

Seek and you shall find

Nine industry figures give **Tim Gillett** the low-down on recent developments in discovery as part of the research process

Please tell us what you understand is meant by 'discovery'

Anne-Marie Viola, discovery usage manager, SAGE: Discovery at SAGE is defined as the easiest possible way that users can find our content, whether through the library ecosystem or through the open web. We want our users to get to the content they need, as smoothly and effectively as possible, wherever their starting point and regardless of their device.

Mitja-Alexander Linss, director of marketing, Reprints Desk: Historically, discovery (as it is understood with respect to research) has been the observation of something not seen before. If research is the action, then discovery is the result. But this traditional definition of discovery is evolving at a rapid pace with the advent of advanced research and discovery technologies.

For example, artificial intelligence (AI) is making it possible to automate the discovery process to make it far less time consuming, while producing superior results that previously have not been imagined. In addition, the ability to mine scientific publishers' metadata with semantic searches, rather than relying on a static keyword searches, yields much more meaningful results.

Marty Picco, vice president of product management, Atypion: Discovery is the process through which researchers find useful information. It can be active or passive – purposeful or recommended. In active discovery, such as a Google search, a list of search results is returned after a user submits a query. In passive discovery, such as an Amazon recommendation, results are provided to a user automatically, based on site behaviour, interests listed in their profile, or their previous search history. Discovery can also be serendipitous – a user might come across something they weren't necessarily looking for in the course of searching for something else.

Mike Roberts, content discovery manager, Emerald: 'Discovery' is the finding of content relevant to the user. This might be an active process, where the user is seeking the content, or passive, where content is suggested as relevant to the user, perhaps through related article suggestions or social groups. It may even be a combination of the two.

For a researcher, how has the discovery experience developed in recent years?

Viola, SAGE: Academic discovery has evolved from representation via metadata in cataloguing, in which end users could only search descriptions of content, to new tools that enable full-text search and most recently, semantic search and entity mapping. With the development of web-scale discovery replacing federated search in the last six to eight years, and as the silos of separate databases are eliminated, researchers are able to search more quickly across a greater pool of content using a Google-like 'one-stop shop' mentality. This means that more and more findings, relevant to a researcher's field of interest, are being surfaced in one place, saving them copious amounts of research time.

Linss, Reprints Desk: One of the major shifts in the discovery realm is the move toward the 24/7 global work philosophy. Geographically dispersed research teams are the norm, and the researcher needs to be able to access information and make sense of it whenever they want to and wherever in the world they are working. This is particularly true of scientific literature access as part of the discovery equation, where round-the-clock, unmediated access to journal articles and citations is a must for many research organisations. Reprints Desk has responded to this trend by seamlessly integrating literature access with the most popular discovery tools such as PubMed, Google Scholar, and more than 70 other discovery portals.

Picco, Atypion: The discovery experience has improved in many ways. Discovery has expanded to encompass more types of content – from words and documents to images, audio, video, multilingual results, and others. Knowledge discovery, via technology such as a Google knowledge graph, can infer what content a user might be interested in. Single-dimensional semantic search leverages tags to return →

“Discovery can be serendipitous – a user might come across something they weren't necessarily looking for”

“The research tools landscape is still very fragmented”

→ ‘more like this’ recommendations, and collaborative filtering to return ‘other users also liked’ recommendations. Multidimensional semantic search leverages a combination of topic tags, semantic information about the relatedness of search terms, keywords, and contextual information that can distinguish between whether the word “apple” refers to the firm or the fruit. User-centred technology uses statistics to predict interests and find results tailored and ordered specifically for each user.

Kent Anderson, CEO, RedLink: For researchers, discovery has evolved, but the goals remain largely the same as they were in the print era: to find what I’m looking for, and to find what I didn’t know I was looking for. In the ‘old days’ this would have been accomplished by flipping through print journals, browsing the library stacks to see what was near the specific title that brought you there, and, perhaps most importantly, getting recommendations from peers. All of these discovery methods have digital equivalents that are becoming more sophisticated. Keyword searches have replaced physical browsing; personalised recommendations driven by AI and recommendation engines provide links to related information, and social media helps us see what people like us might recommend.

