Doing more with ORCID

ORCID IDs in Research Data Management workflows at the University of Cambridge

Summary

Tying a researcher's work to their online identity, unambiguously, universally and persistently is a perennial problem. At the University of Cambridge, the library has implemented a workflow which creates seamless links between researchers and their works using identifiers and different services.

The University of Cambridge research repository (Apollo), uses ORCID IDs as a unique identifier for researchers. When a researcher submits a dataset to Apollo, a DOI is minted for the dataset through the DataCite service. By including the ORCID in the metadata submitted to DataCite, DataCite then populates the ORCID registry entry for the researcher (with their permission) with information about the dataset, using an 'auto-update' feature.

The result is that a link is created between the researcher and their data, through the ORCID ID identifying the researcher, and the DOI for the data assigned by DataCite. The persistent identifiers are used to connect researchers and their achievements, improving visibility and discoverability across different systems. The workflow reduces duplication of effort in entering information and avoids input or identification errors.

Background

Apollo is the open access information repository at the University of Cambridge which holds research outputs and is an established repository considered key to meeting Open Access and research funder requirements. The outputs managed include publications, theses and research datasets. Apollo is hosted and maintained by the University Library in the Office of Scholarly Communications.

A research information system (CRIS) sits alongside a publicly available portal into the repository. The research information system holds information about researchers, research outputs, and contains metadata on people, grants and publications, including publishers' metadata, and grant information from internal and external sources. The information managed in the CRIS is crucial for supporting preparations for REF 2021.

How ORCID IDs are used: a walk-through of the integration

A researcher's institutional identity is connected to their ORCID ID using the Symplectic Elements CRIS system. This uses the ORCID authentication process, in which the user signs in to the ORCID Registry, therefore making sure that the correct connection is made between the researcher, the institutional system, and their information in ORCID. It also gives the institution an opportunity to request permissions from the user to read information from their ORCID record.

Fig 1 Researcher authentication screen for University of Cambridge

Doing more with ORCID - October 2019

Fig 2 A dataset entry in Apollo. The researcher names show the ORCID ID icon

Communications

The University of Cambridge provides information for their researchers on ORCID starting from this help page https://www.research-information.admin.cam.ac.uk/more-resources/orcid
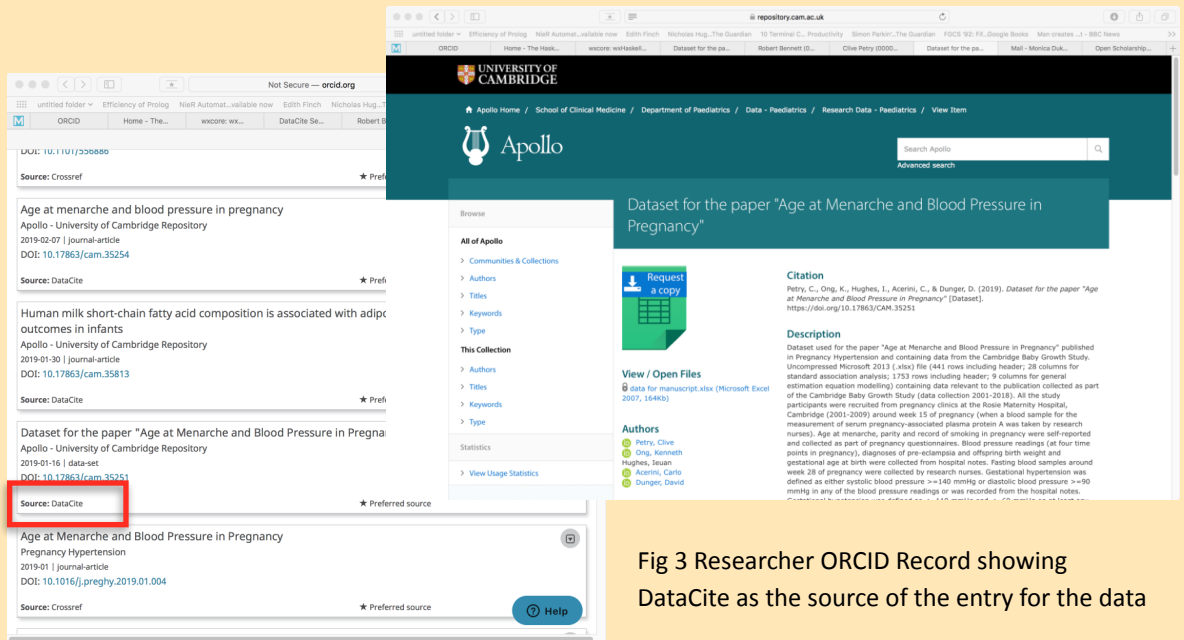

