# What's Next?: The Future of Web Monitoring

"Tis some visitor entreating entrance at my chamber door ... I betook myself to linking fancy unto fancy, thinking what this ominous bird of yore—what this grim, ungainly, ghastly, gaunt, and ominous bird of yore meant in croaking 'Nevermore.'"

-Edgar Allan Poe, "The Raven," 1845

In the first chapter of this book, we said that smart companies make mistakes faster. It's perhaps more accurate to say that smart companies *adapt* faster, learning from their mistakes and forever tightening the feedback loop between what they attempt and what they achieve.

The monitoring tools and measurement techniques we've covered in this book are your eyes and ears, giving you an ever-improving view of your entire online presence. We've tried to lay out some fundamental principles within these pages: determine your goals, set up complete monitoring, baseline internally and externally, write down what you think will happen, make some changes, experiment, rinse, and repeat.

If writing this book has taught us anything, however, it's that the technology of web monitoring is changing more quickly than we can document it. New approaches to improving visibility surface overnight, and things that were nascent yesterday are mainstream today.

In the course of researching and writing the text, we've talked to over a hundred companies building monitoring technologies—some still stealthy—and this has afforded a tantalizing glimpse of what's coming.

# **Accounting and Optimization**

All of the monitoring we've seen in this book serves one of two clear purposes: accounting or optimization.

In many ways, analytics is simply how we account for the online world. After all, we're collecting metrics, such as daily sales, average shopping cart size, and advertising revenue. Accounting-centric analytics is likely to become the domain of the finance department as that department pushes for more visibility into the online business.

Today, analytics is an afterthought in most application development efforts, thrown in as JavaScript tags just before a site launches. This is unlikely to continue. Expect accountants and controllers to define standards for data collection from web systems, and expect investors to demand more real-time data on the progress of the online business. It won't be long before Generally Accepted Accounting Principles (GAAP) are part of the requirements documents that web analysts need to incorporate into their analytics strategies.

The other half of analytics focuses on continuous improvement. This is the domain of product management, interface design, and copywriting. The goal is to get more visitors to do more things you want for more money, more often. Every tweak to your site is an attempt to edge ever so slightly toward that goal, and every visit a hint at whether the attempt was successful.

While much of this optimization can be done manually, at least at first, organizations with sites of any substantial size will soon find this an overwhelming task. They'll turn to multivariate testing, as well as the automation of usability analysis and visitor feedback, allowing web platforms to adjust their content and layout dynamically to maximize conversions based on what works best for a particular segment.

# From Visits to Visitors

It's relatively easy to consolidate monitoring data by time and to segment it by shared metrics. For example, conversions by country can be compared to historical performance within that country, and periods of poor performance can be aligned with bounce rates to see if there's a correlation.

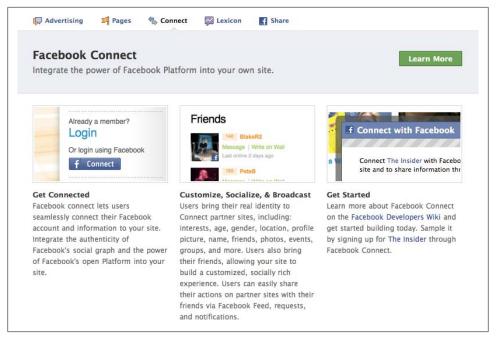


Figure 18-1. Facebook's business solutions include onsite advertising, the Lexicon analytics platform, paid pages for companies, and tools such as sharing and Connect to encourage users to share their actions

A far bigger challenge, however, is linking individual visits to individual outcomes. In essence, the web monitoring industry is moving from a focus on visits to a focus on visitors. The individual visitor—either named or anonymous—must be tracked across social networks and the conversion process in order to understand the return on any marketing investment. It's a return to the days of database marketing.

Identifying an individual visitor is critical. This happens in several ways:

#### Visitor self-identification

Visitors sign up for a federated social model, such as Facebook Connect, Microsoft Passport, or OpenSocial, which shares their activities with others (Figure 18-1). The information may be anonymized by the owner of the social platform, but already, tools like Radian6 rely on OpenSocial to identify key influencers in a social network.

#### Grammatical links

Text from comments and blog posts are analyzed using language-parsing algorithms that try to identify a single person's online identity from the content she creates. This works for prolific web users for whom a considerable amount of content exists.



