

Blog Network [\(L\)](#)

Research in
progress
blog
([https://
blogs.biomedcentral.com/
bmcblog/](https://blogs.biomedcentral.com/bmcblog/))

[\(/twitter.com/BioMedCentral\)](https://twitter.com/BioMedCentral)

About this blog (<https://blogs.biomedcentral.com/bmcblog/about>)

The world needs One Repo

There are many, many repositories out there for scientists to store their research, and although this is great, searching for one thing across them all is not possible. Mike Taylor introduces a new project aiming to overcome this issue, [The One Repo \(https://onerepo.net/\)](https://onerepo.net/).

[Mike Taylor \(https://blogs.biomedcentral.com/bmcblog/author/miketaylor\)](https://blogs.biomedcentral.com/bmcblog/author/miketaylor) 22 Jan 2016



Creating a repository of repositories – The One Repo
Pixabay

Suppose you're a dinosaur lover, and you want to find information about sauropods. You go to Google, type 'sauropod' and hit the Search button, and you get a pretty good list of the best Web-pages about sauropods. Easy.

Now suppose you're a dinosaur *researcher*, and you want to find open-access scholarly papers about sauropods. More than that: suppose you want to text-mine them? Or analyze the papers' metadata to see how open-access publication rates have changed through time, or what proportion of the authors are female?

What if you need to verify that the papers conform to an institutional open-access policy? For this kind of thing, you need structured results, not just links to pages – which means that Google doesn't give you what you need.

Google Scholar is a bit better – it understands the concept of a scholarly paper, as opposed to a general Web page, and it has some metadata. But there is no way to access this data except via the Web-facing user interface. (And you're explicitly not allowed to use crawlers to scrape the metadata out of the page.)

Repositories are the answer

The solution is repositories, mostly run by universities and other institutions and so known as Institutional Repositories (IRs). The problem is that there are at least 4000 of them (<https://roar.eprints.org/>) out there, and no good way to search across them all.

Instead, to find every hit, you'd need to submit your query to 4000 separate IRs – all different cosmetically and often varying in much deeper ways – and somehow collate all the results. Clearly that's a non-starter.

The good news is that some important steps have been taken towards aggregating the contents of IRs. In the UK, JISC's CORE (Connecting Repositories) project (<https://core.ac.uk/>) brings together content from most of Britain's universities; the OpenAIRE (<https://www.openaire.eu/search/find>) repository contains the papers from many European-funded projects; the SHARE Notify (<https://www.share-research.org/projects/share-notify/>) service gathers similar information for federally funded research in the USA.

But while these projects include content from some repositories outside their geographical areas, they understandably focus on their homes: for example, while CORE aggregates 248 UK IRs, only 15 are currently included from Asia; and of the 78 content providers presently aggregated by SHARE, 58 are in the USA.

All in one place

That's perfectly reasonable: they're concentrating on their missions. But what the wider world needs is *one* repo: a single place where you can find any open-access paper, wherever it was written or published. A place where you can search on a Web user interface, or via an API, or just download all the data – and where all the data and services are free to use, redistribute and remix.

Introducing The One Repo

That's what we're working on at Index Data. We're building The One Repo (<https://onerepo.net/>), with sponsorship from SPARC Europe (<https://blog.onerepo.net/2015/07/13/sparc-europe-sponsors-the-one-repo/>) and advice from some of the wisest heads in open access (<https://blog.onerepo.net/advisory-board/>). We're re-aggregating the aggregations from CORE, SHARE and others, and – crucially – filling in the long tail.

We're mopping up all the repositories that slip through the cracks. Some of them, like SSRN, the Social Sciences Research Network (<https://www.ssrn.com/en/>), are big and important, but don't get aggregated by existing services because they don't provide a harvesting API.

But we can handle sites like this, because we have mature tools that let us screen-scrape services that are only available as user-facing Web-sites. It turns out that many of the known IRs lack APIs: for example, more than a thousand of the repositories registered with OpenDOAR do not support the OAI-PMH (<https://www.openarchives.org/OAI/openarchivesprotocol.html>) protocol. The long tail is *long*; but it's hugely important.

Equally important are the many small publishers: for example, scholarly societies that publish one or two journals in their fields. We want to pick all these up, as well as the big, well-known open-access publishers (BioMed Central, PLOS, and so on).

Developing the repository

Needless to say, this is a big job. We have most of the software infrastructure in place, but the much longer-term task of content acquisition stretches out ahead. We need to build harvesters for those thousand-plus repositories that have no APIs, and for

numerous publishers. So we're keeping an eye out for partner organisations that can sponsor the development of some of these harvesters. (Do [drop me a line \(mailto:mike@indexdata.com\)](mailto:mike@indexdata.com) if you're interested.)

Our commitment is to keep the service completely free – not just zero-cost, but making all the data fully open for any and every purpose. There are lots of cool things that we and our partners are looking forward to doing with it. But our fondest hope is that other people will find all sorts of new uses for the data that we've not thought of.

