The Importance of RSS in the Exchange of Medical Information

Frankie Dolan and Nancy Shepherd

[1] MedWorm.com
frankie.dolan@medworm.com
[2] Shepherd Research LLC.
nancy@shepherdresearch.com

Abstract. This paper investigates the role of RSS in providing a solution to the problem of medical information overload, speeding up the dissemination of information and improving communications between all those with an interest in health. It compares the exchange and use of medical information on the Internet before and after the use of RSS and also shares a vision for the future, using MedWorm, a medical search engine and RSS newsfeed provider, as an example. The conclusion highlights how RSS has opened a new dimension of information exchange which has the potential to enable giant steps forward in the field of medicine. To realise its full potential, both publishers and users of medical information need to recognise the importance of RSS, ensure thoughtful implementation of RSS feeds to announce publication, and provide for education regarding its everyday use.

1 Introduction

The Internet has enabled access to a wealth of in depth research and medically related information not previously available, but it has also given rise to a new set of problems for todays physician. Medical practitioners are now inundated with information[1], short of time [2] and yet obliged to keep up to date at all times with the very latest developments. Patients are researching their own conditions and often expect their doctors to have expert and recent knowledge on a vast range of topics.

This paper briefly describes RSS (really simple syndication) [3] and investigates the way in which RSS is starting to provide a solution to the problem of medical information overload, speeding up the dissemination of information across the Internet and improving communications between all those with an interest in health.

Since RSS has been around for several years now, its benefits and situations in which it is appropriate are already known. This paper considers its uptake and re-emphasizes its importance, particularly in the field of medicine where the delivery of timely and highly relevant information is essential. It compares the exchange and use of medical information on the Internet before and after the use of RSS and also shares a vision for the future.

This paper uses MedWorm as an example. MedWorm is a medical search engine and RSS newsfeed provider based on data collected from over 7,000 medical RSS feeds [4].

2 Background

2.1 RSS

RSS is a group of simple data formats that are used to announce new data on the Internet. The first version of RSS was created in 1999 although it wasnt until 2005 to 2006 that it started to gain widespread use [5]. The most commonly used format of RSS today is the latest version, RSS 2.0 [6] which is popular due to its simplicity. An alternative to the RSS format is Atom [7]. Applications used to read RSS feeds are commonly known as RSS readers or aggregators, and RSS is also often referred to as aggregation. Many RSS aggregators, including MedWorm, can parse data from any of the different RSS formats. An overview of the advantages and disadvantages of RSS is presented in Table 1.

2.2 Uptake

RSS implementation has been growing steadily within the field of medicine since its introduction in 2006. A few examples of its use are described below.

It is now common practice for all leading medical publishers to include RSS feeds for each of their medical journals. RSS feeds can also be pulled out from any search run via PubMed [8]. The Ebling Library [9] has compiled a feed directory of over 3,000 medical journals categorised by medical specialty. Making extensive use of RSS are medical websites such as Medscape [10], health consumer websites such as WebMD [11], medical news sites such as Medical News Today [12], all leading newspapers and news services such as Reuters [13]. RSS is also being used effectively in several instances to announce industry alerts as seen by the FDA [14] and recently by the WHO [15] and Flu.gov [16] to keep people up to date with the latest swine flu reports. The CDC [17] uses RSS to announce its morbidity and mortality weekly reports and the NHS Library [18] provides RSS feeds for all of its categories.

Medical universities are now often including RSS feeds for their latest news updates and some, such as Duke University [19], are even offering specialised RSS training on request. RSS is frequently covered as a topic in medical research papers and presentations. The University of Helsinki has developed FeedNavigator [20] which is an online medical aggregator. RSS feeds are also now frequently used on medical websites to display updates from other sources and are published automatically on nearly all medical blogs.

2.3 Room for Improvement

The uptake of RSS within the field of medicine has been widespread within a relatively short space of time. However, the majority of medical professionals,

unless they are keen adopters of Internet and the latest information technologies, have yet to learn about or appreciate the benefits of RSS, and have not yet widely adopted the use of personal RSS readers or personalized RSS aggregation services. In addition, there are still many medical websites not posting RSS feeds for key data updates, in particular for other data types that are not classed as medical journals or news.

3 Past and Present

3.1 Data Collection

In the past, without the use of RSS, an online information distribution service would receive data from different sources using various transfer methods such as email, disc, or online download. Incoming data came in various formats, using different types of identifying tags, if at all. Internal programming teams were often required for the writing of data conversion programs to convert the inconsistent data received into a standard format that had been adopted by the data distribution company. The whole process was labour intensive with different publishers requiring customised data conversion programs. Usually the full data in its entirety was copied across to a central database held by the data distribution company.

