The Future of Research Information: Open, Connected, Seamless

Annette Thomas
Clarivate Analytics, Charlotte.Sullivan@clarivate.com

Follow this and additional works at: <https://docs.lib.purdue.edu/charleston>

Part of the [Library and Information Science Commons](#)

An indexed, print copy of the Proceedings is also available for purchase at:

<http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

Annette Thomas, "The Future of Research Information: Open, Connected, Seamless" (2018). *Proceedings of the Charleston Library Conference*.
<http://dx.doi.org/https://doi.org/10.5703/1288284317000>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

The Future of Research Information: Open, Connected, Seamless

Annette Thomas, CEO, Clarivate Analytics

The following is a transcription of a live presentation given at the 2018 Charleston Conference by Annette Thomas, Chief Executive Officer, Scientific & Academic Research, Clarivate Analytics.

Thank you, Anthony. Good morning, everyone. And I wanted to thank the organizers for inviting me to speak at this truly unique event. It's been six years since I last spoke and it's a pleasure to be back, especially at such a pivotal time in our industry.

Most of us here are information professionals of one kind or another and these days almost every large organization has an information director, a chief information officer, or some sort of head honcho for information, and they are no longer just the people running the computer network. They are helping to shape organizations and strategies in order to make the best possible use of the 21st century's most valuable asset: information. Of course, universities have always had information experts, though they are usually called librarians. Sometimes it seems as if the rest of the world has been playing catch-up with academia when it comes to appreciating the central role of information in our lives. But this isn't to say that we've all got it nailed. We still have a lot to learn about how research could work better than it does today. This is important, not only for us and all of our organizations, but it's important for the whole world because what's at stake is the future of research and research enriches us all. It enriches us intellectually, culturally, and economically.

In my talk this morning, I'd like to talk about some ways that we can make what we do better. But first I wanted to briefly describe some of the current problems that we face in our industry, and I'm not talking about flat library budgets, important though they are, and I'm not talking about battles over copyrights and technological disruptions. All of these problems are certainly important, but to consider them the most important things would be too narrow-minded. I'm talking about even bigger challenges faced by the whole research enterprise.

Over the last few decades, universities have enjoyed unprecedented growth. They've risen in number from about 500 worldwide just after the Second World War to more than 10,000 today. Student

numbers have increased similarly. In the UK nearly half of all students now go on to higher education, a roughly tenfold increase since the 1960s. Around 70% of U.S. high school students go on to college. In the developed world, university enrollment has become a rite of passage. But while universities should be basking in the glow of this remarkable success, many universities instead feel a sense of crisis. This is brought about by three fundamental questions that are being asked more and more. Who are universities for? Should every student aspire to attend or would eliminating elitism from higher education miss the point of it? Who should pay for universities, the current students, their parents, of which I'm one, I have four children? Current taxpayers? Future taxpayers? Past and present benefactors, or some combination of these? And above all, what are universities for? Are they for conducting taxpayer-funded research? Are they for training the next generation of professors? For teaching employable skills? For supplying medical care? For providing infrastructure for the humanities and the arts? For fielding sports teams or even acting as landlords and custodians of real estate? Which, if any of these roles, would be better left to others? But not all of the problems in our industry are on the educational or institutional side. Researchers also face their own set of challenges. How should research respond to political hostility? What can be done to restore the social standing of experts? How can research inform politics without descending into partisanship? And how can research funding avoid becoming a victim of political infighting?

Should research serve the economy or human culture? What is the true purpose of research, and how can we measure and maximize the desired outcomes? How can we best demonstrate and communicate the benefits? And finally, why is so much research impossible to reproduce? What more can be done to counteract the inevitable biases of individual researchers? How can we reward truly robust research, and what incentives can we provide for uncovering invalid findings?

