

From Hydra to Samvera: an open source community journey

The Hydra Project started in 2008 through a partnership between the University of Hull, the University of Virginia, Stanford University and Fedora Commons (now DuraSpace) to create tools that support use of the Fedora digital repository. Hull adopted the software outputs from this collaboration for its institutional repository in 2011 and remains an active Partner in the community, serving on the Steering Group and fostering development of the community and software in the UK and mainland Europe: the community now has 35 formal Partners and over 70 known adopters internationally. In June 2017 Hydra changed its name to Samvera, Icelandic for 'being together', to recognize the value gained from multiple institutions working together to create the underlying common basis upon which multiple different repository solutions have been implemented. Samvera can be adopted through a set of tools to develop your own repository (using a package called Hyrax as the starting point) and is also available as a complete repository solution, hosted or local, through the use of Hyku. The community has been at the heart of making Samvera a success, and will continue to underpin its future direction.

Introduction

The story of digital repositories within UK higher education is one that stretches back to the turn of the 21st century. This article seeks to add to the overall story through the description of the emergence and ongoing development of the Hydra/Samvera digital repository platform and services. It traces the history of the development of this open source repository solution, using the experience of the University of Hull, which was a founding partner in the initiative, as a benchmark example of involvement in the community and how this approach has resulted in implementations that are greater than those that might have been developed individually. It also looks ahead to the future of what Samvera could achieve in fostering openness in the management of digital research outputs and collections.

In the UK two factors in the early 2000s stimulated the advent and development of digital repositories:

- the emergence of the e-prints movement and open access to these papers, which saw the creation of the EPrints software at the University of Southampton¹
- the Jisc Focus on Access to Institutional Resources (FAIR) programme that ran from 2002 to 2005,² which drew attention to the necessity and benefits of managing institutionally generated digital assets, and which saw both EPrints³ and the MIT-developed DSpace⁴ applied to a number of use cases.

The growth of the open access (OA) movement during that decade and the interest in the role of digital repositories sparked by the FAIR and subsequent Jisc programmes led to further investment in the establishment of institutional digital repositories to good effect across universities.

Arriving a little later to the party was the Fedora digital repository system⁵ (not to be confused with the Fedora Linux distribution, with which there is no link). Cornell University had actually developed the first instance of this software in 1996, as part of a project to identify what factors should be taken into account when managing digital content.⁶ It remained largely a research project until 2001–3 when the University of Virginia in partnership with Cornell received funding from the Mellon Foundation to produce a production-worthy version of the software for wider release.⁷



CHRIS AWRE
Head of Information Services
University of Hull



RICHARD GREEN
Independent Consultant

83 Given the rise in implementations of digital repositories, comparisons between the different platforms were not uncommon. DSpace and EPrints focused on delivering a working solution for wide use around a simple use case: open access. Fedora, by contrast, provided a framework that enabled adopters to build the specific digital repository they needed, and all initial implementations were distinct and appropriate to their purpose. Whilst this approach was very flexible and could be applied to the management of many different digital collections, it essentially meant that every repository was being built from scratch: a very resource-intensive approach. The question arose: how to use Fedora in a way that others could more easily benefit from?

'Fedora ... provided a framework that enabled adopters to build the specific digital repository they needed'

The journey begins

Hydra started as an attempt to answer the question above. Discussion amongst some of the interested parties at Open Repositories 2008 in Southampton led to an initial meeting in September that year at the University of Virginia involving the hosts, the University of Hull, Stanford University and Fedora Commons (now part of DuraSpace). A software consultancy, MediaShelf, also joined subsequent meetings. The purpose of coming together was *not* to create an open source software (OSS) project or community, but to come together with a common need and to dedicate time to addressing that need. Namely, what could be done to facilitate use of Fedora? A series of meetings followed over the course of the next 18 months, supported through institutional intent to address the issue and without project funding, all those taking part taking the view that finding a way forward was a good investment in itself.

'what could be done to facilitate use of Fedora?'

