

Available online at www.sciencedirect.com**ScienceDirect**

Procedia Computer Science 106 (2017) 220 – 231

Procedia
Computer Science

13th International Conference on Current Research Information Systems, CRIS2016, 9-11 June
2016, Scotland, UK

Let's Talk – Interoperability between University CRIS/IR and Researchfish : a case study from the UK

Anna Clements*^a, Gavin Reddick^b, Ian Viney^c, Valerie McCutcheon^d, James Toon^e,
Hamish Macandrew^e, Ian McArdle^f, Sophie Collet^g, Juergen Wastl^h

^aUniversity of St Andrews, ^bResearchfish Ltd., ^cMedical Research Council, ^dUniversity of Glasgow,
^eUniversity of Edinburgh, ^fImperial College London, ^gUniversity of Bristol, ^hUniversity of Cambridge

Abstract

Research funders and research organisations both require feedback on the progress, productivity and quality of the research they support. This information originates with researchers, but may be captured in a variety of systems including University CRIS/IR and funder systems. In 2014 all 7 national Research Councils (collectively referred to as RCUK) implemented a harmonised approach to the collection of research output data, currently supported by Researchfish Ltd (referred to as the Researchfish® system). In 2016 this process is gathering feedback from over 60,000 researchers in all UK Universities, and for funders in the USA, Canada, Denmark and the Netherlands, tracking more than £40billion of public and charity research investment and is adding to a dataset of more than 1.5 million outputs.

Researchers, research managers and funders want to find ways to capture this data once and achieve wide re-use of the information. Working together University and Research Council officers, Researchfish Ltd. and Jisc have highlighted that it is important for the “interoperability” between research information systems to be improved. These organisations have started a programme of work to improve the bi-directional flow of information between University and funder systems.

© 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the Organizing Committee of CRIS2016

Keywords: research funding; CRIS; interoperability; Researchfish; research outcomes

1. Pilot project ; bulk upload of publications and grant links from Institutional systems

1.1. RCUK Interoperability Steering Group

To promote discussion between all relevant parties a steering group was established. The group is co-Chaired by Hamish MacAndrew (Head of Research Support Office, University of Edinburgh and Board member, ARMA Association of Research Managers and Administrators) and Ian Viney MBE (Director of Evaluation, Medical Research Council). Membership includes research-intensive universities, funders – including HEFCE (who run the REF in the UK), Researchfish Ltd. (the system supplier), Jisc (the UK's expert body for digital technology and digital resources in higher education, further education and research).

Whilst the remit of the group covers interoperability between institution and funder systems in general, its initial focus was the transfer of publication records linked to funder grant references from Universities to Researchfish®. Publications, in particular journal articles, have widely used unique references (such as DOIs) which can be used to accurately retrieve the publication and associated bibliographic information, and prevent any duplication when exchanging information. However information about the funding that supported the production of publications (linkage to funding references) is incomplete in the bibliographic record. Universities and funders have been actively looking at ways to improve this record in order that both research organisation and funder requirements for a complete publication record attributed accurately to funding sources are met.

From the universities' perspective there is a desire to reduce burden on researchers and support staff by reusing information already captured in their CRIS/IR systems. Open access policies, including those of RCUK and the funding councils, have helped drive a growth of full-text deposits in institutional repositories and CRIS as illustrated in Fig 1. and as referenced in the recent independent advice from the UUK open access co-ordination group to the UK Government (Tickell, 2016). The HEFCE open access policy for the next Research Excellence Framework (REF) exercise (expected in 2020/2021) proposes that the author's accepted manuscript for journal articles and conference proceedings should be 'deposited as soon after the point of acceptance as possible and no later than three months after this date' in an institutional or subject repository. The HEFCE policy was to take effect for all journal articles and conference proceedings accepted after 1 April 2016, but in 2015 this was relaxed to allow for deposit within three months of publication until 1 April 2017. A review is expected in autumn 2016 to determine whether systems have developed sufficiently to support a requirement for deposit within three months of acceptance.