These are all important topics, but let me dwell for a few minutes on the last one. The popular image of research, at least among researchers themselves, has been of a process that transcends human biases

by providing a series of checks and balances through processes like peer review, editorial selection, competitive grant funding, tenure committees, and so on, that reward convincing, novel findings and the researchers who produce them. To a large extent this is true. Witness the long track record of progress over the decades and even centuries, but it's becoming increasingly clear that important aspects of the system are broken. The psychologist Brian Nosek, among others, has shown that researchers routinely fall prey to cognitive biases. Rather than reasoning based on data, they rationalize their preexisting beliefs by selective interpretation of their results. If you believe Karl Popper, scientists are supposed to spend their days trying to prove their pet theories are wrong and the more they fail, the stronger those theories become. But in practice this rarely happens, and in a sense this means that researchers are no different from the rest of us. Surely it is the scientific process itself that helps us to overcome these individual weaknesses and biases but on the contrary. In some ways this system that we all inhabit exacerbates them.

John Ioannidis's infamous paper "Why Most Published Research Findings Are False" showed that, at least in biomedical research, using current approaches as statistical thresholds for significance, we should expect most research findings to be wrong, not some or even many but most, and this isn't unique to biomedicine or even the natural sciences. In more recent work, Ioannidis and his colleagues have reported comparable problems in economics, for example, and there's no reason to think the problem stops there. And given that everyone in this room in some way is involved in that research ecosystem, this important fact makes me at least pause.

Next, I'd like to talk about the world that I inhabited for so many years: journal publishing. Why publish so much that is read by so few? The ever-increasing rates at which papers are published aren't just a sign of higher rates of discovery. They also reflect the fact that researchers and publishers alike have become hooked on a "publish or perish" model that values quantity over quality. The impulse to share is good and even central to science, but the main incentive is to publish in order to satisfy funders' requirements and to extend bibliographies rather than to genuinely contribute to the pool of human knowledge. That's why we get "salami slicing" of research into minimal publishable units, as we used to call them, and vast numbers of papers that are referenced or even perhaps read by literally no one. Of course,

distinguishing between seminal and inconsequential work was a central problem that citation analysis was invented to solve, and it will remain a vital method to achieve this, but we can and should go further. And I'll return to this point in just a moment.

Why publish only positive results? In a way this is an opposite of the first problem, that we publish a lot where in some respects, if we are too selective by publishing only positive or surprising results, we contribute to publication bias, which is one of the key reasons that the system embraces so many misleading conclusions. To give a simple illustrative example, if a given experiment has a one in 100 chance of giving the wrong, though interesting and surprising, result and 100 labs try it, then surely the one lab with the surprising but wrong result shouldn't be the one to publish. We are currently putting too big an emphasis on surprise and novelty when we should be rewarding other less sexy but equally important attributes of good research.

So, how can we be part of the solution rather than part of the problem? And this is the key question. The research enterprise is a monumentally successful human creation and all of us involved in it can take some credit for that, but in certain important ways it is broken, which poses a challenge for all of us, whether we're curators, disseminators, and custodians of research information, but it's an opportunity, too. The coming decades could be a golden age for research and the world has never had more talent, more knowledge, or more wealth. And yet even that can sometimes feel like a barrier. Just because technology allows us to do almost anything doesn't mean that we should try to do everything. What principles can we follow to differentiate between the good, the bad, and the utter waste of time?

Obviously, I don't have all the answers, but I'd like to share some thoughts on this. Over the last 25 years, the Web has transformed human society including research, but it was also a product of research, born of the early academic Internet and the foundational work of Tim Berners-Lee, then at CERN. The worlds of research and technology have always been highly connected and the Web is a wonderful example of that. It can also serve as a source of inspiration as we think about how to meet our current challenges. The explosive success of the Web was in large part thanks to three core principles that can also guide us as we build the systems and organizations, the products and services that will support research in the 21st century: connectedness, openness, and seamlessness.