The discussions centred on two areas:

- how the workflows required to interact with Fedora could be built in a way that allowed re-purposing and reuse
- how any additional software tools could be provided in a modular way that would enable them to be used flexibly by different institutions (recognizing that Stanford and Hull, to name but two, were likely to have different local needs).

By May 2010 the principles of how to address these questions had been established. The name 'Hydra' had also been adopted, epitomizing the overall approach taken: a single body of content (the Fedora repository) accessed through one or more workflows (Hydra 'heads') that could be repurposed by different institutions for local use. The three institutional partners decided to implement the approach individually. A technology decision was taken to use Ruby on Rails to provide much of the modularity (through the code's 'gem' structure), and by November 2011 all had a working version of Hydra up and running.

Through dissemination of the Hydra project work at conferences, other institutions expressed interest in the work and started to become involved. This highlighted a challenge for the original participants, who now formed the Steering Group. What had started as a community discussion had evolved into active OSS development and the challenge was now to find an effective way of sustaining both over time. Recognizing that many OSS projects fail through the lack of community, a focus was placed on creating a community that institutions would want to join and have a stake in maintaining. The formal role of Partner was created, with the requirement that becoming a Partner brought with it a commitment (evidenced through a Letter of Intent and signing a Memorandum of Understanding) to contribute to the ongoing development of Hydra as a community and as a software solution. By mid-2012 Hydra had ten Partners; by the end of 2013 this had grown to 18 and the current total is 35, spanning a range of universities and other types of institution.

'focus was placed on creating a community that institutions would want to join and have a stake in maintaining'

Ongoing discussion amongst the Partners was enabled through periodic meetings, the benefit of face-to-face engagement proving to be vital to the success of the conversations and what came from them. As these meetings grew, it was

84 agreed that limiting the conversation only to Partners was proving to be restrictive, as there were many in the community who had expressed interest in working with and contributing to Hydra but were not from institutions at a point where they could viably become a Partner. To address this issue, a conference, Connect,⁸ was started in 2014 and now runs annually in the autumn: 2016 attendance in Boston, MA was 260 from 90 institutions around the world.

In parallel with the Steering Group and Partners, the third arm of the management of Hydra is the developer community who have driven forward the development of both the tools that came out of the original meetings and other new, improved options that add to the functionality available. All contributors, for technical and non-technical contributions, sign an individual contributor licensing agreement (iCLA), and we require their employing institution to sign a corporate version as well (cCLA).

'The core Hydra software is now at version 10.4 and has had over 50 different contributors'

Both are based on the Apache model, which is the licence that all the Hydra software that is produced is released under. The core Hydra software is now at version 10.4 and has had over 50 different contributors over the years.⁹ This has evidenced the ongoing interest and commitment to the community's work in developing Hydra, which continues to evolve apace.

There have been many steps along the way in how Hydra has developed. The major milestones are described within Table 1. More detailed experiences and activities across the Partnership can be found described through the presentations and posters at the annual Connect conference, as referred to earlier.