How can researchers decide which tools are going to do the right job for them?

Yuval Kiselstein, Ex Libris, a ProQuest company: Users of all types will invariably use several discovery services to accomplish their research needs. Each will learn through trial and error which service works best for them, though there are opportunities to help them determine which is the best resource for them to use. Of course, trial and error can help confirm choice of the most appropriate and best discovery avenue. Though, education and guidance provided by more experienced personnel – especially those that work inside the library, departments and faculty,



Marty Picco, vice president of product management, Atypion

office of student learning, and other fellow researchers – can help facilitate which tool is most appropriate to use.

Babis Marmanis, VP of engineering, Copyright Clearance Center: The first step is to clearly identify the information need that the researcher has. A researcher should ask questions such as: Do I want a tool that can help me identify a specific article? Do I want a tool that provides high-quality metadata information? Do I want my tool to work across multiple data sources and possibly mixed types of content (e.g. text, audio, videos)? Do I want a tool that can provide summarisation of information? Do I want a tool that uses speech recognition? Do I want immersive experiences in my discovery process? Answers to those questions will lead the researcher to the right tools.

Jan Reichelt, managing director at Web of Science, Clarivate: The research tools landscape is still very fragmented. It does probably make sense to first take a look at the more well-known products and explore their individual usefulness. My own experience is that whilst new tools can offer a lot of ‘excitement’, it’s really difficult to build something that gets wide market adoption, which allows for more interoperability and collaborating with colleagues and which means it has all the necessary features. It’s also quite common that researchers talk to each other about which tools they use individually, therefore word-of-mouth is used a lot in the decision making process.

Giuliano Maciucci, head of product, eLife: Ask. Whatever specific need a researcher may encounter, they’re unlikely to be

the first. There’s a vibrant community of researchers out there sharing the ways they work, learn and overcome day-to-day challenges. Whether at meetings or, increasingly, on social media, talking to people in the same boat is a great way to discover new tools. Once some suitable options are found, opt for tools that are open source and community backed, as they will likely evolve alongside your work without locking your ability to perform research to a vendor’s business plan. At eLife, we actively support open-research tools and are always on the lookout for ways to help move them forward.

What is more important for researchers – the overall breadth of data coverage offered by a particular discovery tool, or the user experience?

Kiselstein, Ex Libris: Our vision of discovery doesn’t weigh breadth of data coverage over the user experience. One is not more important than the other, as they go hand in hand. It’s a balancing act, though, as both are critical for helping a user find an appropriate item or explore available resources. From our perspective, it’s important that those implementing a discovery service should determine what’s best for their researchers. As they know what the most appropriate device, and access points, to be used by their researchers to discover the content they want them to discover. That said, our objective is to design a service that enables access to the breadth of data coverage a library wants to be discovered and to design a user experience to ensure

the discovery and delivery of those resources isn’t inhibited by poor design or restricted intentionally.

Marmanis, Copyright Clearance Center: It depends on the use case; for example, in the regulatory context of literature monitoring for pharmacovigilance, the underlying data coverage is crucial to success. That said, researchers have grown accustomed to intuitive online consumer experiences, like those offered by Amazon, Google, Uber and others. If a tool is not initially easy to navigate and straightforward to use, the researcher may not stay around long enough to appreciate the less obvious benefits, like broad coverage of content and data.