Fig 3 Researcher ORCID Record showing DataCite as the source of the entry for the data

Once the researcher grants permission, the information associated with the ORCID iD can be actively read into the institutional research system, and then also populated directly into the repository, Apollo. In Apollo, the ORCID icon is displayed next to the researcher (Fig 2) showing that they have an ORCID iD.

The researcher submits a dataset to the CRIS (Symplectic Elements) which makes a deposit in Apollo. From Apollo, the next step is for the work to be submitted to DataCite. Once the submission is received and approved in the repository, the dataset is registered with DataCite.

At DataCite a DOI is minted for the dataset. The metadata received by DataCite in the registration process includes the researcher's authenticated ORCID iD. Permissions permitting, the work's metadata is then pushed by DataCite directly to the ORCID record in the ORCID Registry using DataCite's ORCID's auto-update feature, based on the researcher ORCID iD.

On looking at the researcher's ORCID record (Fig 3), there is a list of works that comes from the Apollo repository which has been populated by DataCite. DataCite will show as the source.
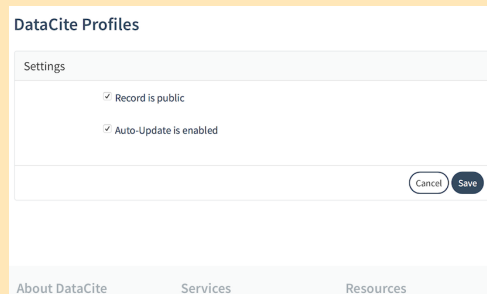
If the researcher grants permission to the research system to read their ORCID record, co-authors can also be pulled from publisher metadata . In order for the workflow to be completed, there is a requirement that the researcher authorises both the research system (Symplectic Elements) to read their ORCID iD, and DataCite to push registered repository works into their ORCID record.

Read More about Auto Update
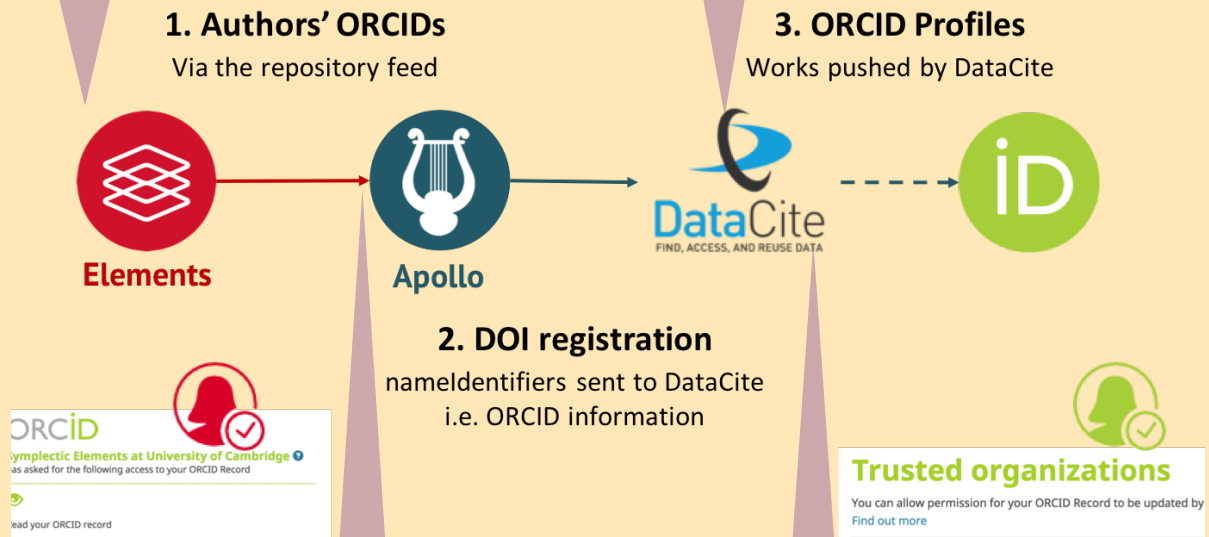https://support.orcid.org/hc/en-us/articles/360006896394-Auto-updates-time-saving-and-trust-building
https://support.orcid.org/hc/en-us/articles/360006894594-Auto-updates-in-third-party-systems-DataCite