Year	Milestone(s)	Note
2008	Initial meeting of project partners.	In September 2008, at the University of Virginia.
2010	First software commit to GitHub in May. Initial project focus to that point was on principles and design.	Key to ensuring that Hydra did not become another open source flash in the pan was deliberate caution about developing any software until we felt comfortable that the model and approach was right and stood on its own feet as value for the community.
2011	All three initial institutional partners have implemented production Hydra repository solution. Code released for others.	The implementations were quite different in their scope and use cases, but demonstrated that the same tools could be applied to different purposes.
2012	Community members invited to become Hydra Partners – committed to help further the development of Hydra. LSE becomes first UK Partner to join Hull.	The interest generated at conferences was central to having other institutions pick up the code and take it in new directions whilst maintaining the core elements to ground any development and keep the different implementations linked.
2013	First release of Avalon, a Hydra head solution for audio-visual resources. First release of Sufia, a general purpose self-deposit repository solution. Number of Partners rises to 18.	Grant funding has enabled specific initiatives like the development of Avalon (and Hyku – see below), whilst Sufia was a generous gift back to the community based on the local implementation at Penn State University. Sufia has subsequently acted as the basis for many other repository solutions.
2014	First annual Hydra Connect conference held at University of California San Diego. Release of Fedora 4 and subsequent adaptation of Hydra to use this. Partners = 23	Partner meetings outgrew themselves during 2013 and interest stretched beyond the Partners, so a larger, more open conference, Connect, was created to enable participation. Fedora 4 was a major rewrite and update of the Fedora system, which Hydra adapted to during 2015.
2015	Hydra-in-a-Box project starts developing a stand-alone solution (now called Hyku). Partners = 28	Hydra had started life as a set of tools that made using Fedora more straightforward, but it still required technical effort. The development work of the Hydra-in-a-Box project has started to move Hydra towards being able to also serve needs through an out-of-the-box solution.
2016	Hydra signs formal agreement with DuraSpace for banking and legal services in support of the developing community. Partners = 32	DuraSpace acts as the supporting and co-ordinating body for Fedora, DSpace, VIVO, as well as offering hosted services. Hence, it is well placed to support Hydra as a growing community organization.
2017	The Hydra Project becomes the Samvera Community. First release of Hyrax, successor to Sufia. Hyku released for local or hosted use. Partners = 35, number of known users >70	Hydra shifts from being a project with a name informed by its technology to a name informed by its community: Samvera means 'togetherness' in Icelandic.

Table 1. Key milestones in the development of Hydra

Hull: a journey within a journey

Hull started its own investigations into a digital repository in 2005, deciding to adopt Fedora for the reasons listed below.

- Fedora is a repository that supports the management of any type of digital content, and we did not know what the University might ask us to manage in the future. It also allowed us to use a single repository for different digital collections without needing to manage multiple solutions. It was felt at the time that EPrints and DSpace were more focused in their capabilities, which we felt might be limiting (though we recognize the developments in both systems since then).
- Fedora could scale to large amounts of content, and we knew that digital content was only going to grow in quantity and size.
- Fedora was open source, like DSpace and EPrints, and had an active international community around it. Hull had previously engaged with other open source communities, implementing uPortal and Sakai, and had found this to be a very useful model of engagement to support what we would never have been able to do by ourselves.

In adopting Fedora, Hull initially followed the same path as others and developed a local solution. However, again as with others, we found this approach to be resource-intensive and non-sustainable, and had started looking for an alternative way forward. Richard Green described our situation at the time at Open Repositories 2008¹⁰ and it was the conversation that followed this that sparked the initial project meeting in September that year.

Participating in this meeting, we quickly identified two aspects that were of value to us as an institution in pursuing this partnership. Firstly, Stanford and Virginia are much larger and better-resourced universities than Hull. However, we all recognized that we had a common set of questions and needs, and this enabled us to work together towards addressing these effectively. Secondly, the international nature of the project meant that whatever came from it was, we felt, more likely to have wide appeal and benefit, and all Partners recognized the value of working together on that basis.

We were very fortunate that the University was willing to support the face-to-face meetings that were used to make the initial progress in defining what Hydra would become. This decision was informed in part because of the success of grant funding to cover the travel costs, but also by the previous success in open source engagement. The work with the project Partners led us to implement an initial version of Hydra@Hull in November 2011 (with additional functionality in March 2012), which was subsequently upgraded in January 2014.¹¹ Specific technical resource to continue development of the local repository has been limited since January 2015 due to staff changes, but support from the community and associated vendors in tandem with the local IT Department has provided valuable input to supporting the ongoing provision of the production service. With the advent of Hyku and other hosted options we are now considering our future plans for upgrading and taking our Fedora repository, started as a service in 2008, to the next iteration of its life.

