1

Managing the Flow of Information: A Crucial Skill

RSS and Atom are about information. More specifically, they are about information flows. They exist in a complex setting that involves many technical issues. Some of these issues are heavily programming orientated, and we will write about them at some length in later chapters of this book. However, there are many other factors that are important in considering the use of RSS and Atom, including interface issues and human usability issues. These issues are important determinants in how RSS and Atom information feeds are best used and influence when and how they work most effectively.

Your success in a rapidly changing information-oriented society is going to depend, at least in part, on how well you handle the flows of information that reach your computer, whether it is a traditional desktop machine, a laptop, or some other device. Similarly, the success of the software you create will depend on how much it helps your users meet their information needs.

> **The processing of information is undergoing a major paradigm shift with progressively wider adoption of RSS in its various versions and the emerging Atom format. The existence of automated information flows is changing fundamentally how we access and use information.**

RSS and Atom are simply two tools that enable developers to provide better access to information for users. One of the big changes that RSS and Atom are bringing about is improved access to all kinds of information for many types of users. It is no exaggeration to say that there is a prospect of everybody with Internet access being able to access all the information that they can handle. The changes in information flow that are already under way are immense. How we handle the flow of information is becoming increasingly important.

RSS and Atom don't exist in a vacuum. They are parts of a larger picture of information management. This chapter gives you a broad overview of several of the important issues that impact the use of information.

In this chapter, you learn:

- ❏ How changes in information flow are taking place
- ❏ How to think about your information needs
- ❏ How to think about the information needs of your customers

New Vistas of Information Flow

The existence of RSS and Atom information feeds, or information flows, are opening up new possibilities in how information can be shared and accessed. Because information is easier and easier to access, software developers and software users must begin thinking about new and different issues.

> **In this chapter, and elsewhere in the book, the term** *information flow* **refers to the notion of movements, or flows, of information. The term** *information feed* **refers to the more specific notion of flows of information that use RSS or Atom.**

The Information Well and Information Flow

For newcomers to the world of information feeds, the terminology and concepts can be far from clear. This section explains our view of some of the changes that have begun and presents them within the larger context of technological changes that have taken place in various localities over hundreds of years. The changes in information flow will take place in a much shorter time span, but we hope the analogy will be helpful to newcomers to this field.

In the sections that follow, we use the analogy of how access to water and supply of water has progressed over time to illustrate what might otherwise seem a very abstract overview of information supply and management.

The Information Well

Consider that water was, and in some places still is, drawn from a well. Access to Web-based information until recently was similar to how a villager in a remote African village accesses water. It had to be done repeatedly and each time one sought out information, he or she had to take time to get to the source of the information. We had to go there every day that we needed information or wanted to see if there was new relevant information available, just as someone has to carry water from the well each time it was needed.

In an information society we need information to survive. Going individually to single Web sites or Web pages to check for updated information is an inefficient use of time. It is a strategy that is deficient in that we only go to places we already know.

Search engines can help reduce the inefficiency of that approach, because we can go to the Google site, or similar search engine, to look in one place for all the new information. Again, since the same search might be done repeatedly, the process is potentially inefficient.

Facilities in browsers, such as bookmarks, help to alleviate the repetitive nature of visiting individual Web sites. More recent features, such as the capability in Mozilla Firefox to open a folder of bookmarks with a single click, help to alleviate the repetitive nature of resuming online research at a particular point in the search.

Fundamentally, though, we are walking each day, or several times each day, to the online information well.

The Information Flow

Most of us live in homes where water comes to us, rather than us having to travel to the water. It makes a lot of sense that information, too, should flow to us. It avoids the repetitive actions of going to visit individual Web sites and, if done well, achieves easier, more efficient and more effective access to information.

Naturally, to set up information flows in an appropriate way takes time. Hopefully, the time spent customizing the information flows you see will be a largely one-time inconvenience, a little like the one-time inconvenience of laying water pipes. Because information is not a single entity, as water is, the situation is more complex and the analogy is too simple. But I hope that you see that having information flow to users is potentially much more efficient than having the users visit individual Web sites multiple times a day.

The Information Flood

At times, once information flows have been set up, typically using RSS or Atom information feeds, the problem is not in accessing information but being overwhelmed by the volume of information.

There are also parallels with the quality of water and of information. After a flood or tsunami water is plentiful but not all water is usable. Just as we want to find clean water in such a situation so, when dealing with information, do we need to assess information quality and reliability.