Design choices and how it all works behind the scenes

There are two institutional systems in this workflow, the institutional CRIS (Symplectic Elements), and the research repository, Apollo (DSpace). Two external services complete the workflow: DataCite, for registering datasets and minting dataset DOIs, and the ORCID registry, which assigns ORCID ID to researchers. The diagram shows the systems working together to create the workflow.

Cambridge's research information system is Symplectic Elements. Symplectic Elements connects to the ORCID API version 2.0. Submission of data to Apollo is managed through the Elements instance. Elements is used to collect ORCID iDs and link them to researchers.

From Apollo, registration of the dataset with DataCite occurs, passing on the metadata including the ORCID ID The step of submitting the data to DataCite from Apollo is completed using a DOI registration module; the *nameIdentifiers* metadata field is used to include the ORCID IDs in the submission to DataCite.

1. Authors' ORCIDs
Via the repository feed

3. ORCID Profiles
Works pushed by DataCite

Elements **Apollo** DataCite iD

2. DOI registration
nameIdentifiers sent to DataCite
i.e. ORCID information

ORCiD
Symplectic Elements at University of Cambridge
as asked for the following access to your ORCID Record

ead your ORCID record

Trusted organizations
You can allow permission for your ORCID Record to be updated by
Find out more

The research repository Apollo is based on the open source DSpace repository. In 2016 the repository (formerly known as dspace at Cambridge) was upgraded from version 3.0 to 5.0 and relaunched as Apollo. The ORCID IDs collected using the Symplectic Elements integration, either connected directly by the researcher or captured from external sources, are submitted to Apollo with the data. A repository tools connector based on metadata crosswalks is used.

The metadata of the work is pushed by DataCite to the ORCID Registry using ORCID's auto-update feature.

The underlying ORCID permissions model is in place and all actions are enabled subject to the researcher granting permission. DataCite is marked as a Trusted Organisation when the researcher agrees to allow DataCite to add data to the ORCID record regularly (known as auto update). The researcher's institution is also a Trusted Organisation when permissions are requested through the research information system.

In terms of technical resource required, a custom metadata crosswalk was implemented between the CRIS and repository and in Apollo. Further (light) customisation of the display view was needed to show the ORCID icon.

The following are some of the factors that influenced the design choices:

A key driver was the availability of technical resources and integration features. This integration approach takes advantage of the CRIS system vendor connection to the ORCID API, combined with DataCite auto update.

The Symplectic Elements instance was chosen to collect the ORCID iDs and link them with researchers since at the time of development a connection to the ORCID API latest version (API 2.0) was possible and came with the system.

The repository software was not yet able to connect to the latest ORCID API version. The repository module was only version 1.0 (although v 2.0 was work in progress) and the module could not read or write to the ORCID registry. Only author disambiguation was

available at the time, which did not offer sufficient benefits.