Hull's early involvement in the community and its management has also continued through its role on the Steering Group and contribution of administrative and organizational support. A Hydra UK User Group has morphed into a European equivalent, and Hydra Europe events were held in 2014 and 2015. In the UK, LSE and York are Partners and there is active engagement and implementation work at Durham, Oxford and Lancaster. Other European activity can be found at the Digital Repository of Ireland and the Royal Library of Denmark, both Partners, as well as the Theatre Museum of Barcelona and sites in Germany. The original bonds forged in 2008 remain strong and have grown: part of this continues to be through the common needs that led to the partnership in the first place, but also in part through the friendships created in actively working together on a shared problem.

'The original bonds forged in 2008 remain strong and have grown'

Hydra to Samvera

As part of maturing as an organization, the community agreed in 2016 to proceed with trademarking the name Hydra and the associated logo. In doing so, we unfortunately encountered a trademark challenge from a software company in Germany. This highlighted a couple of learning points for us as an open source initiative, outlined below.

- The trademark challenge was itself capable of being challenged, but we did not have the funds to support the legal fees involved. To that end we focused on working with the German company's lawyers to reach an acceptable process for ceasing use of the name Hydra for public use.
- The situation was explained to the community and a rebranding process launched through them to identify a new name that would take with it the Hydra approach but also reflect where we now are as an organization and community. Working on this, we discovered that within computer software there is a liking for mythical beasts, but that almost all, from different countries, had already been used!

'Samvera is the Icelandic word for "togetherness" or "being together"'

Following this process, Hydra was relaunched as the Samvera Community in June 2017. Samvera is the Icelandic word for 'togetherness' or 'being together'. This reflects the way in which the community has always operated, whilst the logo (see Figure 1) epitomizes the multiple different solutions that working together is enabling. Internationally, the use of an Icelandic word also bridges the American and European spheres within which Samvera is largely used.



Figure 1. The Samvera logo

Open Samvera

In putting together the ideas and principles that have informed the development of Hydra/Samvera, openness was in-built. Fedora is an open source platform. Tools built on top of this to facilitate its use would not need to be open source per se (it could have been a commercial added-value toolset in theory), but the recognition of the value of working together and sharing what we produced for the good of all users of Fedora was embedded from the start. The initial software implementation also made use of other existing open source components, from the Blacklight discovery tool to the use of the Solr index engine, plus additional components contributed by MediaShelf, the software consultancy initially involved. All software produced is released under an Apache 2 licence to facilitate reuse and adaptation.

'All software produced is released under an Apache 2 licence to facilitate reuse and adaptation'

That is not to say that there are not commercial aspects to the way Samvera now operates. Vendors have been vital to successful implementation support, particularly, in the US, Data Creation Experts, the successor to MediaShelf. In the UK, Cottage Labs offer their own instance of Samvera, called Willow, and ULCC is working on a hosted Hyku offering. The value provided is in both the implementation expertise and ongoing service management, akin to the services offered by Atmire for DSpace and EPrints Services for EPrints. Nothing will stop the software being free to use and open, but, like many systems, buying in a helping hand can be beneficial.

87 How does Samvera relate to other aspects of openness today? It is one of the digital repository offerings within Jisc's Research Data Shared Service initiative¹² (as provided by Cottage Labs), contributing to the provision of open research data. It has been used to support open access at Hull, including the provision of download statistics to the IRUS-UK service, whilst some of the biggest download numbers from Hull's repository have been for a series of open educational resources produced a number of years ago to support secondary science education. The essence of Samvera as a digital repository is to showcase collections, and there are examples of implementations providing open access to images, videos and data sets amongst other materials. Can Samvera support open science? To some extent this will depend on how open science is defined and taken forward. But it is flexible and open to being adapted to meet whatever need comes along.

'it is flexible and open to being adapted to meet whatever need comes along'

Where next?

With the change in name, Samvera will be focusing as a group on the ongoing sustainability of both the community and technology. Both will continue to evolve in ways that meet the community's needs. Growth in the Samvera Community has been steady, which has allowed the community to evolve gradually. Some degree of consolidation is now required to emphasize the professionalism and intent of Samvera, both to those senior decision makers at current Partner institutions to support ongoing commitment and to new prospective Partners and adopters. A new Samvera website¹³ has been released, and the community published its first annual report¹⁴ earlier in 2017. Next steps within the community are to develop further the processes used for technology development, to ensure their ongoing sustainability. This work will in part be informed by a reinvigoration of the role of Partners and how they wish to drive forward the development of the Samvera Community and associated technologies.