Figure 18-2. Social messages, such as these ones on Twitter, contain clues about the spread of social conversations, in some cases even documenting when a social message changes platforms

#### Metadata and fuzzy association

Many social networks include basic information about users, such as their first names or the cities in which they live. By combining this information, analytics systems can make educated guesses about who's who and try to join multiple online profiles into one metaprofile.

#### *Understanding the flow of messages and referrals*

Most social conversations carry threads within them that can be tracked. On Twitter, words like "via" and special characters such as "RT" (reTweet) signify information that's being passed down the long funnel. Tracking tools can follow these threads to examine the social spread of a message (Figure 18-2).

#### Embedding tracers in online activity

For years, analytics companies have relied on cookies to persist information on browsers so they can track new versus returning users. As we've seen, with the emergence of short message interactions like Twitter messages and Facebook status updates, there's nowhere to easily hide tracking information other than within the URL.

#### Asking visitors to log in

Sometimes the simplest solutions are the best. Asking visitors to sign in so that they can be uniquely identified and tracked is the surest approach to tracking. It's also less risky from a legal and ethical standpoint, since site operators have the visitor's permission to know who they are.

Web analytics companies that focus on tracking visits are quickly retooling to track individual visitors, either anonymously or as known individuals, linking user logins to Twitter accounts and blogs. Of course, harvesting and mining every individual's interaction across the Internet has serious privacy concerns, but it also steers web analytics toward the realm of CRM and a one-to-one relationship with a company's customers.

## **Personal Identity Is Credibility**

A corollary to the move from visits to visitors is that personal identity becomes currency. Companies and individuals will start to manage—and defend—their online reputations as personal currency that gets them jobs and gets their voices heard. Online reputation is the twenty-first century's reference letter.

Expect to see certified profiles and recommendations provided by the community. Already, highly ranked contributors on SAP's community site use their communitygranted awards on their Linkedin profiles, leveraging them as proof of their knowledge and ability to work with others when seeking new employment.

These reputations will have a strong influence over which links communities follow, and personal credibility scores will factor into community analytics. Metrics such as friend count, number of followers, and other metrics of popularity are already giving way to more sophisticated measurements on sites like Twitalyzer, such as the reach of a reTweet or the number of community members with whom a person converses regularly.

# From Pages to Places and Tasks

We've looked at the concept of places and tasks several times in this book: as a way to think about your website, as a means of defining analytics metrics, and as a tool for understanding the end user experience you need to measure.

The shift from document-centric monitoring to a places-and-tasks model is inevitable. We simply don't use the Web as a set of pages any more. Startups like Kissmetrics are poised to take advantage of this shift in the fundamental structure of web monitoring.

This will change what we monitor. For places, we'll care about efficiency and productivity; for tasks, we'll focus on effectiveness and abandonment. We'll also use new metrics, such as the number of times a user undoes a particular action, to measure efficiency and effectiveness.

Instrumenting places and tasks requires the collection of many new metrics, often from sources that can't simply be instrumented with JavaScript: email campaigns, URL shorteners, RSS feeds, site crawlers, dynamic ad content, and inbound calls. The collection infrastructure needed to launch a complete web monitoring service will become far more significant and will require the integration of many different disciplines.

# **Mobility**

The majority of devices that can access the Internet today are mobile. That means they're being used in a wide range of locations in conditions that are strikingly different from the desktops most developers assume. iPhone surfers pinch and tap their way through websites, posing unique usability challenges. Wireless devices suffer from sluggish performance, spotty coverage, and high levels of packet loss.

It's not just the surfing conditions that change when the Web goes mobile. Now, location is an important factor. Knowing where a user is physically located will determine which content they see and which ads they're served. Advertisers will want to know which shopping mall a visitor is in so they can tie online traffic to real-world outcomes. Some stores are even launching free Wi-Fi tied to Facebook Connect—a pact in which the visitor reveals his identity in return for Internet access.

# **Blurring Offline and Online Analytics**

Communities will form around locations, much as they coalesce around hashtags today. Transient groups will appear and dissolve as web visitors move from physical location to physical location, and this data will be useful not only for segmentation, but also for mapping social graphs based on physical proximity.

Real-world activities are increasingly tracked online, through devices like the Fitbit or games like Akoha, and they're changing the way we think about analytics. Networks of sensors will soon collect and share our lives with one-time, tacit approval, leading to a new understanding of audiences—and new privacy concerns.