First, connectedness. In the age of data, where we currently sit, we have unparalleled opportunities to construct views, maps, and pictures of information in ways that no previous generation has enjoyed. David Warlock, who I'm sure many of you know, recently reminded me of a quote by Marvin Minsky, which is, "The surprising thing about great libraries of today is that the books are unable to speak to each other." And this serves as a reminder that in another generation much new knowledge can be generated by machine intelligence interactions within existing pools of knowledge. It also reminds us that nothing in research stands on its own. Every paper cites its predecessors and every researcher has collaborators. Every new discovery is connected to an array of associated facts. The genius of the Web was to realize that knowledge itself is a network and that we can represent and navigate it more effectively by creating a virtual space that mimics this. In a sense, hyperlinks existed in people's minds, but this is the first time in human history that they've been clickable. Just as Eugene Garfield realized that counting citations was a good way to measure the importance of a piece of research, Google realized that counting hyperlinks was a remarkably effective way to gauge the importance of a webpage. Google's ranking algorithms now consider hundreds of factors including the particular interest of the person doing the search. But counting links still plays a part because it fits so well with the way the Web works. It was connectedness that gave rise to the Science Citation Index, to the Web of Science, and even the Journal Impact Factor. There is knowledge in understanding connectedness. The Journal Impact Factor quite rightly receives a lot of attention, both positive and negative. But, when used correctly, it remains an important measure of assessment of journals because it uses connectedness in a standardized way to tell us something important about the journal and its discipline and how both are evolving. We also need to be open to a much wider range of indicators beyond citations and a wider range of contributions beyond papers. Researchers are much more than the papers they publish, and the papers they publish contain much more information and knowledge than just the citations. By looking beyond this, it will help us to create a healthier research ecosystem and reduce some of the perverse incentives that lead to the kinds of biases I've recently described. It should also help to address the misuse of metrics and indicators, including the Journal Impact Factor, because we still have a wide range of indicators upon which to draw. There are many efforts underway to address this across the industry, including at Clarivate, and

this is exactly the rationale behind the reestablishment of the Institute for Scientific Information earlier this year. Unlike the Google search algorithm, however, it is vital that research indicators are both transparent and easy to understand. Without this they won't gain the necessary credibility and confidence and they won't deserve to be widely adopted.

And this relates to my next point: openness. At its foundations, the Web is an open system. The underlying protocols are public goods. Anyone can connect to the Internet, register a domain, and set up a website. There've always been and continue to be attempts to control and censor the online world, some of them disturbingly effective, but if we compare the world today with that of a few decades ago it's clear that we've made huge progress, even if we sometimes take two steps forward only to take one step back. Research, too, is founded on the principle of openness, and I don't just mean open access publishing, important though that is. Openness is about transparency. It's about transparency between individuals and transparency between organizations, including service providers like Clarivate. It's about sharing your insights, showing your work, and being willing to take blame as well as credit. It's about a generosity of spirit and a healthy degree of humility. It's about recognizing that important though our work is to each of us individually, it's really just a small contribution to a much greater collective endeavor. Above all, it's about acknowledging that research is not something that any of us does individually but it's something that we all do together.

Clarivate will play its part, especially through the newly reconstituted ISI. But I hope that ISI will contribute to all kinds of developments in research and assessment, from cataloging and curation to new kinds of indicators and incentives, and that these will foster new approaches to research without some of the shortcomings I've just described. But openness is also about being receptive to new uses of content and data. If we want researchers to publish negative results and to share experimental data, while at the same time not drowning in the information deluge, then we need to do a better job in fostering technologies that can help them do this. For each human interface to our products, we should also have a computer interface, an API, and every time we think about how our articles or records might be consumed individually by human readers, we should also consider how they should be consumed in bulk by computers. Because that's the world that we live in now, and that's where so much power remains to

be unlocked, and that is how research information will be used in the future. And none of this is to underestimate the value of human judgment and curation. Those skills will always be essential to our work. But by collaborating closely with computers, we will be able to achieve so much more.

To give one example that I'm familiar with, we've recently expanded the Web of Science API's, which allow our partners to directly access new sources of data and integrate with their own information. We're only at the beginning of this journey, and there is a lot more work to be done, and there are a lot of exciting developments in this API space across the industry. But I believe that we are taking a step as a group in the right direction. It's this kind of interoperability that is the future of online applications, which is to say that it is the future of pretty much everything that we do.