'building digital repositories is a complex task that requires effort'

Samvera has come a long way in its nine years so far. In making the original decision to use Fedora there was an implicit acknowledgement that building digital repositories is a complex task that requires effort. Samvera will continue to seek ways in which that complexity can be eased through the tools provided, so as to benefit from the detailed functionality of a repository whilst facilitating interaction with it. The University of Hull is extremely grateful for the input we have received from the Samvera Community, and the benefits that this active participation has brought.

Abbreviations and Acronyms

A list of the abbreviations and acronyms used in this and other *Insights* articles can be accessed here – click on the URL below and then select the 'Abbreviations and Acronyms' link at the top of the page it directs you to: <http://www.uksg.org/publications#aa>

Competing interests

The authors have declared no competing interests.

References

1. Tansley R and Harnad S, Eprints.org Software for Creating Institutional and Individual Open Archives, *D-Lib*, 2000, 6(10): <http://www.dlib.org/dlib/october00/10inbrief.html#HARNAD> (accessed 3 October 2017).
2. Awre C and Baldwin C, Focus on access to institutional resources: A synthesis of the JISC fair programme, *New Review of Information Networking*, 2005, 11(2), 137–158; DOI: <https://doi.org/10.1080/13614570600573276> (accessed 3 October 2017).
3. EPrints: <http://eprints.org> (accessed 3 October 2017).
4. DSpace: <http://dspace.org> (accessed 10 August 2017).
5. Fedora: <http://fedorarepository.org> (accessed 3 October 2017).

6. Payette S and Lagoze C, Flexible and Extensible Digital Object and Repository Architecture (FEDORA), *Second European Conference on Research and Advanced Technology for Digital Libraries*, Crete, 1998:
<https://arxiv.org/abs/1312.1258> (accessed 3 October 2017).
7. Payette S and Staples T, The Mellon Fedora Project: Digital Library Architecture Meets XML and Web Services, *Sixth European Conference on Research and Advanced Technology for Digital Libraries*, Rome, 2002:
<https://arxiv.org/abs/1312.1385> (accessed 3 October 2017).
8. Connect, a Hydra/Samvera conference:
<https://wiki.duraspace.org/display/samvera/Samvera+Connect+Meetings> (accessed 3 October 2017).
9. Open Hub analysis of the development of Hydra:
<https://www.openhub.net/p/hydra-head> (accessed 3 October 2017).
10. Green R, Awre C and Wallace I, Records Management and Digital Preservation using Fedora and iCalendar, *Third International Conference on Open Repositories*, Southampton, 2008:
<http://pubs.or08.ecs.soton.ac.uk/98/> (accessed 3 October 2017).
11. Hydra@Hull:
<https://hydra.hull.ac.uk> (accessed 3 October 2017).
12. Jisc Research data shared service initiative:
<https://www.jisc.ac.uk/rd/projects/research-data-shared-service> (accessed 3 October 2017).
13. Samvera:
<https://samvera.org> (accessed 3 October 2017).
14. Samvera Community Annual Report 2016:
<https://wiki.duraspace.org/display/samvera/Annual+Reports> (accessed 3 October 2017).

Article copyright: © 2017 Chris Awre and Richard Green. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use and distribution provided the original author and source are credited.



Corresponding author: Chris Awre
Head of Information Services
University of Hull
Kingston upon Hull, UK

E-mail: c.awre@hull.ac.uk

ORCID ID: <http://orcid.org/0000-0002-0964-254X>

To cite this article:

Awre C and Green R, From Hydra to Samvera: an open source community journey, *Insights*, 2017, 30(3), 82–88;
DOI: <https://doi.org/10.1629/uksg.383>

Published by UKSG in association with Ubiquity Press on 08 November 2017