also downloads reports in Excel format, although some of the more complex reports require the user to request they be delivered via e-mail (not ideal for analysis purposes). In order to address the ubiquity of Excel among end-users, SiteCatalyst has an Excel integration tool which, when functioning according to spec, can be a powerful solution to many reporting requirements.

Both HBX and SiteCatalyst offer custom dashboards; however, as with Excel integration, the edge here goes to HBX, with far greater freedom in constructing them. SiteCatalyst's much-hyped Dashboard Viewer we found to be almost useless, since only charts could be included, without any other data.

Search Engine Tracking

With so much emphasis on SEO and PPC these days, both SiteCatalyst and HBX have seen to it that Search Engine Referrals and the distinguishing of paid vs. organic keyword clicks dominate their "Acquisition Sources" reports. However, reconciling this data with anything Google or Yahoo reports, as well as conducting detailed analysis on this information, remains problematic.

Both SiteCatalyst and HBX employ two different strategies towards getting at Paid vs Organic traffic. The first is to turn PPC or Organic activity into campaigns, a functionality built into HBX but needing special implementation in SiteCatalyst. Both traffic is then treated exactly like a campaign, with response and conversion information provided. This method is more seamlessly integrated in HBX than SiteCatalyst.

The second method is through keywords, and by distinguishing between paid and organic keywords. Neither vendor has the edge here, since both do a good job at reporting search traffic, though for depth of analysis SiteCatalyst might have the advantage, since they do not crop the data. Analyzing this traffic, however, is difficult, since it is, by definition, session-based and would require segmentation to take it to the visitor-level. And of course, the lands or responses reported by HBX or SiteCatalyst are always suspiciously lower than those reported by Google or Yahoo. Neither HBX nor SiteCatalyst can account for Content Matching by Yahoo or Google, an increasingly big problem as Content Matching by advertisers is constantly expanding.

It should also be noted that SiteCatalyst has its own "Search Center" which consists of a report suite integrated with Yahoo or Google, as well as a rules-based bid management tool. Like all bid-management tools, it's usefulness depends on the intensity of the user's involvement in day-to-day PPC activity, the intricacy of the user's bid-management strategies, and is subject to the tergiversations of Google and Yahoo API's. One significant drawback to Omniture's Search Center is the fact that its Google-to-Omniture data feed occurs only once per day, so that any real-time bid management is probably just as easily done in the Google interface.

Link Analysis

On many sites, determining the Click-Through-Rate on various links from principal router pages is an important reporting requirement. The solutions by both vendors are only so-so, and the increasing use of Flash technology makes even these solutions difficult to use. There are generally four solutions: a next page report (HBX and SC), Link Analysis (HBX), tagging (SC and HBX), and Browser Integration (HBX and SC).

Next Pages reports (and path analysis) are always a fall-back analytical method, but are problematic for CTR because of the high degree of back-button browsing, which inflates CTRs on particular links. Tagging is always a solution (SiteCatalyst is particularly amenable to on-click handling of link-variables), but requires separate tagging ahead-of-time (as do the "custom links" variable in SiteCatalyst).

Browser Integration and link analysis are the two most useful and direct methods of viewing CTR. HBX's Link Analysis reports are extremely useful in this regard, presenting each link as a separate line and reporting on the number of clicks through it. HBX's Active Viewing and SiteCatalyst's ClickMap are Browser Plug-Ins for Internet Explorer, in which Clicks and CTR of a particular link is superimposed on a snapshot of the page you're interested in. While useful as a overview of page activity, there are two drawbacks to this strategy: first, multi-media and flash technology confuses the application, so that particularly dynamic pages or page-areas (often the most interesting from a CTR perspective) do not parse well, or at all, in these plug-ins. Secondly, these plug-ins require some considerable effort in order to transform them from snapshots into Excel or PowerPoint for digestion by management.

In the end, and with dynamic pages becoming the norm, the best CTR analysis will have to rely on tagging solutions until more sophisticated tools become available. For link analysis we applaud the efforts by both SiteCatalyst and HBX, and give the edge to HBX for the Link Analysis report.

General Usability

At the end of the day neither HBX nor SiteCatalyst are easy-to-use, intuitive programs, but with some experience the user will learn the ins and outs of both. They are both viable and useful enterprise-wide web measurement and analytical solutions.

A significant issue in comparing SiteCatalyst and HBX on a general level is the quality and reliability of the software. No software is perfect, of course, but our experience is that SiteCatalyst is both slower and less reliable than HBX, particularly when it comes to more recently released products such as Omniture Discover or the SiteCatalyst Excel plug-in. We have seen SiteCatalyst Excel-Based dashboards which were populating data perfectly suddenly stop working, with bugginess resembling more an internal beta than a fully released software package. Live technical support for SiteCatalyst can also be elusive, as each client organization is permitted only a handful of designated supportees.

Similarly, running reports takes more time in SiteCatalyst than in HBX. While reports generate fairly quickly in HBX, the user often must wait several minutes for basic reports in SiteCatalyst, or finds to his or her frustration that the system has timed out. Datawarehouse requests, Active Segmentation, data extracts, or even basic report downloads can take a long time, which can be frustrating for analysts struggling against time-sensitive organizational reporting requirements. Javascript-based Omniture Discover™ is only a partial solution, and rumors that planned further expansion and integration of Discover with the Report Suite will only mask, rather than solve, this problem.

We have found HBX much more reliable, both with the software itself and with the live support provided. Our only general frustration with HBX is more analytical, namely, that they only provide the "Top X" results for most of their reports (top pages, top search terms, top paths). While perhaps this makes the software faster, one would like to have the option available if necessary without having to go through datawarehouse.

Comparative Report Card

Feature	HBX	SiteCatalyst	Comment
Implementation Strategy	C+	C+	Poor ability to post-define page hierarchies
Overview Reporting	B+	B+	To some extent a matter of taste.
Conversion Analysis	B+	B+	Wider range of custom variables and commerce events in SiteCatalyst.
Path Analysis	B-	B	A slight edge based largely on presentation preference.
Visitor Segmentation	B+	B-	SiteCatalyst provides better analytical potential but is difficult to use.
Report Creation and Distribution	A-	B	HBX has done a better job integrating with Excel.
Search Engine Tracking	B+	B	HBX's campaigns are more tuned for keyword tracking; Omniture's Search Center still remains predominantly another bid management tool among many.
Link Analysis	B+	B	HBX has a useful Link Analysis report.
General Usability	B+	C+	SiteCatalyst is slower and less reliable than HBX