



**Better Data is Better Publishing**  
**Interview with Chuck Hemenway, Copyright Clearance Center**

**For podcast release**  
**Monday, September 30, 2019**

**KENNEALLY:** Data-driven solutions for publishing can lead to improvement in many areas, from manuscript workflow to peer review, audience development to market reach. Yet publishing has a data problem – a deficit of accurate, relevant data necessary to manage in a world of change.

Welcome to Copyright Clearance Center’s podcast series. I’m Christopher Kenneally for Beyond the Book. Publishing is a profession notorious for relying on gut – business and editorial decisions made on instinct and intuition. But it’s 2019, and it’s time for your gut to retire. Data-driven decision-making, DDDM, is considered a serious discipline today in academia and in organizations. DDDM may not guarantee for business success, but as use cases demonstrate, better data and better use of data can lead to better publishing.

At Frankfurt Book Fair on Thursday, October 17<sup>th</sup>, CCC’s Chuck Hemenway offers insights on why better data is better publishing. He joins me now with a preview. Welcome back to Beyond the Book, Chuck.

**HEMENWAY:** Glad to be here.

**KENNEALLY:** Well, you and I both attend our share – our fair share of trade shows, including the world’s largest publishing trade show, the Frankfurt Book Fair, where we look forward to seeing many of our listeners in just a couple of weeks. And we have been hearing a lot at those programs about big data over the last few years. You’re trying to move us, I think, to the next level, talking about better data. But we should start by setting the scene here and recognize that data, at least when it comes to publishing, is a real catchall term. So what are we talking about?

**HEMENWAY:** In most cases, we’re talking about either the content that we all consume and need or the metadata that describes it. In many cases, those two have been lumped into one box and labeled data, but they’re distinctively different. But I think for our purposes, we’re going to talk about them in a way that will make them seem very similar.



## BEYOND the BOOK

A Copyright Clearance Center Podcast

**KENNEALLY:** Metadata is the data about data. It's the descriptions. It's the signifiers for all of these various aspects of publishing. And it can be used in marketing. It can be used in a variety of workflow activities – in peer review when it comes to scholarly publishing. Metadata really is an almost inexhaustible area of data itself.

**HEMENWAY:** Yeah, and I think to your point, it is almost inexhaustible. There are categories of metadata which serve different purposes, whether that be for describing the content itself, tying it back to apparent work, or describing the contributors to the content – the authors, their institutions and affiliations are all manner of metadata, and they all have their purpose.

So when I talk with publishers, sometimes I find it stunning that there's a disinterest – sometimes – in talking about data, talking about metadata. I think for a lot of publishers, they may feel that it's not for them. It's something that commercial publishers worry about. But it's something that all publishers should be taking seriously and from time to time auditing and putting on their calendar for review and making it a real core part of their process and their planning, more accurately.

**KENNEALLY:** Because it can really make a difference not only in workflow, in getting the product to market, so to speak, but once the product is there, understanding better how your customers are using it.

**HEMENWAY:** Well, that's absolutely true. First of all, they've got to find it, right? And in many cases, it's well documented that products that reach market with a substandard descriptive metadata do poorly. They're not borrowed as much in interlibrary loan if it's books. They're not discovered well if they're journal products. The sales numbers prove that out and back that up. So I think regardless of your stripes as a publisher, if your intentions are purely good and non-commercial, then discoverability, dissemination, all of those things are front of mind. Those all require the same care and rigor and effort as the commercial publishers apply. So whether you're commercial, whether you're non-commercial, content products with better metadata go further.

**KENNEALLY:** And beyond this excuse that, oh, that's for the commercial publishers, the other excuse – well, maybe sort of the rationalization for not attacking this problem – is just it can be fairly overwhelming when you start to think about it.

**HEMENWAY:** Absolutely. As a topic, it's pretty dizzying. It's the kind of thing that can put people to sleep at a cocktail party. Trust me, it's happened. (laughter) But when we look at it in increments – we don't need to build the Great Wall of China



## BEYOND the BOOK

A Copyright Clearance Center Podcast

all in one week, right? Publishers can carve it up into segments and look at it over time. How are we managing our metadata for this purpose or for that purpose? I think regardless of business models, folks want to get the discoverability questions answered early, because that's key. To be found is key. So they can break this down into very small, manageable pieces that they can address over time.

**KENNEALLY:** It's a great point. So you're saying prioritize first and then kind of draw some boundaries around the role that data is going to play for a specific objective.