The flow of information is becoming an overwhelming flood. At the time of writing the number of information feeds is measured in millions. The total number of feeds may be approaching 10,000,000 of which roughly half may be active. Estimates vary. The reality is that nobody really knows exactly how many feeds there are. It is clear that the number is so large that the individual user has to have mechanisms to prevent the flood of information from preventing them from locating the information they need.

Managing Information

Efficiently and effectively managing information is a crucial skill both for individuals and for businesses. In a highly competitive business environment, efficiency in handling information is a key competitive tool. Businesses (or individuals) that handle information inefficiently compared to their competitors will fall behind.

The ability to locate relevant information in a timely and efficient way can help you succeed in the information economy in which we all now have to work. Several general questions are important with regard to the how information is best handled.

What Do You Want to Do with Information?

Information is not, of course, identical to water but the parallels continue. Just as you might want to use some water for drinking and some for washing, so some information is relevant for a particular purpose and some is not. One of the fundamental steps in information management is to define what you want to do with the information you access.

One line of thinking about information feeds is that the information that is brought to you is purely of transitory interest. In other words, it's temporarily useful and not worth storing. It's news, and will be replaced later today or tomorrow or next week by other information. That line of thinking looks on information as essentially disposable. However, it's not as simple as that. The following sections explore different responses to individual pieces of information.

Browse and Discard

For some information you will very likely simply want to browse the headlines and discard the information (or allow it to be replaced automatically when a feed is updated). Information such as conventional news headlines might fit this category. The majority of them are likely to be skimmed, many are likely to be of little interest if any, and much of the associated content will be left unread. A small proportion of the information (which will likely vary according to your situation) will be sufficiently interesting or relevant to be worth reading. However, much of it won't be of interest and you won't want or need to see it again.

The browsing of headlines is, perhaps, an inescapable chore. Many of them will be irrelevant, but it is possible to scan a large quantity of headlines in a short period of time. Keeping this irrelevant information for lengthy periods of time is not an issue.

Read

Some information will be more important or interesting and you will want to read it. Only after you have begun to read it will you be in a position to assess its current and possible future relevance to you.

Depending on the pressures on your time, you might not have time to read an entire article. Alternatively some articles you read may have content of more than transitory interest. The question of whether and how to keep the information comes into the picture. You have, explicitly or implicitly, to decide whether you want to keep it. Not all tools currently available make that process of longer-term retention of information easy. If you keep it, where are you going to put it so that it is easily retrieved and accessed when you want to review it?

Study and Keep

Some types of information, or information from certain sources, you might want to study in detail and want to keep, perhaps indefinitely. Long-term storage of this information is a significant issue. How should you group items? What search functionality is available in different tools that you can use for storage?

With these different information needs, you must find tools that give you ways to handle all these information types.

Taking Control of Information

You have specific information needs. Chances are you won't find a single tool that meets all of those needs.

> The term *aggregator* is used here for a piece of software that collects, or aggregates, information feeds. Typically you will be able to use an aggregator to subscribe to selected information feeds. Some aggregators have a built-in Web browser.

Aggregators have weaknesses and strengths just like any other class of applications. For example, when you open NewzCrawler or several other aggregators, in their default configurations, they will pop up a frenetic flurry of "pieces of toast" in the right side of your screen, over screen real estate that you might be using for another purpose. Whether listing screens of information that the aggregator has brought to you justifies the visual intrusion will depend on how you use information feeds. If you are interested in information being absolutely up to the minute you will probably want to review new information as soon as it is available and so some visual indication that new information is available is an advantage. Alternatively, you may be keeping a watching brief on a topic and don't need instant notification of new information. You need to be able to configure an aggregator, at an appropriate level of granularity, to give you information that you really want to see and avoid intrusion with information that isn't of a level of importance that justifies interrupting what you are currently doing.

Determining What Is Important to You

There is no shortage of information out there. In fact, the problem in many information domains is that there is too much information. Techniques for finding relevant information are key to successful and effective information flows.

So where do you find useful information? That's a topic we will return to in Chapter 4. For now, your bookmarks and Google searches are obvious places to start. If those pages match your interests there is a reasonable, or better, chance that any information feeds for those sites will also contain information that is interesting and relevant to you.

Avoiding Irrelevant Information

With increased amounts of available information, there is a growing need to filter out information that is not relevant to you. Time spent scanning headlines can be significant. Time spent exploring the "Oh that looks interesting" items can consume inordinate amounts of time, but not actually achieve anything useful.

One way to avoid, or reduce, irrelevant information is to choose carefully what feeds you subscribe to. However, many feeds contain diverse information and only a relatively small proportion of the information may be of direct relevance to you. Some aggregators enable the user to create filters so that certain types of post are hidden or only posts that contain keywords specified by the user are displayed. These techniques reduce the amount of irrelevant information, but at the risk of excluding interesting information present in the information feed.