And this brings me to the third important characteristic of the Web. Although it's decentralized, it's also remarkably seamless. Jumping from one website to another doesn't feel any different from moving between pages on the same website. Users rarely need to think about where in the world any piece of content is located. Even more impressively, the Web can seamlessly blend text, images and video, data and code into rich interactive publications or applications of a kind that aren't possible in any other medium. We are still at the beginning of this evolution from static to interactive content and we all have a lot to learn, but this is the future of knowledge dissemination. Inspired by the example of the Web, we need to become seamless, too. Of course that means our own products and services should work well with each other, but that's not enough. We also need to achieve seamlessness between organizations. And as you all know better than I, seamlessness also applies to the library, which isn't stand-alone, but increasingly working in a seamless way across the institutions, often with a view to supporting communication between scholars, assessing the research and communication of research discoveries, and so the products and services that we provide need to support all of you in this broadest of remits, and they need to work in a seamless and interoperable way.

One area that I and many of us are focused on is creating a seamless way for researchers to access content using technologies to simplify the hurdles and barriers that exist to accessing journal articles. This is important because these barriers are frustrating for researchers and they slow down progress for

all of us. Our mission is to help people make sense of the vital but complex world of research, taking an unbiased and independent approach without fear and without favor. Above all, we want to understand and support the various participants from institutions, funders and governments, to publishers and corporations, but ultimately and most importantly we want to support the researchers themselves. If only the world of research looked as simple as this it would be an easier task, but as we all know it's nothing like this but it's more like this [*showing diagram*]. Like the Web itself, research is constantly moving. It's a constantly moving maelstrom of ideas, of interactions and innovations, but also like the Web, that's precisely its strength and the reason that it's so productive.

So, I hope I've managed to explain some of the lessons that I think we can take from the huge success of the World Wide Web. These are lessons that its creators, many of whom are researchers, often inherited from academia. The world of the Web and research are not just mutually reinforcing. They're inextricably linked. So, it was perhaps natural a year on from my appointment to Clarivate to think about how I would use these characteristics of the Web: connectedness, seamlessness, and openness, to think about what we would do and how we can make our contribution. And just to give you a little bit of a preview of where we're headed, you can see that now in our name, the Web of Science Group. So, this won't formally be announced until January, but for me the Web of Science is the Web of Science in a connected, seamless, and open way. It's about bringing what we do together but also making it interoperable and more flexible with what you're doing.

But where does research go next? No one organization, let alone one individual, has all the answers. So, what I've tried to cover in this talk isn't by any means a comprehensive solution. What I've tried to do is to provide some pointers, some ideas, some ways of thinking that I hope can move us collectively in the right direction and allow those of us that are working alongside the academic research and scholarly community to find our true purpose in enabling excellence going forward, to enable excellence in a way that maximizes the impact that research has on society at large.

And finally I'd like to end on a personal note. As Anthony mentioned, for many years I was a publisher, but before that I was an editor and before that I was a researcher, a cell biologist to be exact. I'm

lucky to have been able to maintain my academic ties over the years and even most recently as a trustee at Yale. These and other interactions with universities all over the world provide me with constant reminders of how much has changed.

Scientific communication is changing. We've been talking about this for decades, but I think we are truly at a pivotal point. New tools, new approaches are being explored by new players, by publishers small and large, by librarians, funders, and organizations like the Web of Science Group involved in workflow data analytics and assessment. And yet in the last 15 years perhaps the most significant trend in our community has been the increasing consolidation and opportunities of scale including so-called vertical integration, but in my view this is no answer to the challenges described. Arguably, it compounds the problem and it compounds some of the challenges as what we need most of all is not size and conformity. What we need most of all

is talent. What we need is diversity of thinking and creative approaches, which are not always comfortable partners with a unified corporate view. Ours is an industry built on the vision and daring of people like Eugene Garfield. He was a true pioneer. He and his colleagues brought order to a chaotic world of research information. Now we need to rise to a similar challenge.

Cynics often tell me that today's big innovation opportunities lie in a different realm, especially in technology, and we shouldn't have such high expectations for our own industry. I disagree. The domain in which we operate is still full of opportunities and good ideas, if you know where to look. These can drive progress for many more decades to come. We just need to have the insight to recognize them, the courage to act, and the integrity to place the needs of researchers and scholars at the heart of what we do. If we can do this, then the rest will take care of itself. Thank you.