This blurring of online and offline worlds is aided by real-world tracking tools such as RFID (radio frequency identification) tags, barcodes that can be scanned by PDAs, VOIP services that link calls to analytical data, caller ID systems, and so on.

## **Standardization**

One of the reasons it's hard to get a complete picture of your web presence is the lack of standards. With the exception of rudimentary logging (like CLF), some standard terms, and some industry initiatives for scoring (such as Apdex), there's no consistent way to collect, aggregate, or report web monitoring data.

This is particularly true for performance monitoring. It wouldn't take much to improve the state of the art with a few simple additions to the technologies we use today. If, for example, the world's browser makers decided to cooperate on a single DOM object that marked the final click on the preceding page, it would dramatically improve the usefulness of JavaScript-based RUM. Unfortunately, such standards are woefully absent.

Synthetic testing vendors differentiate their offerings through their instrumentation (browser puppetry versus scripting), the number of test points they have, the detail they offer, and the reports they generate. There's no reason why we can't have a standard language for defining scripts, or one for annotating page load times—this wouldn't undermine any vendor's uniqueness, but it would be a boon for web operators who need to manage and interpret test data.

Similarly, a common format for associating analytics, WIA, RUM, and VOC data across service producers would make it far easier for web analysts to do their jobs; yet data is seldom available for export. We need a common format for visitor records that can work across products.

The same is true of community monitoring. There's no standard measurement of ranking across social network analysis tools, and it's up to customers to demand openness and interoperability.

Increased concerns over data privacy and online activity may spur such standardization. It's unlikely, however, that web monitoring companies will voluntarily standardize the log formats and metrics that lock in customers and make migrating from provider to provider more difficult.

# **Agencies Versus Individuals**

Public relations (PR) has traditionally been an outbound business, controlling the flow and the message. Today, the new PR is community interaction. Lobbying and advocacy are so passé; we'd rather they change sentiment and tweak the zeitgeist.

While many companies have started community management as an in-house effort often part of user groups, support communities, or blogging—PR agencies are taking notice. Seeing their roles as keepers of the corporate message at risk, agencies want to own the interaction.

This heralds an arms race between agencies and the organizations they serve. On one side of this pitched battle will be agencies, which have the ear of the marketing department and can spread worries of liability and spin control. They'll employ mass-messaging tools—many of which we've seen but have yet to be launched publicly—to send personalized mass messages to followers. On the other side will be the righteously indignant individuals claiming the moral high ground of the genuine and devoting themselves to nurturing real relationships with their markets and followers.

How this will play out is uncertain. Several agencies we spoke with are building their own proprietary tools to mine communities on behalf of their clients; meanwhile, conversation listening platforms are turning support teams into skunk works PR agencies.

The outcome of this battle is in the hands of the communities themselves, who will either embrace the new communications channels even if they're disingenuous, or will reject them, exposing agencies as an insulating layer that impedes interaction to an organization and its audience.

# **Monetizing Analytics**

Silicon Valley is littered with business plans based on ad revenue. In the era of free software, companies expected to make their money from advertisements. Yet pay-perclick advertising margins are notoriously thin, and often insufficient to support a business model.

Knowing which segments are most likely to do what you want—and how to reach them—is critical for any online business. Much of this information flows across Internet connections, social networks, search engines, and websites, and the keepers of these sites see analytics as the new AdWords: in the world of free, analytics is the cash register, and many companies seek to monetize their services by selling insights into online visitor behavior.

Let's look at some of the industries hoping to cash in on analytics and web intelligence.

### **Carriers**

Mobile phone carriers had a great business. They provided services such as voicemail, text messaging, videoconferencing, newsfeeds, and more, for a fee. With the advent of the consumer Web on smartphones, however, they've lost the ability to unbundle individual services and charge for each of them. Apple finally capitulated and allowed Skype on the iPhone, which will have a material impact on long-distance revenues. Carriers have lost control of their original business models.

Instead of just becoming "dumb wires" for data applications, one way carriers can reclaim their revenues is to watch what mobile users do. Already, some firms (like Neuralitic) make analytical tools that capture mobile user traffic, aggregate it based on the applications and services those users access, and help the carriers analyze their businesses. The data could easily be sold to third-party clients who want to better understand mobile usage patterns.

## **Search Engines**

Search engines are the most obvious candidates for analytics revenues, since paid search is their primary source of revenue, and giving customers tools to understand the effectiveness of their campaigns is a part of the sales cycle. Search engines can do more than just analyze keywords, however: they have data on trends, geographic spread, and more. We've covered many of the free tools that search engines—particularly Google make available, and most of them are subsidized by advertising revenues. Search engines do have premium offerings for some of their tools, however.