**HEMENWAY:** Absolutely. They need to set out clear goals over time – what do we want to accomplish? What do we want to get from our metadata now? What do we want to get in 12 months, 18 months, 36 months? Some of those goals and aspirations may require more effort than others. But we've seen publishers achieve incredible results by adding one or two metadata items to their product suites. Even if it's just the cover of their book, even if it's just a simple affiliation for a journal article, it has incredible downstream impact. So understanding a publisher's business, understanding their goals and aspirations, can definitely shape a recommendation about what to do first, what to start with. Because obviously high impact and results are what everyone is looking for, regardless of business model.

**KENNEALLY:** When you give our – sorry.

Well, Chuck, when you give your presentation at the Frankfurt Book Fair on Thursday of the book fair week, you're going to be sharing some insights based on experiences speaking with publishers of all types. Just share with us what the situation is when you go in to talk to a publisher. What do you see behind the curtain or under the carpet?

**HEMENWAY:** I think I see a lot of misunderstanding about the role and value of metadata, whether it's books or journals publishers. Some distinct examples would be journals publishers, now more than ever in the age of transforming business models or journals flipping to open access, the ability to collect quality metadata about the authors, their affiliations, their sponsors in publishing – putting a small amount of effort into that part of the production process pays off in spades downstream in terms of keeping that publisher out of Dutch with funders and institutions and everybody else involved in that endeavor of publishing. So a small effort at the headwaters of the process has huge dividends in the tidewater. And book publishers similarly – we're seeing real data around taking good care with two or three metadata elements really having a tremendous impact about their discoverability and their sales numbers.



My takeaway from these meetings and these discussions is that this thing is fairly unknown to many executives in publishing. They don't understand the relationship that metadata has with their products and with their success and with their profit and loss statements. And to the degree that we can shine some light on that and get folks to think about these things as a strategic imperative and not just a damned mechanical thing that needs to be addressed when it breaks, but rather to see it as a strategic imperative to have their metadata strategy sewn up properly and have a plan of attack, we think that that's going to pay off for them handsomely over time.

**KENNEALLY:** And if you think about scholarly publishing in particular, Chuck, the business models there are under a lot of stress, a lot of pressure, particularly coming from the drive to make open access a sort of ubiquitous reality. One of the ways that publishers are responding to those demands from the funders and from the institutions is with so-called transformative agreements to move from subscription models to some kind of read and publish or publish and read, depending on who you talk to, that allows for institutions to have access to materials, but also to see their scholars published in various journals. I would imagine data is going to play a role in sorting all that out for publishers. If they're going to make that a sustainable business model moving forward, they're going to need data about authors, about institutions, and all the rest.

**HEMENWAY:** Truly. And the needs, I think, will only deepen in the journals space. Right now, many publishers are playing catch-up trying to get good, clean affiliation data about their number one – their primary or corresponding author to figure out who are they faculty for, and further from that, who is sponsoring the research? So many publishers are playing catch-up rather quickly, and they're getting those items battened down and handled.

But over time, we're going to see that there will be a desire for information about authors number two through X. Who are they on faculty with? How do we create machine-readable identifiers that give us that data, or how do we add that to the record? Because there is an existing need now – folks would like to understand that better – but I think in three to five years, it will be demanded that we understand authors two through X and be able to create maps. Where does this author fit in the scholarly landscape? Where does this paper fit in the scholarly landscape? At the paper level, that's easier to do. But when you're author number five on a paper, there's not a lot in the way of mechanics today, and there's a desire for that sort of understanding of the space.



# BEYOND the BOOK

A Copyright Clearance Center Podcast

KENNEALLY: Well, if publishers want to be there in two or three years, they need to be there on Thursday, October 17<sup>th</sup>, at the Frankfurt Book Fair in Hall 4.2, joining Chuck Hemenway for his presentation Better (laughter) – there's the Bostonian.

Well, if publishers want to be there in two or three years, Chuck, they need to be there on Thursday, October 17<sup>th</sup>, to join you for your presentation in Hall 4.2, Better Data is Better Publishing (and Maybe Better Science, As Well). Chuck Hemenway from Copyright Clearance Center, thanks so much for joining me, and we look forward to seeing you in Frankfurt.

HEMENWAY: Glad to do it.

KENNEALLY: Beyond the Book is produced by Copyright Clearance Center. Our co-producer and recording engineer is Jeremy Brieske of Burst Marketing. Subscribe to the program wherever you go for podcasts and follow us on Twitter and Facebook. The complete Beyond the Book podcast archive is available at [beyondthebook.com](http://beyondthebook.com). I'm Christopher Kenneally. Thanks for listening and join us again soon on CCC's Beyond the Book.

END OF FILE