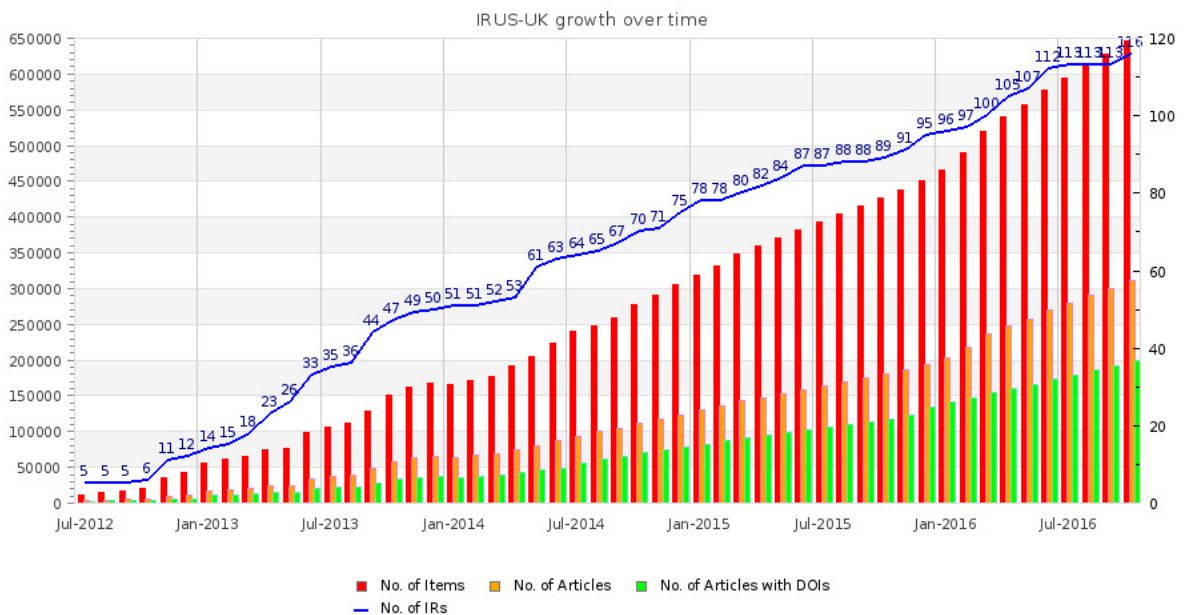


Fig. 1. Growth in repositories and CRIS registered with IRUS-UK (<http://www.irus.mimas.ac.uk>), adapted from “Open access to research publications (Tickell, 2016)”

The RCUK open access policy provides block grant funds to institutions to support the transition to full open access. In order to qualify for this money to support an APC payment an author has to confirm that the publication results from RCUK funding. There is therefore not only the business requirement for institutions to capture information on articles at acceptance for REF, but also information on who has funded the work for allocation of APC block grants. Although these requirements and processes have been implemented by many institutions, these are still new and systems and behaviours can take time to embed. As such it is acknowledged that for many institutions, initially there will be a low proportion of articles for which this information will be available, though with an expectation of future growth.

Funders have been addressing their need to know which publications resulted from their funding by asking researchers to attribute their papers to grant references via Researchfish®. Over 100 funders and UK research organisations subscribe to Researchfish®, and the outputs dataset contains more than 1.5 million records. Research publications currently represent roughly half of the outputs that are linked to funder grant references via Researchfish®.

It was decided to pilot the transfer of publication data from Universities to Researchfish®. Ahead of the funders’ annual output collection exercise in 2016 there was limited time for Universities to carry out preparatory work, and no opportunity to develop any automated approaches to data validation. The worked with data that could be compiled by six pilot institutions in a short six-week period. However it was decided that the pilot work should extend over two collection cycles (years) with the second cycle benefitting from learning from the first.

The aims of the year 1 pilot included:

- Testing the ability of Universities to provide suitable exports of grant-linked publications data to a level of quality required for submission into Researchfish®.
- Monitoring for any changes in researcher behaviours with respect to reporting non-publication outputs separately via Researchfish®.

Based on learning from the year 1 pilot, in year 2 Universities would have more time to compile data, additional Universities may be added to the pilot, and the exercise would benefit from some automation of the process.

Due to the short time period available to prepare data in year 1, the group agreed that quantity of data received from participating institutions would not be considered when assessing the success or failure of the pilot. The pilot would focus on assuring that data was complete, accurate and exchanged without errors. The pilot was also an opportunity to trial communications to researchers and pick up any indications that the behaviour of researchers in reporting outputs had changed.