Today, using RSS, publishers simply update an RSS text file online when new information is released. A data distribution service that uses RSS, such as MedWorm, can automatically check the RSS files regularly for updates. Data is tagged with standard RSS formatting which means that customized data conversion programs are no longer needed, instead only one RSS data parser is required. Providing concise summaries with hyperlinks to the full text of articles means it is no longer necessary to submit data to a central database. The full text of the data can now remain at the point of origin with the RSS feed items acting as pointers.

3.2 Search Engines

The comparison of a search engine based on data collected from web crawling to that of one based on data gathered through syndication may not at first glance seem a valid exercise, since a search engine such as Google indexes the content of web pages themselves, whereas an engine relying on RSS feeds focus only on updates. However, from the end user's perspective, the end result and purpose of the search engine looks similar, and in medicine both are used for the same purpose, to get information to answer questions. End users frequently ask for clarification on why they should use any other search engine to that with which they are familiar. It is therefore important that the informaticist clearly understands the benefits of the RSS populated database if they are to encourage users to make use of such a resource.

Without the use of RSS, a search engine such as Google uses a Web crawler to find information to populate its database by following hyperlinks from one page

to another. All data found is cached and indexed. The possibility of duplicate date is large since webpages can have many different formats and yet contain the same content, making it difficult for duplicate data to be identified within the database. Descriptive data is collected from metatags, which may or may not have been included, and which are prone to manipulation for search engine optimization techniques. It is sometimes difficult to determine any clear title and frequently impossible to identify any summary text. Most limiting of all is the common lack of a distinct publication date. In many cases the most accurate date that the Web crawler can provide will be the date on which the article is discovered. Web crawlers waste a lot of processing power by revisiting pages looking for new content and it can take a long time for new pages to be discovered and indexed. In addition, the amount of data to be included can have an impact on performance and slow down the time it takes to get relevant and up to date content indexed for the audience to find.

Now with RSS, a search engine such as MedWorm uses RSS feeds to populate its database with benefits over that of a Web crawler. All items include a clear title and published date, and often there is also provided a concise summary of the publication. Data is less prone to manipulation by keyword cramming, since if the content does not make sense, the user will unsubscribe. Using RSS, when unformatted text and article summaries instead of full text are indexed, processing is quicker [21]. This enables frequent re-indexing to ensure that queries always return the most up to date information. In addition, processing power is not wasted from the revisiting of pages to look for updates, since it is understood that newly published data will be announced through a new RSS item that will be automatically indexed for inclusion.

For the user of medical information, the greatest advantage that an RSS based search engine has to offer over a standard Web crawled search engine is the speed at which data is indexed and the ability to return search results by date order. Up to date information can often be more important than link popularity in the context of medically related information. Medical professionals and scientific researchers need to place information in the context of a publication date and in relation to other information known previously.

3.3   Subscription to Data

The traditional method of subscribing to data is via email update. Email communications suffer from numerous problems, including spam [22] and misdirection. Providing updates via email subscription is known as push technology, since the publisher sends out the emails and could keep sending them whether the recipient wanted to receive them or not.

RSS subscriptions are known as pull technology since the user chooses which RSS files they wish to receive and how frequently the information will be pulled. If a recipient chooses to unsubscribe, they simply remove the link to the RSS file that they no longer wish to include in their RSS aggregator, which eliminates the problem of spam.

3.4 Sharing Data Between Sites

To share data from one website to another it used to be necessary to either use a customised plug-in or to learn how to program an API that had been designed specifically for that site. Such an approach can be difficult to implement requiring programming skills, or can be limiting.

With RSS it is relatively simple to install one of many RSS widgets that can read RSS files and output the contents in many different formats. Data can be set to cache and refresh as wished to ensure that data is always up to date but bandwidth is saved.

Registered MedWorm users may create customized RSS feeds to place on their own websites. For this, complex queries may be combined and then optionally manually filtered further if need be. Descriptions and homepage links for the RSS feeds can also be set and there is the option to track which feed items have been read from the customized feeds.

4 Looking to the Future

The benefits of RSS as discussed above are well known within the IT community and the prevalence of RSS, particularly within the field of medicine, is growing.

Here we list some areas that have been identified through experience with MedWorm as having potential for future development.

4.1 Clinical Trials

Although clinical trial data may be found online, in particular in the central repository ClinicalTrials.gov [23], it can be difficult for physicians and their patients to keep up to date with clinical trials that are relevant. It is now possible to create an RSS feed for clinical trials for any category via ClinicalTrials.gov and MedWorm is considering how it might integrate such data. If all sites that published data relevant to clinical trials were to include RSS feeds for such data, this would help improve communication.

