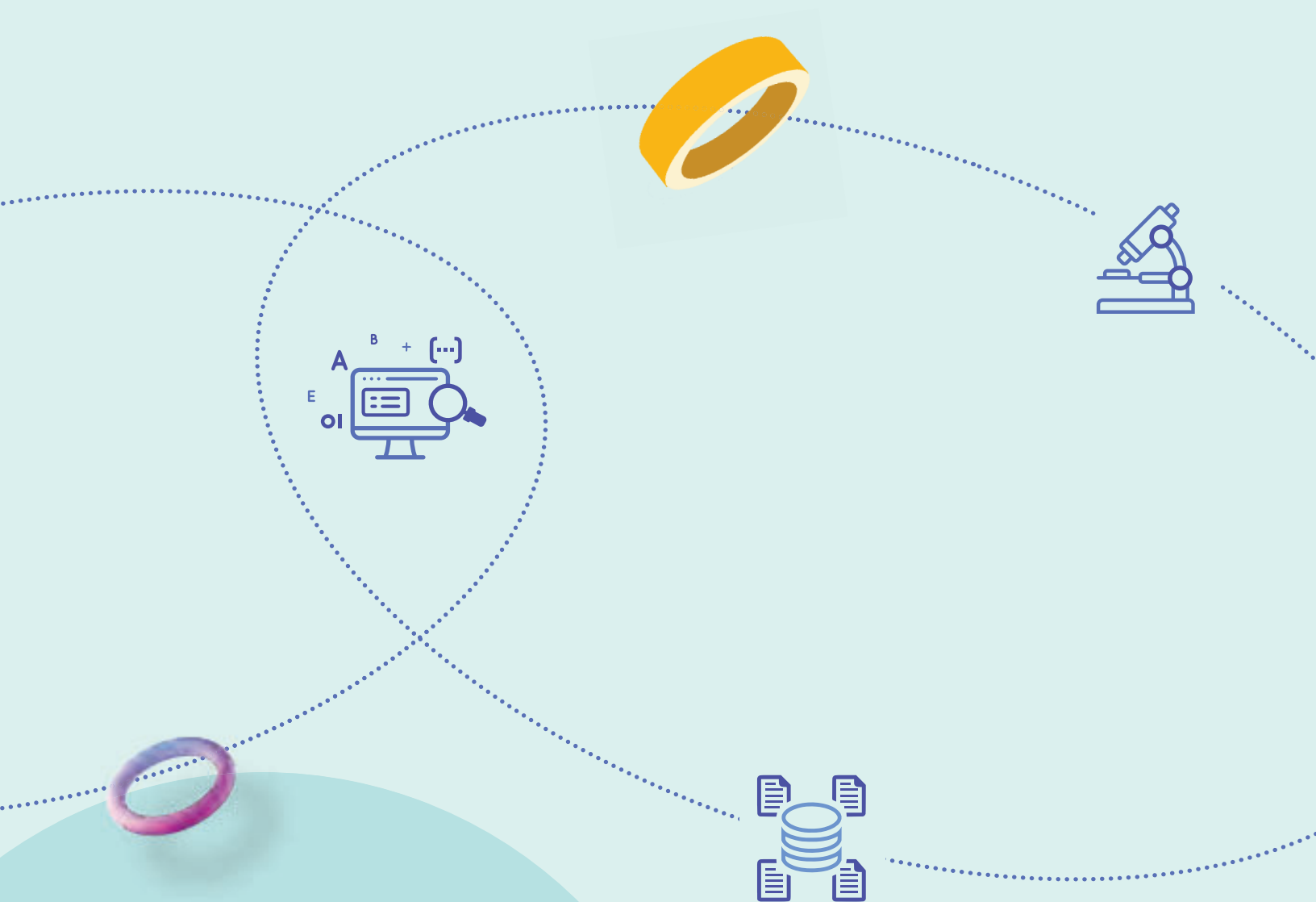# The Need for a Next-Generation Research Repository

Adi Alter, Director of Product Management, Ex Libris

Eddie Neuwirth, Director of Product Management, Ex Libris

ExLibris®
a ProQuest Company

Current institutional repositories are not meeting the needs of research universities effectively. Here's how a next-generation research repository can solve this challenge.