1.2. Pilot Project Timeline

A project team to co-ordinate the pilot was setup, co-Chaired by Anna Clements (St Andrews) and James Toon (Edinburgh). The project team included representation from the 6 Universities involved directly in the pilot (Bristol,

Cambridge, Edinburgh, Glasgow, Imperial College London & St Andrews), Researchfish Ltd. and Research Councils.

Table 1. Pilot timeline.

Task	Who	Time Period
Agree scope of data	Project Team	15 th Sep 2015
Prepare data upload	ROs	By Nov 6 th 2015
Validate data	Researchfish	November – December 2015
Upload data into Researchfish	Researchfish	December 2015
RCUK submission period		February – March 2016
Analyse submitted data	Researchfish	April – May 2016

The project team met to agree the scope in Summer 2015 and there was limited time for data to be prepared by institutions for the pilot submission in early November. This timetable was required to allow sufficient time to check the data and upload it into Researchfish® in advance of the annual RCUK output submission period (Feb/Mar 2016), thus prioritising quality over quantity of data submitted. The aim was that researchers would be able to see publications contributed by Universities, correctly attributed to grants, when they logged into Researchfish® to review their outputs.

1.3. Pilot Project Data

To be within scope for the pilot, publications had to have a one to one relationship with a unique identifier (either a DOI or a PubMed ID). This meant that the publications would tend to be journal articles but could include other types (e.g. book chapters) providing that they had their own unique identifier. The publications also had to be attributed to an award reference and funder in Researchfish. The targeted date range for the publications was 2015-2016 to correspond with the increase in CRIS use following the HEFCE mandate, but publications outside of this range were also accepted.

Three universities supplied information on all of their researchers and the other three chose to supply information on either a number of schools within their institution or a single department as shown below. The data was restricted to funders taking part in the pilot (Arts and Humanities Research Council, Biotechnology and Biological Sciences Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Medical Research Council, Natural Environment Research Council, Science and Technology Facilities Council, National Centre for the Replacement, Refinement and Reduction of Animals in Research, Academy of Medical Sciences, Cancer Research UK).

Table 2. Scope of data provided by pilot research organisations.

Research Organisation	Scope of Data Provided
Imperial College London	Whole University
University of Bristol	Whole University
University of Cambridge	1 Department +

	small number of additional awards
University of Edinburgh	Multiple Schools
University of Glasgow	Whole University
University of St Andrews	Multiple Schools

2. Results of Pilot Project

2.1. Pilot Project Analysis

The initial outcomes from the pilot were encouraging with almost every record of the around 2,800 publications supplied by the 6 universities passing validity checks. Some records were not imported either because the Principal Investigator (PI) had already added the publication in Researchfish or because the publication attribution was already captured as part of the routine Researchfish collection of attributions from publication acknowledgements.

As a result of the pilot submissions made by the 6 participating Universities, around 2,500 publication attributions were added that would otherwise have been the responsibility of PIs to enter and grant assign manually within Researchfish. The following graph shows the distribution by research organisation and the proportion of the overall data supplied that was subsequently imported.