#### **URL Shorteners**

The many URL shortening sites—dominated by bit.ly and Tinyurl—are well poised to make money from analytics. They sit between a web user's intentions (clicking the link) and the outcome (delivering that user to a site) and can build a great deal of information on the spread of social messages, as well as which content works best for which demographics, social networks, times of day, or geographies.

These URL shorteners also threaten social news aggregators. Sites like Digg, reddit, and Slashdot rely on surfers to submit stories they find interesting, and despite the efforts of these companies to make the submission process easy—through toolbars, embedded buttons, and so on—it still requires effort to submit a story.

bit.ly, awe.sm, and Tinyurl require no effort beyond the shortened URL, however, to learn what people find interesting. They can see which content is spreading most quickly, and segment it by user demographics, simply because they're acting as middlemen between the links and the content.

While today many of these sites offer analytics for free, bit.ly is already forming business relationships with marketers who want to track the spread of their messages. As organizations seek to do multivariate testing of community messages, they'll turn to these URL shortening companies for metrics and reporting data.

## Social Networks

There's a war brewing between open, opt-in tracking through sharing and participation across many community sites, and the all-inclusive walled gardens of social network giants.

When you receive a message from a friend on Facebook, you're notified by email. To respond to that message, however, you need to click on the link provided in the message and respond using Facebook's tools. The same is true, to a large degree, for other walled garden social networks. LinkedIn is somewhat more open—email messages can be taken out of the platform and into traditional email—but social networks in general have been reluctant to integrate with systems outside their walls. It's only in the face of competitive pressure from sites like Twitter that social networks have been willing to put even carefully guarded openings in their garden walls.

Walled garden social networks want to be a one-stop social platform whose users don't need to venture elsewhere. Even "open" features like Facebook Connect are aimed at pulling content from elsewhere on the Web inside those walls, where members can further discuss them.

It's far easier to track—and monetize—community activity when that activity all happens in one place. A picture, leading to a Wall post, leading to a message, resulting in an invite, can all be tracked back to a user and segmented by demographics. As soon as the user's online experience transcends a single site, tracking him becomes far more difficult.

The tension between walled gardens and open communities isn't just technical—it's cultural. Early adopters and the web-savvy favor an archipelago of online communities loosely connected through APIs and hyperlinks. More mainstream consumers, however, want a unified set of tools built around their social graphs, a group hug with the friends they know. The social makeup of communities makes this distinction clear: a real-world meeting of Twitter friends feels like speed dating; a Facebook meet-up is more like a class reunion.

Whether a site is a lone island in an archipelago of social sites or a one-stop continent, it's still got to make money. To do so, the site's operators will deploy monitoring tools that help advertisers understand the relationship between social network activity and business outcomes. This will happen either by keeping them in the walled garden the Facebook model—or by finding ways to connect community members' many online interactions.

This will in turn cause (justified) concern over the tracking of personal data online. Our network of families, friends, and acquaintances will become increasingly valuable to marketers who seek to optimize online activity, but we'll be increasingly concerned about sharing that network with others.

This picture of social networks might seem bleak: rather than them becoming a tool for collective human consciousness, they'll slide inexorably toward their role as yet another marketing tool. But this has happened throughout much of the Web already, from websites to blogs to email, without entirely tragic consequences.

Online communities will continue to thrive and connect, but when organizations want to monitor their roles within their communities and tie this information back to business outcomes, they'll start tracking them.

#### SaaS Providers

One area in which analytics has been glaringly absent is within hosted applications. SaaS offers an unprecedented opportunity for employers to measure the productivity of their employees, yet many SaaS offerings lack even rudimentary visibility into who is using the application and how they're doing so.

Sales is a numbers game, and sales force automation products, such as Salesforce.com, already give managers the ability to see which employees are meeting their goals. Call center tracking applications have similar analytics. With more and more enterprise applications available in SaaS formats, it's likely that we'll see "productivity analytics" become a more commonplace feature of SaaS sites.

Today, however, it's hard to implement this type of tracking yourself. Most modern SaaS platforms won't let customers embed tracking tools within their pages, so you're at the mercy of the SaaS platform to find out how much your employees are using the service, where they're getting stuck, or where they're pressing Delete the most.