Institutional repositories that collect a university's research assets in one place and make them publicly available serve an important role.

These repositories, which are usually managed by the library, help disseminate the work of faculty members to the broader academic community, making it easier for other researchers to find, use, and build on the knowledge generated by a university. They also help universities comply with rules requiring research funded with public tax money to be made publicly available.

But in many ways, the institutional repositories that exist today are not meeting the needs of libraries or the research community effectively. For instance, they often lack a clear and cohesive structure, which makes research assets hard to find. They are also largely cumbersome to maintain, with inefficient workflows that make it difficult to deposit new research outputs, link publications with their underlying data sets, and add comprehensive metadata to make these assets discoverable.

As a result, the research assets of universities are not being showcased as well as they could be—and staff are spending too much time on these labor-intensive tasks.

In this white paper, we will make the case for how a next-generation research repository can solve these challenges. With input from research universities around the world, we will outline a vision for a next-generation research repository that meets the needs of both researchers and institutions far more effectively.

THE LIMITATIONS OF CURRENT REPOSITORIES

Institutional repositories can mean different things to different universities. But generally, they are intended to store the content created by scholars at the university and make this material accessible to a wider audience. In most cases, university libraries are tasked with managing these repositories—sometimes with the help of their IT department.

Here are some of the many ways the repositories being used today fall short of meeting institutional needs:

Lack of a clear strategy

Many universities have no clear strategy for how they will store and manage their research assets. This often results in the creation of multiple repositories within the same institution.

According to one research article, "Many institutions describe a situation where they have as many as five different platforms … that have characteristics of [an institutional repository]." This "fragmented environment"

results in a duplication of effort across the university and means that "standards are often implemented a little differently from one repository to another." [1]

In some cases, universities have lost sight of what the goals for their institutional repository were in the first place—or else these goals have evolved over time. In other cases, the personnel who created the institutional repository might not be the same people who manage it today, and so the policies defining what belongs in the repository may not be clear.

At some institutions, the repository began as a way to showcase only open-access materials, which make up a just small fraction of their research assets. As leaders looked to expand the kinds of research assets they were collecting and managing, they developed additional repositories for other kinds of works. Sometimes, universities are forced to create multiple repositories because the systems they are using will not support new asset types. Therefore, a university might have one repository for publications, another for data sets, and so on.

Instead of having a single, unified place where all research assets are stored, universities frequently end up with separate silos of research content. Not only are the data in these silos disconnected, but having disparate systems makes it harder to standardize the collection of research assets and apply metadata to these assets consistently. Managing multiple repositories also takes more time and effort.

And even when institutions don't have multiple repositories, the lack of a clear strategy can be problematic and can result in a repository that is a hodgepodge of unrelated content, thus undermining the value of the repository and its content.

Inefficient workflows

Populating an institutional repository can be a challenge. Typically, faculty are encouraged to deposit their research outputs by filling out an online form. But many researchers don't follow through on this step, either because they don't have the time or they don't see the value in doing so.

And even if faculty do deposit their research outputs, they might add the wrong version of their publication— or they are likely to upload it with incomplete metadata. Library staff frequently have to fix incorrect entries or enrich the metadata associated with these entries to ensure that research assets are easily discoverable. With no easy, systematic way of doing this, the process can take several hours of staff time.

Incomplete and disconnected data

Inefficient processes tend to discourage researchers from uploading their work into an institutional repository, and this results in the omission of many research assets.