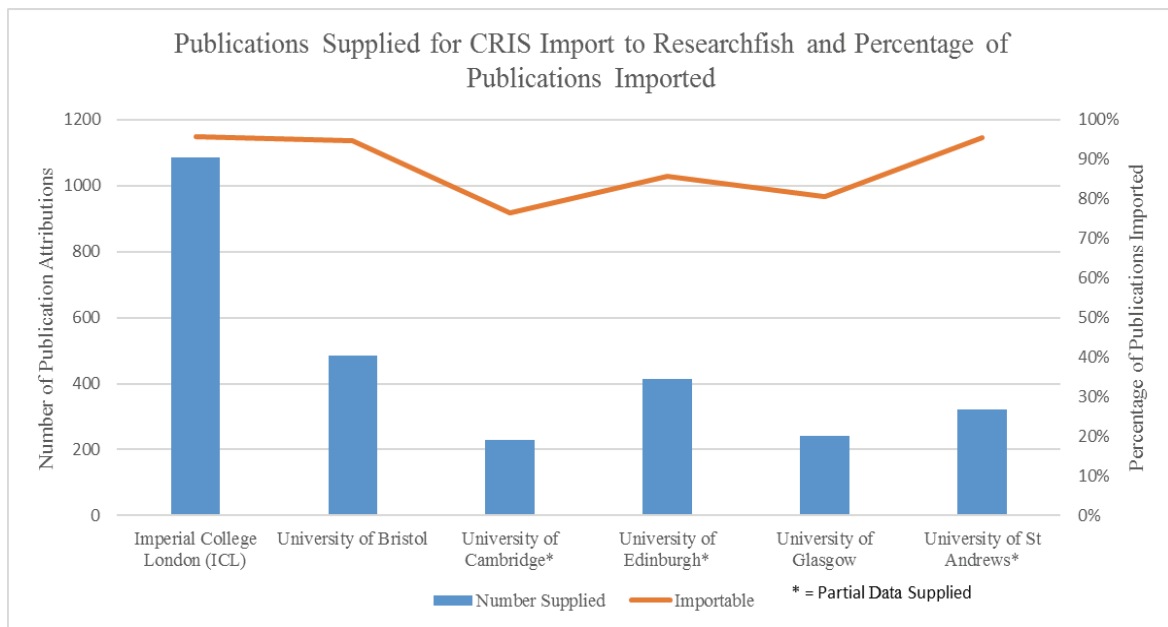


Fig. 2. Distribution by Research Organisation and the Proportion of the Supplied Set Imported.

It is generally accepted good research ethics to include attribution to funding sources in a publication. Many publishers and funders (including all of those in the pilot) have specific policies requiring authors to include the source of funding and funding reference. Researchfish searches publications for those acknowledgements and imports them wherever possible. In 2015 Researchfish added over 31,000 additional publication attributions by this method, including an additional 3,000 publications to the 6 universities in the pilot, without PIs needing to take any action in Researchfish or in local CRIS. However due to the business requirements of institutions to comply with funder OA policies, PIs and/or Library support staff do take action in local CRIS and repositories. If this occurs then there is duplication of data capture effort.

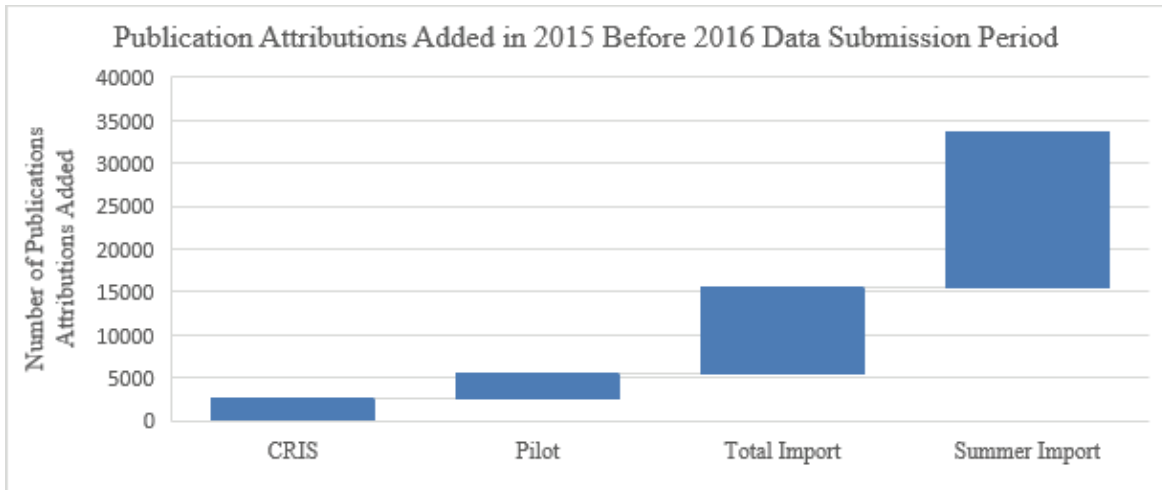


Fig. 3. Publications Imported from CRIS or Publication Attribution.