As businesses start to expect the transparency and accountability of analytics, we expect them to demand this kind of visibility from all of their software tools.

## A Holistic View

What's most obvious from today's monitoring industry is that siloed perspectives will not last. Vendors are rushing to broaden their feature offerings, either by building all the components themselves, acquiring other companies, or using creative mashup approaches to integrate their reports into entrenched solutions, such as Google Analytics. As Figure 18-3 shows, the distinctions of analytics, WIA, EUEM, and VOC are fuzzy at best, with many firms offering strong products in several segments.

Analytics companies are positioning themselves as the anchors of all this analysis. They are the cornerstones of monitoring strategies because they are the tool that sees the payoff. Whether it's an improvement in performance, positive visitor sentiment, a better page design, or a groundswell of community support, everything that happens online has an outcome that ties back to the organization's goals, and the analytics tool is where those outcomes are defined.

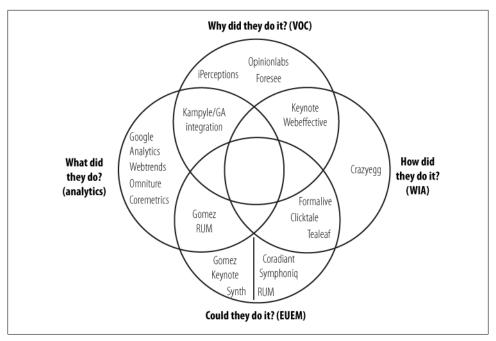


Figure 18-3. Examples of blended product functionality as the monitoring industry consolidates and vendors merge

Web analytics tools, however, are really a subset of the broader category of BI. For larger enterprises, it's likely that a central BI repository will include not only the online outcomes tracked by analytics platforms, but also call center traffic and retail transactions. In other words, the analytics giants of today may be consumed by even bigger companies that are the central repositories of a firm's interactions with all of its customers.

## The Move to a Long Funnel

Increasingly, web users learn about destinations from their peers and their friends. Referrals have been the basis of much of the Internet economy, driving online advertising and paying for many of the websites we take for granted. This, in turn, has fueled the analytics industry. A referral from a website, however, isn't as genuine as one from a trusted friend or an admired celebrity. Already, Twitter mega-users rival Digg and reddit in their ability to send thousands of visitors to a site in an instant, bringing it to its knees.

Community monitoring tools will have no choice but to partner, as marketing executives demand more details about how their money is being spent. The community monitoring industry today is where the analytics industry was only a few short years ago (Figure 18-4).

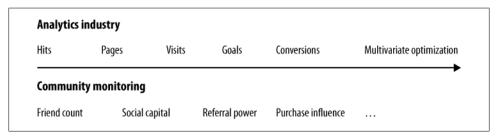


Figure 18-4. The community monitoring industry lags behind web analytics by a few years, and will undergo many of the same changes

Then, analytics focused on hits and pages; today, it's about conversions, segmentation, lifetime visitor value, viral coefficient, and optimization. Similarly, today's community monitoring tools focus on popularity and sentiment, but are quickly moving toward an analysis of users' long-term impact on the business.

The shift from web referrals to social referrals has far-reaching consequences for the analytics industry. While analytics continues to be important for onsite activity and optimization, such as reducing abandonment and deciding which content works best, it's becoming less useful for offsite analysis. Expect to see a flurry of partnerships between established analytics vendors, as well as changes to the offerings of leading advertising network providers who try to insert themselves into the social graph.

# A Complete Maturity Model

As we've looked at various forms of monitoring, we've considered a maturity model that begins with basic technical visibility and moves toward a more integrated, comprehensive perspective of your entire online presence. Table 18-1 shows the questions an organization is trying to answer at each level of maturity.

Table 18-1. The questions an organization answers at each level of web monitoring maturity across various forms of monitoring

