

# RightFind™ XML for Mining

## A Q&A with Copyright Clearance Center (CCC) (RightsDirect's parent company)

Copyright Clearance Center's (CCC) RightFind™ XML for Mining solution enables researchers to identify and download full-text article collections including content from multiple publishers in XML format and import those results into their preferred text mining software.

Text miners can use content that their organisations already license, as well as discover and pay for access to additional relevant content they might not have otherwise known about.

At launch in mid-2015, [RightFind XML for Mining](#) includes content from more than 30 publishers including [Wiley](#), [BMJ](#), the [Royal Society of Chemistry](#), [Taylor](#)

[& Francis](#), [SAGE](#), [Cambridge University Press](#), [American Diabetes Association](#), [American Society for Nutrition](#), [Future Medicine](#) and more. CCC is actively growing the pool of participating publishers (both major brands and smaller society publishers) with an initial focus on the life sciences.

FreePint interviewed Lauren Tulloch, Director of Corporate Products and Services for CCC, about the development process of the new offering and the challenges it solves for information professionals and publishers alike.

### Incubation of an Idea

**FreePint: Tell me about when you first started working on the concept for RightFind XML for Mining**

**Tulloch:** It initially started through conversations with both our corporate licensees and our publishers.

Many corporate researchers were already mining abstracts from sources like PubMed. But they knew they were missing out on

important facts, information, and connections that could only be found in the full-text. In addition, they were trying to get rights to mine full-text within their publisher contracts and were sometimes encountering resistance. Even when they were able to obtain those rights, they often didn't have an easy way to get content in a mineable format.

Publishers told us they were getting inquiries about making their content available for text mining purposes, and they wanted help managing these requests.

In the summer of 2012, CCC hosted a meeting in Amsterdam, which included a selection of publishers and researchers engaged in text mining within corporate environments. The purpose was to identify the roadblocks and the opportunities relating to text mining. Coming out of that meeting was the idea that became our pilot program.

Both groups — the publishers and corporate researchers — were interested in seeing a centralized solution from CCC. We weren't trying to create text-mining



**Lauren Tulloch**  
Director of Corporate Products and Services  
Copyright Clearance Center

software; we were trying to meet the need for a single platform where text miners could access mineable, licensed content across lots of publishers.

The next step was to construct a pilot with a handful of publishers and corporate users to test the idea from a technology perspective as well as from a business model and licensing perspective.

## A Pilot Project Takes Off

**FreePint: From the initial meeting to the point of product launch today, what were some of the key milestones along the way?**

**Tulloch:** We engaged in about nine months of research and prep work. This involved identifying publishers and users who wanted to participate as well as laying the technology ground work. We kicked off the pilot in 2013, a little over a year after we had that initial meeting.

We had about half dozen folks on either side. First, we needed to get a critical mass of content from our publishers. That was actually more time-consuming than you might realize.

In parallel, we had the usual software development milestones, the first of which was getting the software to the point where pilot participants could test it and provide input.

Our pilot testers gave us some really good feedback, from which we made a number of tweaks to the solution itself. Through the process, we came to understand much better the needs of the end user. We understood first-hand their challenges with trying to create corpora of content ready for text mining.

**FreePint: Talk a little bit more about the challenge of getting full-text XML articles for text mining**

**Tulloch:** The typical process is a substantial burden for corporate researchers, and this is where CCC comes in to help.

Today, corporate researchers have to obtain permission to mine subscribed content. Many publishers don't include text mining licenses within their subscription agreements, nor do they have a business line dedicated to licensing text mining activity. Even finding the right person to talk to can be a challenge.

Once they get permission, they have to get the actual content. The preferred format for mining is full-text XML, but not every publisher can provide that easily. So, researchers may be forced to convert PDFs from their own holdings, which results in significant loss of fidelity. The conversion process sometimes introduces errors and often removes tags that indicate sections of the article, such as introduction, conclusion, methods and materials, etc.

**FreePint: Yes, we've heard a lot about those challenges as well. Pulling together content from many publishers in a single platform could really accelerate the process and reduce upfront manual effort associated with text mining.**

**Tulloch:** Exactly.

**FreePint: Can you describe the benefits to the publishers? Why do they want to participate in the service?**

**Tulloch:** First and most importantly, this service reinforces the importance of subscriptions. With our service, companies recognize an additional value in being subscribed. That said, large R&D organisations subscribe to an enormous amount of content, but they can't license everything. With our service, customers can discover content that they're not subscribed to. They may choose to purchase the content within our solution or upgrade to a subscription.