Reichelt, Clarivate: This is a very difficult answer. Historically research tools and related services have been more focused on the breadth/depth of data coverage, and have neglected user experience. In the last couple of years we have seen new tools that are a lot more focused on user experience, but lack the breadth/depth, which means those products don’t really solve the problem as well, but they ‘feel good’. For professional research, however, in the end the data and quality count. Therefore, I believe a good approach is looking at high-quality products that have all the data breadth and depth, and work on improving their UX. Whilst it’s not easy, it’s probably easier than building the data quality that’s needed to meet the necessary professional standards. Our ambition has to be high data coverage combined with a great user experience.

Linss, Reprints Desk: The breadth of data coverage and the user experience cannot be mutually exclusive. In fact, today’s researcher must be able to access and manage data in ever-increasing amounts and complexities, within a systematic workflow designed specifically for them. Just as internet search technology increases the awareness of the massive amount of available scientific data, new knowledge management technology that incorporates any familiar discovery tool into a sensible research methodology is finally making access to the right data time- and cost-effective.

“The breadth of data coverage and the user experience cannot be mutually exclusive”

Far more valuable than simply discovering then presenting information, these new personalised access tools will help the researcher make the best of available content by pinpointing what’s important. Personalisation is dramatically improving the way researchers in all fields are able to work.

With an eye on AI and machine learning, what developments are likely to be important over the next year or two?

Roberts, Emerald: Data mining, of content itself and of citation networks, is going to help create discovery tools which are tailored to the individual. As an example, we might expect integration with virtual personal assistants (like Deakin Genie) to help with tasks such as alerting to new content. Voice search is going to become more important in this arena, as providers like Amazon and Google work out how to improve and tailor the performance of these systems. The importance of the source publication brand is likely to diminish as AI helps identify ‘nuggets’ of information highly relevant to the individual and the work they’re doing.

With awareness and need for open data increasing, there will be better incentives and improved infrastructure for researchers to share and discover their field of interest in the near future. The value chain will soon increase focus on the quality of data and new routes to discoverability in the research landscape.

Anderson, Redlink: Artificial Intelligence gets more intelligent all the time, and consumers of all types, including researchers, increasingly expect organisations they interact with to anticipate what they want. The danger is that we lose the element of surprise and the opportunity to try something completely new and unexpected, if our future behaviour is driven by predictions based solely on our past behaviour. For researchers, looking for a new solution, or perhaps a completely new project, it will be important that these tools don’t limit the horizons of inquiry.

Maciucci, eLife: AI may soon be able to go some way towards helping to connect and contextualise the somewhat fragmented landscape of online research artefacts, and helping to translate more of that mass of information into useful knowledge. We’re currently looking at how neural networks and computer vision may help extract



Mitja-Alexander Linss, director of marketing, Reprints Desk

semantic information from PDF research manuscripts, in an effort to provide a solution that may one day help publishers and preprint repositories transition their legacy research publications to more metadata-friendly, data-mineable formats such as JATS XML. This will ultimately help move us all one step closer to finally shedding the centuries-old legacy of print the research publishing industry is still carrying today, even in its digital incarnations.

The really exciting aspect of AI is how it’s proving itself a remarkably effective tool in accelerating key aspects of research, from drug candidate screening and cancer diagnostics, to the search for undiscovered exoplanets around distant stars. AI has the potential to cut through the labour of research, leaving us mere humans to deal with the discovery, that final cognitive leap that turns data into knowledge. I believe AI should be seen as a complement to human insight, not a rival. If used properly, AI could well turn out to be the most powerful discovery tool we’ve ever built.

Reichelt, Clarivate: Machine learning and AI will have an impact on several developments: firstly, I think they will help to automate certain processes that were manual and cumbersome before. For example, Kopernio uses AI algorithms to bring down the number of clicks needed to access a journal article to only one click, where previously a researcher would have to click 10+ times. Secondly, I believe that machine learning will help us remix content and data to create new solutions and products that we didn’t have before, when we were unable to look at data points at scale. And lastly, AI will improve how new discoveries are made. Artificial intelligence and related algorithms allow us to sort through vast amounts of data, leading to new discoveries that we never thought possible. **Ri**