During the first few months of 2016 researchers added over 220,000 new publication attributions in Researchfish across all funders and research organisations. Figure 3 shows the new 2015-2016 publications added during this period and the publications added as part of the pilot (both from CRIS and attribution within the publications themselves) for the six universities in the pilot. Researchfish includes a number of publication look-ups and integrations (CrossRef, INSPIRE – HEP, ISBN, NASA-ADS, ORCID, PubMed, Scopus, EThOS and Web of Science), as well as the capacity to manually enter a publication without any identifier, and this distribution is also included for five of the six universities. The percentage of publications sourced from CRIS are noted, but the universities supplying partial data may have had more data within scope, but did not supply this as part of the pilot. As stated earlier, the pilot at no stage attempted to judge success or failure on the quantity of grant-DOI/PubmedID links uploaded from local CRIS and repositories. What Fig 4 highlights is the large number of links that are made by PIs directly in Researchfish. This means that links are being created post-publication which is inefficient – ideally the links should be captured as early as possible in the publication lifecycle.

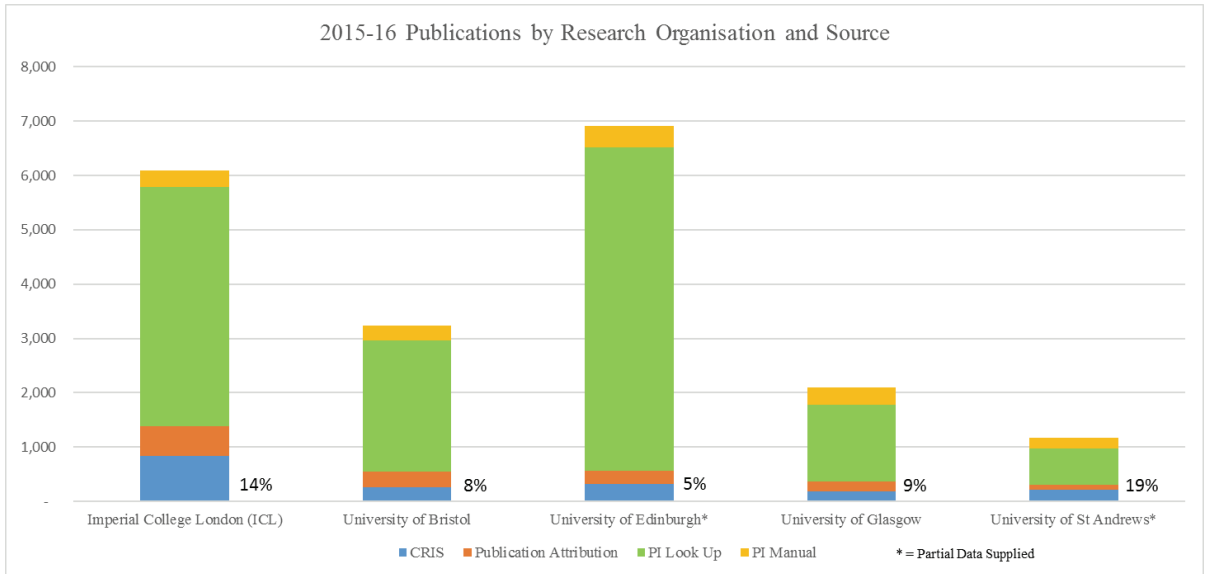


Fig. 4. Publications Imported from CRIS, Publication Attribution or PI Report in Researchfish.

Publication attributions added by automatic processes are sometimes subsequently removed by the award holder. This may be for a variety of reasons including errors in adding the data, errors in the manuscript processing, mistakes in the recording of the award reference in the manuscript, etc. Publications added as part of the pilot were monitored to see how many of them had been removed by the end of the submission period in March. Around 4% of publications added either from publication attribution or CRIS were removed. There was no significant difference in the disattribution rates, though the proportions between the pilot universities varied as can be seen in Figure 5 below. Caution should be exercised in interpreting these figures as the absolute number of publications added by some universities was so low that big differences might be the result of a few individuals. Particular care should be taken with the University of Cambridge data where roughly three quarters of the publications imported came from a single department (the Medical Research Council Epidemiology Unit at the University of Cambridge) though these made up only roughly one third of the University of Cambridge disattributions.