If the topics of interest to you are rapidly changing ones and having a lead, however slender, over your competitors is a key issue, you might be willing to pay the price of screening all new feed items. For

example, you might be a stock trader and want to be aware of rapidly changing stock prices, to help you buy or sell at an advantageous price. In this case, screening a stock ticker feed might be preferable in some circumstances to missing out on a key lead that happens not to use the key terms that you have put on your watch list. I don't think there is any way to avoid this trade-off of scanning time versus the risk of missing interesting material.

Determining the Quality of Information

One of the important issues of online information is whether or not it is trustworthy. There are many levels of trust. For example, it would a mistake to give your bank account details in response to an e-mail without checking very carefully that the e-mail is from a legitimate source. But trust in the context of information feeds is likely to consist more of the idea of a source of information being reliable for accuracy rather than moral trustworthiness.

However, there are situations where trust comes into play. If you are reading the blog of a corporate blogger, how can you know whether or not he or she will give you an honest view? Being a corporate blogger is a difficult position. The blogger, almost inevitably, needs to more or less toe the company line. But if the corporate blogger does that too tidily then he or she is unlikely to retain readers for long. Over time, a reader has to form a judgment about whether the individual blogger is trustworthy.

Judgment of a more technical nature comes into play if you are following a blog about possible upcoming technical trends. Does the blogger have a good grasp of technical trends? Does he or she have a track record of getting things right in the past? To apply the water metaphor again, is the water of good quality or not?

Information Flows Other Than the Web

The issues regarding access to information are not new. Some have been present for centuries, but in different forms than those of the present day. Access to information today is unprecedented. A century or two ago it is almost unthinkable that anyone except the privileged few would have had access to the amount of information available on the Web today. In fact, even the richest and most powerful in society would operate in deep ignorance compared to average citizens in today's world.

This section briefly explores some information-related issues that have contributed to the success of other information media over the centuries. Depending on nature of the information conveyed, any proposed solution that uses RSS and Atom must be examined in the light of the strengths of more established media. Sometimes there will be no contest. For example, books cannot be in print instantly in the way that an RSS or Atom information feed can be available globally almost instantly after it is created.

Users, typically, won't care whether an information flow uses RSS or Atom. They want their information needs met. Seeing things from the user's point of view is key to designing and implementing quality information flow solutions.

Books

Books have been around for some 500 years. There must have been good reasons to keep a pretty stable format for such an extended period. One possibility is that no superior technology came along during that time. Until the advent of radio that was, perhaps, true. Another factor is that a book has a huge

convenience factor that online reading can't yet match. Try reading online during a power outage. During a power outage a computer becomes a useless piece of metal and plastic. Yet the low-tech book can, at least during daylight hours, continue to be read as normal.

Books still offer advantages for very large documents, because you can easily create your own bookmarks. In Web pages you have to rely on bookmarks that the document author has created. In addition, many people simply don't want to read for sustained periods of time onscreen. Books are more convenient to read for long periods of time. Many people simply don't like reading text onscreen for long periods.

Magazines

Magazines have, typically, different content from many non-fiction books. The articles are typically shorter, although often longer than a typical technically orientated blog post. Will information feeds, if a suitable business model can be found, increasingly replace magazines? I suspect that process is already happening. I used to subscribe to several computing-related magazines. Currently I subscribe to two. The others fell by the wayside, in part because I increasingly found them repeating information that I had seen on the Web or in information feeds some weeks earlier. The two magazines that I continue to subscribe to have what I would view as relatively meaty technical articles. I guess that similar articles are available online, but there is something satisfying about reading something in print rather than onscreen.

Newspapers

Newspapers, with rare exceptions, publish on a much faster timescale than a typical book or magazine. Twenty-four hours later that same newspaper and its content may have been discarded, used to wrap fish and chips (at least in the United Kingdom), or be burned or composted.

As mentioned elsewhere, handling news is where some protagonists of information feeds see their only role—providing information that is up-to-date and essentially disposable. Information feeds can compete with, and exceed, the speed of publication of newspapers.

Attitudes toward newspapers are slowly changing. People are finding more of the up-to-date information on the Web, through blogs and online news sites. The issue of reliability of information again features highly when choosing an online source of information. An increasing number of people are beginning to replace their newspapers with their newspapers' corresponding Web sites or a diverse range of online news sites. Surveys, such as the Online Publishers Association Generation study (`www.online-publishers.org/pdf/opa_generational_study_sep04.pdf`), suggest that newspapers are potentially under threat from online media including aggregators. Some data (for example, `www.wired.com/news/culture/0,1284,65813,00.html`) suggest that in some age groups newspaper subscriptions can't even be given away for free anymore.