Topics:

[Medical Evidence \(/blog/topics/medical-evidence\)](/blog/topics/medical-evidence)

[Open Access \(/blog/topics/open-access\)](/blog/topics/open-access)

[Publishing \(/blog/topics/publishing\)](/blog/topics/publishing)

Tags:

[data repository \(https://blogs.biomedcentral.com/bmcblog/tag/data-repository/\)](https://blogs.biomedcentral.com/bmcblog/tag/data-repository/)

[medical evidence \(https://blogs.biomedcentral.com/bmcblog/tag/medical-evidence/\)](https://blogs.biomedcentral.com/bmcblog/tag/medical-evidence/)

SHARE THIS POST

< PREVIOUS POST

Leaving academia: what to look for in a new career

13 Jan 2016

(<https://blogs.biomedcentral.com/bmcblog/2016/01/13/leaving-academia-new-career/>)

NEXT POST >

Planning ahead: what to think about prior to submission

25 Jan 2016

(<https://blogs.biomedcentral.com/bmcblog/2016/01/25/planning-ahead-think-prior-submission/>)

[View the latest posts on the Research in progress blog homepage \(https://blogs.biomedcentral.com/bmcblog\)](https://blogs.biomedcentral.com/bmcblog)

COMMENTS

About

[About our blog network \(https://blogs.biomedcentral.com/about/\)](https://blogs.biomedcentral.com/about/)

[Copyright information \(https://blogs.biomedcentral.com/about/copyright-information/\)](https://blogs.biomedcentral.com/about/copyright-information/)

[Community Guidelines \(/about/community-guidelines\)](/about/community-guidelines)

[About BioMed Central \(https://www.biomedcentral.com/about\)](https://www.biomedcentral.com/about)

Network Blogs

[On Biology \(/on-biology/\)](#)

[On Medicine \(/on-medicine/\)](#)

[On Health \(/on-health/\)](#)

[On Society \(https://blogs.biomedcentral.com/on-society/\)](https://blogs.biomedcentral.com/on-society/)

[On Physical Sciences \(https://blogs.biomedcentral.com/on-physicalsciences/\)](https://blogs.biomedcentral.com/on-physicalsciences/)

[Research in progress blog \(/bmcblog/\)](#)

[BMC Series blog \(/bmcseriesblog/\)](#)

[Bugbitten \(/bugbitten/\)](#)

Related Blogs

[SpringerOpen blog \(https://blogs.springeropen.com/springeropen\)](https://blogs.springeropen.com/springeropen)

Network Topics

[Biology \(https://blogs.biomedcentral.com/blog/topics/biology/\)](https://blogs.biomedcentral.com/blog/topics/biology/)

[Medicine \(https://blogs.biomedcentral.com/blog/topics/medicine/\)](https://blogs.biomedcentral.com/blog/topics/medicine/)

[Health \(https://blogs.biomedcentral.com/blog/topics/health/\)](https://blogs.biomedcentral.com/blog/topics/health/)

[Developing World \(https://blogs.biomedcentral.com/blog/topics/developing-world/\)](https://blogs.biomedcentral.com/blog/topics/developing-world/)

[Open Access \(https://blogs.biomedcentral.com/blog/topics/open-access/\)](https://blogs.biomedcentral.com/blog/topics/open-access/)

[Technology \(https://blogs.biomedcentral.com/blog/topics/technology/\)](https://blogs.biomedcentral.com/blog/topics/technology/)

[Publishing \(https://blogs.biomedcentral.com/blog/topics/publishing/\)](https://blogs.biomedcentral.com/blog/topics/publishing/)

Community

[Network RSS Feeds \(https://blogs.biomedcentral.com/network-rss-feeds/\)](https://blogs.biomedcentral.com/network-rss-feeds/)

[Follow us on Twitter \(https://twitter.com/BioMedCentral\)](https://twitter.com/BioMedCentral)

[Follow us on Facebook \(https://www.facebook.com/BioMedCentral\)](https://www.facebook.com/BioMedCentral)

Follow us on LinkedIn (https://www.linkedin.com/groups?gid=878087&trk=hb_side_g).

Follow us on YouTube (https://www.youtube.com/channel/UCNowWtw7oCZH3f0u2tF_wbw).

By continuing to use this website, you agree to our [Terms and Conditions](https://www.biomedcentral.com/terms-and-conditions) (<https://www.biomedcentral.com/terms-and-conditions>), [Community Guidelines](#) ([/about/community-guidelines/](#)), [Privacy statement](https://www.biomedcentral.com/terms-and-conditions/privacy-statement) (<https://www.biomedcentral.com/terms-and-conditions/privacy-statement>) and [Cookies](https://www.biomedcentral.com/cookies) (<https://www.biomedcentral.com/cookies>) policy. [Manage cookies](#)

SPRINGER NATURE (<http://www.springernature.com>)

© 2024 BioMed Central Ltd unless otherwise stated. Part of [Springer Nature](http://www.springernature.com) (<http://www.springernature.com>).