Using RSS to announce new clinical trials does not solve the problem of complexities of data structure or provide reporting functionality to cater for different terminology or search by eligibility. However, it does provide a very simple way of keeping physicians constantly and broadly aware of new opportunities within their field as they arise. Having clinical trial feeds indexed by MedWorm would enable trial updates to easily reach their target audience, since medical professionals and patients already use MedWorm to get daily updates on topics of their interest, either via visiting the site or through email or RSS subscription.

4.2 Funding Opportunities

Currently many worthy organisations, medical scientists and physicians miss out on vital grants and other funding possibilities, often finding out about them

only after a closing date has passed. Email subscription is commonly used but as described previously this method has its weaknesses. MedWorm has been working closely with the Hope Center for Neurological Disorders at Washington University to ensure key funding RSS feeds are included within MedWorm. Utilizing RSS and MedWorm to announce funding updates would ensure that these opportunities reach their target audience in a timely manner.

4.3 Discussions & Commentaries

Medical discussion forums on the Internet are valuable in their facilitation of building new relationships and sharing knowledge. However, forums are often prone to spam and manipulation. Important knowledge can be buried with time and much repetition can take place with the same questions being asked and answered in multiple places. Forums sometimes lack focus and webmasters are often faced with the problem of ensuring enough traffic to keep on stimulating forums with new discussions over time.

Items from an RSS feed could now be used as a basis for discussion, bringing people together to share ideas on the most recent medical research and news. Any query run on MedWorm to pull out an RSS feed of publication items on a particular topic also returns an RSS feed for the discussion items connected to the publication records returned.

In addition, MedWorm registered users may now use the discussion functionality to pull out RSS feeds on just the items on which they have commented. This functionality can be used for medical professionals to start to build up their own running commentaries on topics of their choosing, which could be displayed on their own websites.

4.4 Social Networking, Digital Journal Clubs and Focus Groups

MedWorm now has the potential to link up users reading the same items to other users reading the same data, should users opt in to share their reading habits. RSS use increases readership to levels where such connections start to become meaningful. The opening of new communication lines leads to new ways of individuals working together. MedWorm already provides a platform for discussion on current articles. RSS use has brought about the concept of running a digital journal club, through the selection of articles to review from the RSS database and then the creation of new RSS feeds for reviewed articles.

Similar to a digital journal club, focus groups could be formed to concentrate on particular illnesses. MedWorm provides a novel meeting ground for different user types, from medical scientist through to patient, to discuss common areas of interest but from differing perspectives.

4.5 Trend Identification

Due to the increased exchange of data and readership that RSS provides, MedWorm is starting to identify trends within the field of medicine, and to provide

RSS feeds for these trends. The value of such statistics comes from the combination of the amount and variety of data being parsed, the MedWorm taxonomy and the wide spread use of MedWorm provided data: on the MedWorm site itself, in the RSS feeds that individuals subscribe to via their own RSS readers, and also on other websites that utilize the MedWorm feeds.

Currently reports include the most frequently read items, by each category as well as overall. All reports are provided with their own RSS feed. In the future it will also be possible to obtain feeds for the most frequently arising topics within the data itself, as well as the ability to see which topics have seen increased activity within a specified time.

4.6 An Added Dimension

Just like the advances that came to society with the building of roads, the invention of the telephone and the development of the Internet, RSS provides a new dimension of communication that could see rapid advances being made in medical knowledge. For the first time ever, publishers all across the globe can easily and almost instantaneously route their data to interested audiences throughout all sectors of society. Users will be able to connect to other people reading the same data and exchange thoughts in real time, and those discussions and thoughts will not be lost but will be fed back into the growing body of evidence in which we can all participate.

5 MedWorm Recommendations

RSS promotion within the field of medicine has seen real progress over the past four years. In order to reach the full potential of RSS, there must be a concerted effort from all in the field of medical informatics to further promote the use of RSS in their setting.

5.1 Integration

The first step towards the RSS ideal is to ensure that all new medical information which is meant to be shared broadly gets submitted effectively into the RSS highway.

Here we discuss some of the issues that the developer often faces when looking at how best to integrate RSS with their website:

Data Types When adding any new medical information to the Internet, it should be a priority to ensure that the information is announced by way of a new RSS item posted to an RSS feed, not restricting RSS use to that of medical journals, blogs and news updates. Other types of information that should also post updates to an RSS feed include: all publications that have not been included in the feed of a journal; events such as conferences, exhibitions, and

training courses; clinical trials; online courses; guidelines; press releases; directory items; funding opportunities; podcasts; alerts and recalls; legal outcomes; medical images and videos.