Broadcast Media

Broadcast media have characteristics that are different from any paper-based media such as books, magazines, and newspapers. They can achieve instant coverage of chosen issues but suffer from a linear manner of presentation. For example, if I want a weather forecast from my favorite radio station I need to be listening at the exact time it is broadcast. Sometimes that is inconvenient for me. An online source of that information is more convenient.

Broadcast media also have significant production costs and other production issues. This has, for such practical reasons, limited the number of broadcasters. In some countries a business model of broadcasts serving up audiences to advertisers has been prominent. Blogging and the dissemination of blogged information through information feeds have greatly reduced production costs and therefore have opened up an online analog of broadcasting to a potentially wide audience.

The Web and Information Feeds

The Web when it was first widely available was in a way both wonderful and at the same time difficult to use. I first used the Web in 1994 in a text-based browser. At that time finding information was, as I recall, slow and clumsy. There was no Google although there were search facilities. Just finding my way around what then seemed to be a flood of information was intimidating and slow, in part because I had to navigate in a way wholly unlike anything I had done when using books, library catalogs, and so on. I can still remember the feelings of finding something interesting then, a little later, being totally unable to find it again. Tools have changed in the intervening decade and my familiarity with the metaphor of the Web means that things I found clumsy and disorientating in the early days are now routine.

Similarly, the arrival of information feeds will, I expect, seem confusing to some audiences. Currently I think it's likely that those using information feeds are largely a technically literate audience. When a wider, less technically literate, audience has information feeds made available to them, the usability of aggregators and other tools will be key factors in uptake. If the aggregator or other tool is hard for a beginner to use, then it is likely to lose out in market share to more usable products, whatever positive attributes your product may otherwise have.

New Information Opportunities

The existence of RSS and Atom and their use on a wide scale bring exciting new information opportunities. The aim, typically, is that you distribute information to significant numbers of people through an information feed. A host of issues then arises. How do you get your site or feed noticed? How do you keep the interest of those who subscribe to your feed? Such issues are similar to those that affect every Web site, but in an information feed the importance of the headline or title is crucial. If that doesn't catch the attention of users as they scan a list of titles, readers will be drawn into viewing your blog entries or Web pages. Getting people to subscribe to the information feed is essential but it's not enough. You need to be able to draw them into the content that lies behind the headlines.

New Information Problems

With increased flows of information one practical problem is how to handle the flood of information that is available to us. Some ubergeeks, Robert Scoble of Microsoft being one (`http://radio.weblogs .com/0001011/`), claim to review several hundred information feeds each day. At the time of writing, Scoble claimed to review 957 feeds each day.

Most users of information feeds won't attempt to follow that number of feeds, since doing so will demand a substantial time commitment each day. Many users will simply not be able to devote the required amount of time to scan and absorb relevant information.

The Need to Keep Up-to-Date

The knowledge that your competitors, in whatever field you work, can have access to essentially unlimited information flows can create a feeling of having to keep up. That pressure to keep up combined with the practical difficulties of absorbing and handling such continual large flows of information can cause so-called information anxiety. Attempting to monitor an unrealistically large number of information feeds can produce so-called *feed fatigue*.

Distractions

Another human factor in information feed management is distraction. If you subscribe to large numbers of feeds it is all too easy to spend time exploring articles or blog posts that are interesting but probably not directly relevant to what you do. When you have large numbers of diverse feed subscriptions, being disciplined and not following distracting posts can be far from easy. This issue is not dissimilar to the reasons that some companies banned employees from surfing the Web during working hours. The costs to companies of distracting information feed items could be substantial.

Summary

Information flows are likely to be the key to efficient information management over the next several years. Delivery of information through RSS and Atom information feeds opens up new opportunities for delivery of content to users. At the same time, the volume of information that's delivered to users through information feeds raises new issues of information management for individuals and for businesses; for example, how to select reliable information and how to cope with large numbers of feed items.

In this chapter you learned that...

❏ RSS and Atom provide a new metaphor of access to Web-based information — instead of going to the information well, user can now access information as easy as piped water in their homes.

❏ Selecting appropriate information is key to efficient information management.

❏ New issues arise as the spectrum of users who will use aggregators moves from being a highly technical audience to a more general one.

Exercise

1. Review the bookmarks and favorites in the browser(s) that you use to see which sites provide information that you might want to include in an information feed. Consider how you might need to filter the information from some sites so that you get the information that you want.