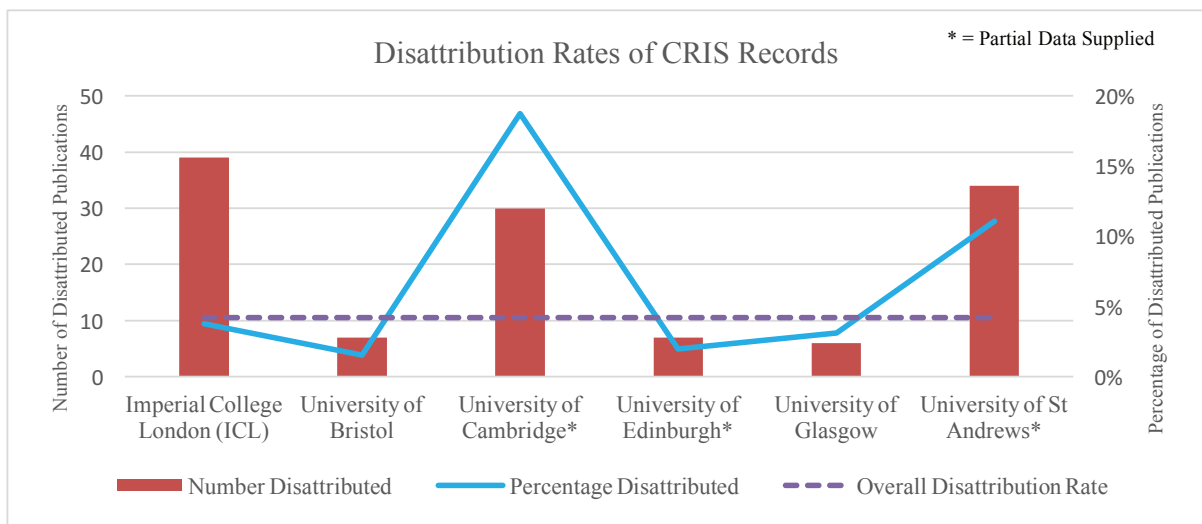


Fig. 5. Disattribution of CRIS Publications January-March 2016 in Researchfish.

3. Feedback, conclusions and next steps

Returning to the original aims of the pilot:

- The ability of Universities to provide suitable exports of grant linked publications data to a level of quality required for submission into Researchfish®.

This was achieved. The data quality provided by pilot research institutions met the requirements for import.

Although the number of outputs was small – largely due to the very short timeframe in which to prepare outputs for phase 1 – only a handful of the supplied data were invalid i.e. grant reference and/or publication ID was incorrect.

- Monitoring for any changes in researcher behaviours with respect to reporting non-publication outputs separately via Researchfish®.

The pilot institutions gathered feedback from their researchers asking the following:

1. Did you notice that we had bulk uploaded some items to your awards prior to the submission period?
2. Did the transfer of information from [CRIS/IR] to Researchfish save you time in your Researchfish submission? If yes, how much?
3. Were there any issues arising specifically from the pilot?
4. Did you return more / less / the same information in other areas as you would have without the transfer?
5. Do you want this transfer to be continued in future? Why / why not?
6. Has your opinion of Researchfish changed as a result of this pilot?
7. Do you have any further views on the pilot?

This sample size involved was small but the overwhelming feedback was that time was saved, the researchers wanted the transfer to be continued in the future and that researchers returned the same, or more information in other areas (i.e. outputs and outcomes other than publications). See Appendix A for summary and example quotes from feedback received.

Actual analysis of the data in Researchfish is inconclusive due the small size of the data uploaded in phase 1. More analysis should be possible following the second phase, which should include more publication data from all of the six universities.

3.1. Next Steps for the Pilot

Given the positive outcome of phase 1 of the pilot the next phase will automate checking and import of the publication-grant links. This will require new functionality in Researchfish to automate upload and validation of csv data, allow multiple upload iterations and allow uploads closer to the start of the submission period. The pilot will include all funders using Researchfish and the six institutions will aim to extend capture of new publication-grant links to all Departments/Faculties.

A decision has been made not to expand the number of pilot institutions in phase 2. This will ensure that if there are any problems with the new more automated process then Researchfish will still have the capacity to revert to the phase 1 manual upload process, which is considerably more time intensive.

If phase 2 is successful and does not adversely affect other information reported by researchers, then the automated bulk upload could be made available more widely.