Using a separate RSS feed for each different type of data, as listed above, enables clearer data definition within MedWorm. When several different feeds exist for a site it can be useful to also have one RSS feed that combines all of your RSS releases into one catch all feed, for those subscribers that want to be kept up to date with everything on your site but dont want to have to subscribe to many different feeds separately. However, other than that the one catch all feed, items should not be duplicated across RSS feeds but rather appear in just one RSS feed to help avoid duplicate items appearing in systems that use your feeds for data population.

Intent to Share Including a lengthy RSS terms and conditions legal document can discourage other websites from using your feeds, which could limit your readership. It is better to include content in your feeds that you are happy to be shared anywhere, anyhow, without the need for legal notices. If there is reluctance to share data within your organization, include a description of the data rather than the data itself.

Content Include a descriptive title and date of release as well as a description wherever possible. Ensure the key words relevant to the item are included somewhere within the feed item, without resorting to 'keyword cramming' that would degrade the quality of the data. A concise description of an article is preferable to the whole article itself. It is also better to include a short paragraph that is of use on its own rather than a half sentence 'teaser' that tells the user little.

Do not include feed items that present the user with nothing more than a subscription page when clicked through. If a subscription page is to be used, add some value to the feed by providing some free information when people click through on a feed item, before logon is necessary. Such an approach provides an incentive for readers to continue to subscribe to your feed even when they do not think they want to become a customer, and will likely result in additional sales over time.

Discovery All RSS feeds should be clearly labelled with an orange RSS logo. Any webpage that has an associated RSS feed on it should also include RSS auto-discovery mark-up in its html header [24] There should be an RSS feed on every home page of every medical website. If there are more than one RSS feeds available for a site, there should always be a link from the homepage to a page listing all the feeds available for that site with descriptions, except in the case where the number of RSS feeds for a site is enormous due to dynamically created feeds from different data categories, as seen on MedWorm.

The addition of new RSS feeds for a site should also be announced via an RSS posting on an existing RSS feed on the same site. It is good practice to also

include an OPML file [25] to list all feeds on a site when the number of feeds is numerous. A suggested standard that could be adopted is to include an OPML file for every site with numerous RSS feeds, that includes the file name and location of the file as /rss.opml, similar to the standard of adding a robots.txt file to a site, as suggested by Jeremy Zawodny back in 2003 [26]. At the time the suggestion was not warmly welcomed, but since then it has become common practice to pass lists of RSS feeds from one system to another using the OPML format, and also the number of RSS feeds offered by any one site has typically increased, so this might now be considered as a logical step forward.

Continuation When redesigning a site, give considerable thought to RSS items that have previously been posted to items on the existing site. Ideally any move of data from one link to another should include with it auto redirections for visitors to the previous data links. You may wish to plan ahead for potential future site redesigns by using dynamic urls in your RSS feeds that can be mapped to different physical urls through script.

5.2 Education

The second step towards the RSS ideal is to ensure that end users are trained and encouraged to use RSS to receive data wherever possible.

This should include the provision of RSS training either within the work setting, at medical school, or in continuing education courses. It may be difficult to change the habits of an older generation when it comes to data retrieval and the integration of new technologies to assist with current awareness, but it should be relatively easy to encourage good habits at the start of their medical careers with the younger generation.

The incorporation of RSS into the medical syllabus at university would give all newly qualified medical professionals a basic understanding of what RSS is and its benefits, and could introduce them to a number of tools and services that would help them to better manage their data and keep abreast of the latest research without the daunting feeling of information overload. Employers and managers should also be educated as to the importance of RSS and requested to implement incentives to get their entire medical workforce familiar with its application. This may include pre-installing RSS software on all computers and handheld devices and ensuring that employees receive training and incentives for its use. As RSS is encouraged and adopted the benefits will become apparent to all.

In order to ensure that RSS updates are posted to the RSS feeds, the process must be included in the submission of any new data that is to be posted online.

6   Conclusion

RSS opens up a new dimension of information exchange that has the potential to enable giant steps forward in the field of medicine. Medical information published without the use of RSS may result in lost opportunities to quickly further

knowledge. For RSS to realise its full potential in the field of medicine, everyone needs to participate in the promotion of its use, by ensuring the provision of RSS feeds to announce the publication of all medically related data and the provision of RSS training wherever possible.