Submitting the works through DataCite was used as a way to overcome the limitations of the repository and CRIS system as neither of those two systems were able to write to the ORCID registry at the time of implementation.

One key factor was to remove the need for developer support and future maintenance. The professional units involved (Library, Research Office) have been trying to reduce their load on Developer support and this solution allowed for this to happen without the need for building in additional developer backup for a production service, nor did it leave the outcome fragile or reliant on a single skilled individual to support prospectively.

Doing more with ORCID - October 2019

Outcomes

As of September 2019, 25,550 articles, 1,329 conference proceedings and 1,100 datasets in Apollo have ORCID IDs. Some examples of the top entries containing several ORCID ID associations come from the field of high energy physics, since there are often multiple contributors e.g. https://www.repository.cam.ac.uk/handle/1810/277892

The following benefits have been described:

For the researcher They only need to submit data to one system and the data propagates across different systems automatically (once the researcher has granted permissions)

For the scholarly communications ecosystem the ORCID record is better populated so the information about the researcher's works deposited in Apollo are also available to third party systems through the ORCID record, together with the associated identifiers e.g. the DOI minted by Datacite. This results in richer metadata in the ecosystem.

For the institution Higher visibility for the outputs in Apollo, and by providing this service that is useful to academics, leveraging their efforts to help work out who all the ORCIDs registered to a Cambridge domain in the ORCID registry belong to.

Low technical resources were needed due to relying on the CRIS vendor integration and the DataCite integration which provided the required interactions with the ORCID registry.

The team implementing this solution were recognised for a particularly successful cross-departmental team partnership between **Agustina Martinez Garcia** of the Office of Scholarly Communication, **Owen Roberson** of the Research Office, and **Dean Johnson** of UIS.

Next Steps

The institution is planning to continue upgrading the platforms (both CRIS and repository) and may revisit the integration. In particular with Symplectic Elements, a full integration in the CRIS system has just been released in October (v5.18), with the ability to write to (as well as read from) the ORCID registry. This would enable stating affiliations, and an easier process for granting rights for researchers.

Institutions want to offer a good user experience for researchers by supporting them to manage permissions and control their own information, provide workflows

"These three staff were instrumental in preparing, implementing and evaluating the synchronisation of two internal research systems, a complex and challenging project that succeeded because of effective collaborative working."
From https://www.staff.admin.cam.ac.uk/general-news/professional-services-recognition-scheme-winners

which deliver efficiencies and present them with good usability in systems. For researchers there is currently a non-intuitive, multi-step process and an improvement would be to make this easier. They currently need to pull information into Symplectic Elements (authorising Elements) and they need to authorise DataCite to push to the ORCID registry, therefore doing the authorisation twice in two different systems and interfaces.

Some researchers have reported duplicate publications in their ORCID records due to clashes between Publisher and the repository DOIs, which they have had to remove manually. To help prevent such situations the team has looked into changing the crosswalks used to send metadata to DataCite so that Publisher DOIs are included in the alternate identifiers section of the DataCite profile. This way ORCID can use this information to prevent duplicate publications in researcher records. This fix has just been released for Cambridge repository.

Find out more

Contact:
Dr Agustina Martinez Garcia
Open Research Systems Manager,
Office of Scholarly Communication

University of Cambridge Apollo repository
https://www.repository.cam.ac.uk

ORCID EMEA webinar 'Unlocking the power of ORCID integrations', 31st October 2018 Slides available from https://doi.org/10.23640/07243.7286159.v1
Recording from 18 mins to 33 mins - quick registration required
https://www.gotostage.com/channel/3065235363405651462/recording/ofwebinarsho032b426867f446489878a0dfe09b3ca2/watch

For more information on the UK ORCID Consortium visit http://ukorcidsupport.jisc.ac.uk

For all enquiries from UK Higher Education institutions and services please email help@jisc.ac.uk and the request will be routed to the dedicated Jisc UK ORCID support team   Further information about Jisc jisc.ac.uk