Alongside the pilot, Researchfish has been working with ORCID and introduced the ability to associate ORCID and Researchfish accounts early in 2016. This enables publications linked with an ORCID to be passed into Researchfish and vice versa. The publication-grant reference links are not currently passed between these systems as this is not currently part of the ORCID data model. However, this work is recognized as important and discussions are ongoing with ORCID as to how this might be expanded. If successful, this could provide a more robust way (through the use of APIs rather than csv download/upload) of transferring publications already linked to research grants in institutional CRIS to Researchfish. University CRIS systems have already, or are in the process of integrating with ORCID. For example, Pure now pushes content to ORCID if, of course, the researcher agrees for this to happen.

In addition to passing information from institutional CRIS to Researchfish there is also the possibility of taking information from Researchfish and adding it to institutional CRIS. Imperial College London has used the data from Researchfish to help backfill its CRIS, as has the University of St Andrews.

3.2. Widening the Discussion; DOIs and Funding Data on Acceptance?

As can be seen from Figure 4 there is a considerable gap between those publications reported to funders and those publications containing attribution to that funding. If this information was included in the publication itself, in line with funder policies, then researchers would not have to do anything further in any other system as this data would be available for all to use.

A recent sector-wide consultation was carried out by Crossref on the merits or otherwise of assigning DOIs before the online availability of publications e.g. on acceptance. There were 104 responses and 90 were positive, 7 neutral and 7 negative. If all stakeholders in scholarly publishing can support the implementation of workflows to achieve the provision of DOIs as early in the process as possible, capturing the metadata, including funding information, and linking to that DOI, this would help support the business requirements of institutions and funders and avoid duplicate data entry.

4. Beyond bulk upload of publications with DOIs and PMIDs

Publications account for approximately half of the outcomes recorded in Researchfish and successfully implementing interoperability for publications for all funders and research organisations will be a big step in reducing duplication of effort across the sector.

Some institutional CRIS/research systems also contain information on additional outcomes with overlap with those required by the Common Question Set used in Researchfish. This ranges from publications without DOIs e.g. Books, Chapters, Working Papers through datasets, impact, further funding, engagement activities and IP. There is also work ongoing across the HEI community on reviewing and extending data models to collect more structured research activity, outcome and impact information in institutional CRIS/research systems. The time seems opportune therefore to continue to work together on agreeing standard definitions for additional research outcomes as a step towards further interoperability, whilst remaining mindful of the challenges created by exchanging data without unique identifiers.

The recent kick-off meeting for the CASRAI-UK chapter (16th May 2016), attending by over 50 HEIs, highlighted research outcomes reporting, and specifically Researchfish, as one of the top 5 priorities which could benefit from an open standards-based approach to agreeing definitions. The others were Open Access APCs, Organisational Identifiers, Research Excellence Framework and Research Career level nomenclature.

During the project, as well as research outcomes reporting, the benefits of exchanging data with grant management systems (such as the RCUK's JeS system) has been highlighted. The Department for Business Innovation and Skills is currently working with its partner organisations on the design of a new single grant management system, which may provide an opportunity for greater interoperability in future. The primary benefit of interoperability with this system would be direct submission of grant applications from an institutional pre-award research management system. In relation to research outputs reporting, automated approaches to update Principal Investigator status codes (response codes) in Researchfish directly via an API, rather than via the current circulation of spreadsheets, would be of benefit.

Appendix A. Feedback from researchers involved in the pilot

Methodology for Institution 1: Those academics that had the most publications transferred for them as part of the pilot were contacted individually by the university research office for feedback, so this was not a group that represented all academics that had any publications uploaded for them. Where a time could be arranged, a telephone interview was conducted whereas responses were obtained by email in cases where this was not possible. There was a very limited timescale to obtain this feedback so the volume was low – eight responses were received from fifteen academics contacted. Quotes that the collector feels are representative of the range of responses received and were pertinent specifically to the interoperability pilot are given below beneath each of the questions asked.

Methodology for Institution 2: All (current) academic staff who had made links between their publications and grants in our CRIS, and for whom the publication-grant links were transferred to Researchfish as part of the pilot, were invited to provide feedback via a short SurveyMonkey survey. A total of 114 academics were contacted by email and 14 responded – a response rate of 12%. A representative selection of responses are given below.