	Level 1	Level 2	Level 3	Level 4	Level 5
	Technical details	Minding your own house	Engaging the Internet	Building relationships	Web business strategy
Analytics	How many hits did I get?	How many conversions have I had?	What segments or sites are sending me the most, best traffic?	How loyal are my users, and what's their lifetime value? What content works best for them?	What KPIs should I use to manage my organization? How are online and offline channels related?
WIA	What did visitors see and dick?	Where in a form or a page do users get stuck or abandon? How can I im- prove my design?	How do different visitor segments interact with my site?	How long does it take for people to learn the application? Where do they get stuck? Can I resolve their prob- lems by having their sessions in my CRM?	Can I automatically serve optimal content and navigation for different segments and levels of visitor experience? How efficient/productive are users?
NOC	What do emails and "contact us" forms say about visitor satisfaction?	When I intercept visitors, what do they tell me about my site, their perception of me, and their loyalty to my products or services?	When I engage my audience elsewhereon theInternet, what do they say? What's their senti- ment? How does word ofmouth spread?	How does visitor feedback correlate with conversions? What do visitors tell me about my competitors?	How do I tie visitor sentiment to employee performance, software release success, and product road maps?
Synthetic monitoring	Is my site up? How fast is it?	Are transactions working? When things are slow, which tier of the infrastructure is guilty?	How are the communities and third parties on which I rely working? How's my CDN? Which parts of the Internet are slowest? Did I make my overall SLA?	Howdoes my performance and availability affect my business? What are my aggregate scorecards?	What's the best balance of performance, capacity cost, and revenue? What's the business plan for uptime? Does IT get bonuses based on delivery? How cost-effective is our elastic capacity?

	Level 1	Level 2	Level 3	Level 4	Level 5
RUM	What are the traffic flows? When a problem happens, what does the trace show me?	How fast are pages loading for actual users? Which pages, servers, and objects are slowest? How healthy are transactions for actual users?	Which segments of visitors are most correlated with performance issues? Who are my worst-served customers? For whom were SLAs violated?	How did the latest release affect user experience? What's a given user's lifetime SLA?	How do online incidents tie back to CRM and support response? Have I automated the refund and dispute resolution process with RUM data?
External community	Am I being discussed?	Do I have a presence? Are people engaging with me on my own site?	Are they listening to & amplifying me? What's sentiment like? Which messages work best with which communities?	How does the "long funnel" affect my revenues? What's the lifetime engagement of a visitor across all social networks? Who are my most vocal/supportive community members?	What will the community buy? How do I automatically match the right messages to each audience? How does my community help itself? Is virality a part of business planning?
Internal community	How much information have we generated?	How good is the information we're generating? What's most and least used? How easily can employees find it?	How integrated is KM with corporate culture? What are people sharing most and least?	What's the payoff? How does KM improve per-employee contribution or reduce per-employee cost? How are people connecting with one another? Are our collective guesses good?	What does our organization know? What is our organizational knowledge worth? How can we use better KM posture to improve competitive position? Are we hiring and firing based on employee knowledge contribution?
Competitors	What are my com- petitors doing?	How does my site com- pare to competitors'? Is it as fast? As easy?	Am I more popular than com- petitors? Better ranked? More adopted by certain segments?	How loyal are my customers? What can they tell me about my positioning versus others?	Is competitive analysis pulling data from VOC, community, and EUEM? How should competitors' strategies affect my own?

Ultimately, your complete web monitoring strategy needs to span both your site and the other sites and communities that affect your business. It needs to track visitors through their offsite activity, arrival, usage, and long-term engagement. It must monitor how they help your business, and how they refer others to you.

# A Complete Perspective

Figure 18-5 is a quick reference for many of the things you need to consider when formulating your web monitoring strategy.

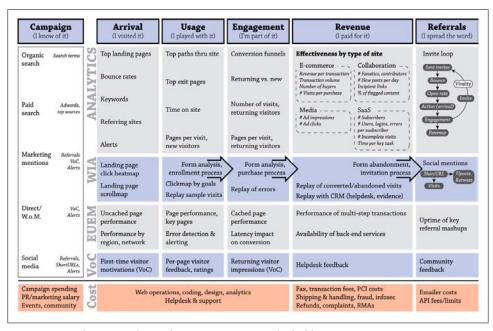


Figure 18-5. What a complete web monitoring strategy looks like

Complete monitoring includes analytics, WIA, EUEM, and VOC. It also includes looking at the many stages of visitor engagement, from initial arrival and use of the application through revenue growth and referral management. It encompasses effectiveness and efficiency measurements, as well as the instrumentation of virality and social spread. Finally, monitoring must be considered in the context of the many costs that are involved.

# **The Unfinished Ending**

We're hoping this text provides a good background on web monitoring and that it has armed you with enough detail to dramatically improve the visibility you have into every website that affects your business. But it doesn't end here; by the time you read this book, the information is out of date.

We hope you'll join us at www.watchingwebsites.com for updates and perspectives on this dynamic topic. Together, we can continue the conversation.