

# Addressing the unharmonised metadata of RSS feeds would support research discovery and speed up science



*RSS feeds allow academics and others to keep track of the latest papers to publish in a chosen selection of research journals. They are machine-readable and aggregable, thus presenting a potentially simple solution for promoting content awareness on a large scale. However, **Andreas Pacher** flags a handful of problems which continue to limit the effectiveness of RSS feeds; namely a lack of metadata, a lack of harmonised metadata, and a more general lack of awareness. If efforts are focused on addressing these problems, RSS potentially speeds up science by aiding research discovery for scholars and other publics, representing a positive contribution to scholarly communications as a whole.*

[Keeping track](#) with one's research field may be cumbersome, but this can be mitigated by [research discovery services](#) which aggregate the latest papers of selected journals. To name a voluminous example, [JournalTOCs](#) automatically collects the tables of contents from almost 32,000 journals in 73 subject areas, and offers customisable alert services. In narrower fields, [The Philosophy Paperboy](#) lists the latest papers from 563 philosophy journals in a freely available service initiated by Andrea Raimondi, a PhD student at the University of Nottingham. Finally, my own recently launched project, the [Observatory of International Relations \(OOIR\)](#), makes use of 217 political science journals to automatically enumerate their newest publications. Users of such platforms do not [need to check each journal individually](#) anymore.

What all these services have in common is their underlying technological tool: RSS. [Most academic journals provide RSS feeds](#) with metadata about articles which are updated whenever a new publication appears. What makes them so powerful is that they are machine-readable and aggregable, thus presenting a potentially simple solution for promoting content awareness on a large scale.

Given these advantages, one could assume that RSS plays an indispensable role in scholarly communication. Alas, when I created OOIR, I encountered various problems which can be clustered along the main strands of: (i) lack of metadata; (ii) lack of *harmonised* metadata; and (iii) lack of awareness about RSS feeds in general.

## Metadata in RSS feeds – unharmonised or absent

The technical side of RSS feeds is fairly simple: whenever an academic journal publishes a new article, it automatically adds a few metadata to its RSS feed. A simple code writes “<item>”, followed, for instance, by the article's title inserted between <title> and </title>, its authors between <creator> and </creator>, its link between <doi> and </doi> etc., until all metadata are listed, closing the section with “</item>”. This plain syntax makes them effortlessly readable by programmed scripts – RSS, in fact, stands for “Really Simple Syndication”.

But this simplicity can be deceptive, for minor variations can immediately disrupt the aggregation process. Instead of listing an article's authors in “<creator>”, a publisher could also name them in “<author>”, or somewhere under “<description>” alongside the whole abstract and date of publication (as Elsevier does, making it impossible to retrieve precise author data through a machine). Once RSS feeds do not follow uniform standards, they can cause irritations since automated processes are not always programmed to detect irregularities.

This is why Crossref published a set of [best practices on RSS feeds for scholarly publishers](#) in 2009, to recommend the use of standardised metadata. While some publishers do indeed follow these requirements, there has been no monitoring of this issue, and Crossref seems to have [lost interest in RSS](#) since then.

As a consequence, academic RSS feeds are today beset by inconsistencies across publishers. I already mentioned the multifarious tags which can contain author names, but it gets even more complicated: do they start with the given name or with the family name? And in the latter case, are they separated by a comma or not? There is also no uniform practice regarding multiple authors, and dates are not always formatted according to [W3's recommendation](#) of YYYY-MM-DD.

While RSS feeds are relatively scarce on data, one could automatically fetch more article-level information by inter-operationalising them with other metadata repositories (like [Crossref](#) or [Altmetric](#)). The key to access their opulent treasures – ranging from [citation data](#) to the number of tweets mentioning an article – would be the paper's DOI (digital object identifier). However, this is often hampered by the fact that many RSS feeds link their articles not through their DOI but via the journal homepage's URL.

While such outliers can be adapted to through a bit of (re)programming, what I also encountered is the complete absence of core metadata. Author information, for instance, is left out of Springer Nature journals' RSS feeds.

## Disinterest towards RSS

Most fundamentally, I encountered a general lack of awareness about RSS feeds. A few RSS feeds were discontinued some years ago. Two journals which shared the same publisher accidentally had identical feeds without anyone noticing. Other journals offer RSS feeds, but fail to link them on their websites. A few publishers occasionally update their feed with new articles hours before they are accessible through their DOIs (something I have observed of articles from Wiley, Springer Nature, and Taylor & Francis). Finally, at OOIR's launch, 33 of 217 political science journals still had no RSS feed at all (this number is now down to 30 after some awareness-raising).

Ironically, there is also the opposite problem of journals having three or more feeds. Oxford University Press journals, for instance, have RSS feeds respectively for "Latest Issue", "Advance Articles", "Open Access", and, optionally, for "Editor's Choice" (see, for example, [The Chinese Journal of International Politics](#)). One could wonder whether this quadruplication really addresses a popular demand.

The lack of awareness does not only pertain to the journals and publishers, but also to the broader open science movement. Top scientometric journals have not discussed RSS in years; Crossref's work on RSS basically ended in 2009; the venerable DOAJ contains a list of 16 [best practices for open access journals](#), ranging from online marketing to aspects of scholarly communication, but does not mention RSS – hence the strange outcome that almost none of the few open access political science journals have RSS feeds. Only JournalTOCs occasionally lobbies for a more consistent use of metadata, leading to their recent success of [integrating clear labels identifying open access articles in RSS feeds](#) among some publishers.

Greater awareness about RSS could lead to a more conscious utilisation of metadata and to better services harnessing these machine-readable tools for the benefit of the research community and its stakeholders. As RSS potentially speeds up science by aiding research discovery for scholars and other publics, this would be very positive to scholarly communication as a whole.

*Featured image credit: [Clicker-Free-Vector-Images](#), via Pixabay (licensed under a [CC0 1.0](#) license).*

*Note: This article gives the views of the author, and not the position of the LSE Impact Blog, nor of the London School of Economics. Please review our [comments policy](#) if you have any concerns on posting a comment below.*

## About the author

**Andreas Pacher** initiated the [Observatory of International Relations \(OOIR\)](#), a website which tracks political science journals to continually list their latest papers. He studied law and international relations in Vienna (University of Vienna), Paris (SciencesPo), and Shanghai (Fudan University), and works in the public sector of the Republic of Austria. His research interests revolve around the everyday practice of diplomacy. You can follow OOIR at [@ObserveIR](#).