---

[1] Arlitsch, K., & Grant, C. (2018). Why So Many Repositories? Examining the Limitations and Possibilities of the Institutional Repositories Landscape. *Journal of Library Administration*, 58(3), 264–281. doi:10.1080/01930826.2018.1436778

Even when articles or publications are deposited, there is seldom an easy way to add metadata or link to the data this work derives from. This means other researchers cannot easily verify or build on this scholarship. And if metadata are incomplete, research assets are less likely to be discovered.

WHY DOES THIS MATTER?

These shortcomings in how research outputs are collected, stored, and managed have important implications for everyone involved in the research lifecycle.

- *For researchers,* not promoting their work as effectively as possible in a research repository means they could be missing opportunities for professional recognition, connecting with colleagues at other institutions who are doing similar research, and even securing future grant funding.

- *For research librarians,* having to spend countless hours manually updating the institutional repository and adding or enriching metadata takes away from the time they could be spending on more strategic work instead, such as scaling up their activities to support more researchers and projects, helping researchers develop data management plans, and advising on and ensuring compliance with open-access policies.

- *For universities,* when assets are missing from a repository, leaders have an incomplete picture of what their research outputs are. If the institution is not showcasing its work effectively for potential students, donors, or faculty, this could make fundraising, student recruitment, and attracting and retaining top-notch talent more difficult. Noncompliance also could become an issue in cases where grant programs require research results to be made publicly available.

- *For the research community and the public as a whole,* when a university's research outputs are not easily discoverable, this deprives researchers from other institutions of the opportunity to validate, challenge, support, or expand upon this body of knowledge, as well as the opportunity to find collaborators for their own work. It also prevents the public from engaging in "citizen science" and even just learning about the work of academics. The entire sum of knowledge is less rich as a result.

UNIVERSITY OF
SURREY

The University of Surrey's experience shows the challenges that are common among research institutions as they try to archive, manage, and promote faculty research.

A public research university in England, Surrey has had an institutional repository showcasing faculty research papers for more than a decade, says Fiona Greig, Head of E-Strategy and Resources for the university's library. But when the U.K. came out with a new rule requiring universities to make the data from research projects publicly available, Greig and her colleagues realized this system would not be sufficient to meet their needs. "Any data set produced with government funding must be made publicly available when a research article is published—and the data have to be reusable and accessible for at least 10 years," Greig explains. But the repository that Surrey was using was not designed to store data sets as well as journal articles. This forced the university to use a separate repository for its research data.

Managing the original repository was challenging enough from a resource standpoint, without adding a second one to manage.

Because academics were often confused about how to deposit their research and add the proper metadata, the library had made this a fully mediated service for faculty. But this resulted in a significant workload for library staff. "I have three team members whose job is to upload open-access materials," Greig says.

Having a separate repository for research data adds to this work, and the systems' shortcomings create further challenges. For instance, it's not easy to link data sets with the journal articles they go with—and there is no hierarchical structure to these repositories. "Our existing repositories are just flat boxes with metadata," Greig observes. "They take everything as non-relational documents."

In addition, these repositories do not have advanced metrics for measuring the impact of the university's research. With all of the time spent uploading and maintaining research assets, librarians can't focus on this task themselves. As a result, "we are not measuring our outputs very well," she says.

Faced with all of these challenges, Greig and her colleagues are looking for a new approach to managing research assets. "We need to be smarter about the technologies we use," she concludes. "It's too big a risk for institutions like ours not to comply with open-access requirements. The current workload needed to ensure compliance is not sustainable, and we need something that is easier and more efficient moving forward."

WHAT AN IDEAL RESEARCH REPOSITORY SHOULD LOOK LIKE

A next-generation research repository can solve these challenges by making it much easier to collect, manage, promote, and track the impact of a university's research outputs. The entire research community would benefit from these improvements.

Here are the key characteristics that we believe such a repository should include.