References

1. Glasziou, P.: Information overload: whats behind it, whats beyond it?, http://www.mja.com.au/public/issues/189_02_210708/gla10552_fm.html
2. Greenhalgh, T.: A comparative case study of two models of a clinical informaticist service. Volume 324., http://www.bmj.com/cgi/content/full/324/7336/524 (2002) 524–29
3. Hart, L.: Library 2.0: Rss feeds dynamic uses for special libraries., http://www.sla.org/pdfs/sla2007/hartrssfeeds.pdf
4. MedWorm: http://www.medworm.com
5. Wikipedia: Rss, http://en.wikipedia.org/wiki/RSS
6. Winner, D.: Rss 2.0 at harvard law., http://cyber.law.harvard.edu/rss/rss.html
7. Nottingham, M., Sayer, R.: Rfc 4278: The atom syndication format, http://tools.ietf.org/html/rfc4287 (Dec 2005)
8. NLM: Pubmed, http://www.ncbi.nlm.nih.gov/pubmed
9. WISC: Ebling library, http://ebling.library.wisc.edu/rss/index.cfm
10. WebMD: Medscape, http://www.medscape.com/
11. WebMD: http://www.webmd.com/
12. MediLexicon: Medical news today, http://www.medicalnewstoday.com/
13. Reuters: Reuters: Healthcare, http://www.reuters.com/sectors/healthcare
14. FDA: Fda: Rss feeds, http://www.fda.gov/AboutFDA/ContactFDA/StayInformed/RSSFeeds/default.htm
15. WHO: Who: Rss feeds., http://www.who.int/about/licensing/rss/en/
16. US_government: Flu.gov., http://www.pandemicflu.gov/
17. CDC: Cdc: Rss feeds, http://www2c.cdc.gov/podcasts/rss.asp
18. NHS: Nhs library, http://www.library.nhs.uk/
19. Powers, A.: Duke university medical library: Rss, http://www.mclibrary.duke.edu/training/rss
20. : The university of helsinki: Feednavigator, http://www.terkko.helsinki.fi/feednavigator/
21. Wittenbrink, H.: Rss and atom: Understanding and implementing content feeds and syndication
22. Hayes, B.: Spam, spam, spam, lovely spam. Volume 91(3)., http://www.americanscientist.org/issues/pub/2003/5/spam-spam-spam-lovely-spam (May-June 2003)
23. ClinicalTrials.gov: http://www.clinicaltrials.gov
24. RSS_Advisory_Board: Rss autodiscovery, http://www.rssboard.org/rss-autodiscovery (Nov 2006)
25. OPML: Userland software. opml 2.0, http://www.opml.org/spec2 (Nov 2007)
26. Zarodwny, J.: Rss auto-discovery 2.0, http://jeremy.zawodny.com/blog/archives/000967.html (Sep 2003)

Advantages	Disadvantages
A simple, accepted, widely used standard defining basic elements so that data can be easily shared across different platforms.	Defines only basic elements, so more complex data structures may be lost across platforms. Mal-formatted data and non-standard characters often cause RSS parsing to break.
It acts as an alert to the latest updates, saving the need to crawl the whole web to find the latest information.	Not all websites use RSS and others do not use it effectively. It depends on RSS items being posted to RSS feeds when new information is published.
The simple data structure removes problems related to redundant and poor quality data that is often added for the search engine optimization purposes.	Usually not all data is posted to a feed item, so a search on feed item content may miss important information and keywords are not always included. It does not therefore replace the need for a web crawler.
It can save bandwidth and speed up processing times by passing basic information and summaries of data rather than heavily formatted full text webpages.	It can result in an additional processing burden on a web server when some of the RSS feeds become heavily subscribed to, especially by other websites that may call the RSS feeds very frequently.
The manual collection of RSS feeds ensures the use of only quality data sources.	It is a manual process to find and add newly created RSS feeds to aggregators.
It gives the subscribers more privacy without the use of tracking cookies and scripts to log reader activity.	Readership of RSS items is difficult to track, due to the many different platforms through which data can pass and the lack of cookies and data tracking scripts.
It is known as 'pull' technology which allows the subscriber to decide what they wish to read and to have full control of which sources they no longer wish to subscribe to, hence the avoidance of spam.	RSS is not a push technology, which means that important updates can be missed if users have not subscribed.
The use of RSS encourages publishers to be more generous in the sharing of their data and hence promotes the advance of universal knowledge.	There is uncertainty and concern over the sharing of potentially copyrighted information via RSS.
Due to the passing of simply formatted text, RSS reader enable the user to scan through large amounts of data very quickly.	Formatting may be lost when data is passed through an RSS feed.
It can save RSS subscribers time and help keep them better informed on a topics of their interest.	It is hard be hard to motivate users to learn to use an RSS reader in addition to the use of email and there is the perceived risk of information overload.

Table 1. Advantages and Disadvantages of RSS