Our solution enables researchers to build a corpus of full-text content from a wide range of publishers using a single platform. Text miners get visibility into what's out there that might yield important results, whether or not they subscribe to it today.

**FreePint: Does the solution take into account the customer's current licensing arrangements?**

**Tulloch:** Yes, the system automatically checks the subscription status for all the results within a project based

on the particular customer's subscriptions. Users can immediately download subscribed content. Where they don't have a subscription, they have the option to pay to obtain a full-text copy of the XML article or to download only the abstract and metadata that would normally be available publicly.

**FreePint: Can you describe the business model for the solution?**

**Tulloch:** The solution is offered for an annual license fee, which covers the following:

- Use of the software to search across CCC's entire repertoire of participating titles
- The downloading of a set number of XML articles
- Permission to use the downloaded XML content for internal text mining projects.

Publishers are paid based on how often their content is used within the solution. They also have the opportunity to sell unsubscribed articles that text miners discover in the course of creating a project.

**FreePint: We talked about the feedback from users and publishers, and you mentioned that CCC is not in the business of providing the text mining technology itself. Can you tell me a little bit about the technology companies your customers are working with? How did you start to bring them into the conversation?**

**Tulloch:** As we were researching the marketplace for text mining, we paid close attention to which

software solutions were being mentioned most often by our customers. That's why we started a partnership with Linguamatics. They were most often mentioned by the people in our pilot group, as well as by the additional customers we spoke to as part of our research.

That said, our solution can be used with any commercially available text mining solution, and we'll be working on integrations with other companies as well.

## Supporting Education and Shifting Perceptions

**FreePint: What's the level of awareness among corporate information professionals related to text mining at their organisations?**

**Tulloch:** It really depends on the organization and the individual information managers. Some of the customers we talk with demonstrate a lot of alignment between information managers and the researchers who are driving text mining projects. Many information managers are highly engaged with text mining and working hard on the issues of access and licensing. Then there are others where text mining is run out of a completely independent department. They rarely touch base with their corporate librarians.

One of the ways that our users have told us that we're helping them is simply by educating them about what's happening out there. The information professionals who are more involved in text mining at

their organisations are finding that this is really a career opportunity for them.

**FreePint: There seems to be both a lot of excitement around it, as well as a lot of anxiety.**

**Tulloch:** The folks that aren't yet involved want to become more educated so they can provide value in this important and growing area.

**FreePint: What kinds of activities are you doing to provide that education?**

**Tulloch:** We have already conducted a number of educational webinars and produced a white paper. Throughout the upcoming year we will continue to offer programs to build awareness.

We also have advisory groups here at CCC and RightsDirect for groups of customers. This is a topic that's been on the agenda for our annual meeting for the last three years, and we continue to have an active dialogue with our customers on this topic.

## The Next Phase of Development

**FreePint: Now that you've launched, what's next for development?**

**Tulloch:** As you know, our initial focus for the solution was developing a central source for biomedical content to serve customers in the life sciences industry.

One of our focus areas throughout the next year will be identifying additional industries and related content to broaden our offering.

We're also developing new features including the ability to get alerts and updates for projects that have been completed. This addresses researchers' desire to be on top of the latest published research as soon as it's available.

**FreePint: We'll look forward to taking a closer look. Lauren, thank you very much for your time today.**

### About Lauren Tulloch

Lauren Tulloch is the Director of Corporate Products and Services for CCC. In this role, Lauren oversees product development for CCC's corporate licences and software products. She has been with CCC for three years. Prior to joining CCC, Lauren spent 12 years at HCPro, a publishing and training company focused on the healthcare market. As Group Publisher and Editorial Director there, she managed a team of 35 content developers and educators and oversaw product development for a wide range of education and compliance solutions. Lauren began her career as a newspaper reporter and editor in the greater Boston area. She attended Boston University where she earned a Bachelor of Science in journalism and Bachelor of Arts in political science.

This article is part of the FreePint Topic Series "[All About Usage — From Content Sharing to Data Mining](#)," register for free updates and more.

## GET MORE INFORMATION

To learn more about RightFind XML for Mining, please visit [www.rightsdirect.com/rightfind-xml-for-mining/](http://www.rightsdirect.com/rightfind-xml-for-mining/)