Comprehensiveness

A next-generation research repository should enable universities to collect, manage, and showcase all of their faculty outputs and data within a single repository. To meet this requirement, it must be able to support a wide variety of asset types across a full range of academic disciplines. The repository should not be limited to publications but should also include pre-prints, data sets, audiovisual media, creative works, computer code, blog posts, and other formats of materials.

Connectedness

The ideal research repository would give universities an easy way to link research outputs with the data sets, presentations, blog posts, press coverage, social media mentions, awards, and other materials and activities associated with these outputs. That way, anyone who is reading a faculty member's research paper would have access not just to the paper itself, but to a wide range of information that could help them better understand and make use of this scholarship. Users would be able to navigate easily from one related asset to another.

Openness

The ideal research repository would apply the FAIR Data Principles to make data findable, accessible, interoperable, and reusable by other researchers and institutions[2]. It should also integrate seamlessly with a university's existing workflows and technology systems through application programming interfaces (APIs) and well-known standards.

Automation

A next-generation research repository would use automated processes to capture information and make it easier to deposit research assets wherever possible, thus reducing the workload on librarians and faculty. For instance, it should be able to identify journal articles published by faculty, capture the metadata associated with these articles, and add this research to the repository automatically. By using flexible data models to support various asset types, a next-generation research repository would give faculty and librarians a very structured way of adding metadata to ensure that research outputs are fully discoverable.

Advanced analytics

The ideal research repository should use advanced analytics to give leaders greater insight into the impact of their institution's research. Provosts, deans, research office staff, and others should be able to glean insights that go beyond just how many papers their faculty have published in academic journals and how often these papers have been cited. For instance, leaders should be able to track faculty publications in non-academic channels as well, such as traditional and social media. They should be able to measure the impact of a variety of asset types and get an accurate picture of the research collaboration that is occurring.

Easy to scale and support

A next-generation research repository must be easy to scale and support. It must be able to grow with the institution as needs change. Ideally, it should be cloud-based, so universities are automatically using the latest

---

[2] See "Implementing FAIR Data Principles: The Role of Libraries" from the Association of European Research Libraries: https://libereurope.eu/wp-content/uploads/2017/12/LIBER-FAIR-Data.pdf.

version without having to upgrade or re-implement the repository over time. In essence, the solution should allow library and IT staff to work more efficiently by focusing their efforts on supporting researchers instead of managing and maintaining the technology.

USE CASE: DREXEL UNIVERSITY

Like the University of Surrey, Drexel University has been using an institutional repository that was not intended to store and manage research data. As a result, Drexel has maintained its research data in separate collections—but the university did not have a way to make these data collections searchable.

Drexel leaders wanted to adopt a more thoughtful, integrated approach to managing research assets in a way that would ensure regulatory compliance, while also making all assets easily discoverable.

Recognizing the need for a more strategic approach, stakeholders from Drexel's research office, graduate schools, and libraries engaged in conversations to address these challenges. Their discussions focused on three areas in particular: the policies, communication, and technology needed for success.

"Managing research outputs is not a solo activity," says Dean of Libraries Danuta Nitecki. "It takes a village."

With respect to the technology that Drexel would use for this task, university leaders hope to implement a system that can integrate research outputs with their underlying data sets, while also making the data discoverable. A key consideration is that the system must be easy for faculty to use in promoting their research to a global community. "If you make the process too complicated, nobody is going to do it," Nitecki says.

Nitecki and her colleagues believe they have found such a system in Esploro, a new cloud-based research services platform from Ex Libris.

How Esploro Meets the Need for a Next-Generation Research Repository

In consultation with library and research staff at universities around the world, Ex Libris has developed a next-generation research repository that meets these key criteria. Called Esploro, this cloud-based solution helps universities showcase their research and measure its impact by systematically capturing, managing, and disseminating research outputs and data using a unified repository, integrating research evidence from diverse systems.

Esploro is based on open standards, so that it integrates easily with other existing research systems—and it includes automated processes to save time and simplify data capture for librarians and researchers.

For example, by leveraging integrated researcher profiles and Ex Libris' discovery services, Esploro can automatically identify research published by faculty in journals, external repositories, and other sources; capture

the relevant metadata associated with a source; and create a record within Esploro for that research asset. This saves librarians and researchers from having to manually enter all of this information for themselves.

Esploro also broadens the scope of research assets that can be stored in a research repository. The solution's flexible data model supports a wide variety of research assets, including data sets, creative works, activities, and other materials. Each type of asset has its own unique schema for capturing metadata, with data fields that are relevant to that particular asset type. Having a unique schema for each asset type helps institutions add these various research assets to the repository easily, while also capturing rich, high-quality metadata to ensure discoverability.

In addition, Esploro contains built-in metrics to demonstrate the use and impact of research outputs in much more sophisticated ways. The solution does this in two ways: (1) by broadening the traditional impact metrics such as h-index to take into account multiple asset types, and (2) by measuring impact across both academic publications and traditional and social media.

One of the guiding principles in developing Esploro was that universities are not going to change the habits of their researchers. If a faculty member is already depositing her research into Arxiv, for instance, because that is where all of her peers are working, she is not likely to repeat this step in an institutional repository. With Esploro, she won't have to: The system can capture information about her research directly from Arxiv and create a record for that research automatically. In this way, populating the research repository is no longer dependent on the actions of researchers themselves.

With rule-based workflows that universities can customize according to their needs, Esploro is intuitive out of the box, yet highly configurable for each institution—helping librarians and researchers alike work much more effectively and efficiently.

Use Case:
University of Denver

The University of Denver has a Digital Commons system to store faculty research, as well as faculty profile pages to highlight their work. But these two systems were not connected, which required faculty to enter their research in two different places. Not surprisingly, "we have had very slow uptake in faculty wanting to engage with that," says Dean of Libraries Michael Levine-Clark.

Even faculty who were depositing their research were doing so inconsistently, often with incomplete metadata. University leaders also wanted the ability to capture a broad range of faculty outputs, including creative works—something that wasn't easy with the systems they had in place.

Levine-Clark and his colleagues were determined to find a better solution. "Esploro happened to come along at just the right time, as we were beginning those conversations," he says.

The University of Denver has joined Surrey, Drexel, and other leading research universities from around the

world in becoming early adopters of Esploro. Levine-Clark is looking forward to having a single, unified system for storing and disseminating many different kinds of faculty outputs, with automated processes for capturing information.

"It's important for us to present this as less work for our faculty, not more," he says. "If we can tell them that we'll automatically populate their faculty profile pages with information generated by the system, and all they have to do is review and approve this information, that's a key way we can provide value for them."

Besides simplifying workflows, Esploro will improve the university's ability to measure the impact of its research. "We are currently tracking the number of articles our faculty publish and the number of citations their work receives," Levine-Clark says. "But we would like to have a much better picture of the full scholarly and social impact of their work."

Transforming Research Outcomes

Collecting evidence of a university's research outputs in one place and making them easily discoverable is an important goal. But there are many challenges that stand in the way of achieving this goal, and current institutional repositories are limited in how effectively they can solve these challenges.

A next-generation research repository holds the answer. By automating workflows, broadening the scope of research assets that can be stored and managed, making assets more discoverable, and measuring the impact of research in new and exciting ways, a next-generation research repository such as Esploro can transform how universities manage and promote critical outputs, leading to benefits for everyone involved.

Want to learn more about Esploro? Contact one of our experts and hear what customers are saying.

About Ex Libris

Ex Libris, a ProQuest company, is a leading global provider of cloud-based solutions for higher education. Offering SaaS solutions for the management and discovery of the full spectrum of library and scholarly materials, as well as mobile campus solutions driving student engagement and success, Ex Libris serves thousands of customers in 90 countries. For more information about Ex Libris, see our website, and join us on Twitter, Facebook, YouTube, and LinkedIn.