Q1 Did you notice that we had bulk uploaded some items to your awards prior to the submission period? [Inst 2 only]

11 x "Yes"

2 x "No"

1 x "I don't remember, but I think so"

Q2 Did the transfer of information from [Symplectic/Pure etc] to Researchfish save you time in your Researchfish submission? If yes, how much?

"[The interoperability pilot] saved a lot of time and having it automated really is the way to go. I'd say it saved me personally several hours and then time saved from chasing others to upload their stuff."

"...this scheme is helpful to reduce the time required, especially when there are many papers and grants to deal with."

"Hard to say but it saved a lot of pain of repetition"

“An hour ish”, “2-4 hours”, “A lot!”

“Probably 10 min- 10 papers per year; so 1 - 2 hours”

“Hours!!!!”, “significant”, “30 mins”, “20 minutes”, “2h”,

“No idea”

Q3 Were there any issues arising specifically from the pilot?

“Not really, I thought it worked very well.”

“No problems with the pilot.”

“I think things become more complicated when, as in our case, there are multiple co-investigators on a grant.”

Q4 Did you return more/less/the same information in other areas as you would have without the transfer?

“Much more, indeed without this link I doubt we would have come anywhere near getting this level of engagement.”

“More or less the same information”

4 x More (31%)

0 x Less

9 x The same (69%)

Q5 Do you want this transfer to be continued in future? Why / why not

“It is practically essential to have this interface.”

“Yes - saves time and is more accurate.”

“Yes - because there should be no double-entry of data!”

Q6 Has your opinion of Researchfish changed as a result of this pilot?

“Yes, for the better”

“Yes, I'm happier to participate as it's a whole lot easier now”

“Not really”

Q7 Do you have any further views on the pilot?

“My experience is that Symplectic does a pretty good job of catching publications, so if everything can essentially be done through that single interface, so much the better.”

“...The maximum automation makes sense and also ensures some level of accuracy.”

"...We log our publications in our CVs, in Pure, so doing that in any other place is a complete waste of time and holds up income generating activities."

"...Having the bulk upload massively reduces the concern that I have (or have not) captured all the information"

"Please, please, please carry on minimising workload that gets pushed down to academic. After a little tussle, Pure is an excellent [database] and resource that helps with such uploads/exchanges. A one-stop-shop for this sort of thing is much appreciated..."

"I was relieved that the pilot took place, it was very successful and should, in my opinion, continue."

References

1. Acknowledgement of Funders in Scholarly Journal Articles. (2008, February). Retrieved May 20, 2016, from <http://www.rin.ac.uk/system/files/attachments/Acknowledgement-funders-guidance.pdf>
2. Adding a publication in Researchfish, (2006). Retrieved May 29, 2016 from <https://youtu.be/tgwyXDGqeuQ?t=249>
3. Author Responsibilities - Conflicts of Interest. Retrieved May 20, 2016, from <http://www.icmje.org/recommendations/browse/roles-and-responsibilities/author-responsibilities--conflicts-of-interest.html>
4. Best practices for depositing funding data. Retrieved May 27, 2016.
5. <http://blog.crossref.org/2015/09/best-practices-for-depositing-funding-data.html>
6. CASRAI-UK First Chapter Meeting a Success. Retrieved May 27, 2016.
7. <http://symplectic.co.uk/events-conferences/hosted-casrai-uk-chapter-meeting/>
8. Community responses to our proposal for early content registration. Retrieved May 27, 2016, from <http://blog.crossref.org/category/accepted-manuscripts>
9. Conditions of Research Council Training Grants. (2016, April). Retrieved May 20, 2016, from <http://www.rcuk.ac.uk/documents/documents/termsconditionstraininggrants-pdf/>
10. Graf C, Wager E, Bowman A, Fiack S, Scott-Lichter D, Robinson, Andrew. (2007, 1). Best Practice Guidelines on Publication Ethics: a Publisher's Perspective. *Int J Clin Pract* 2007; 61(s152): 1–26
11. Guide for Authors – Research Policy. Retrieved May 20, 2016, from <https://www.elsevier.com/journals/research-policy/0048-7333/guide-for-authors#12000>
12. Policy for open access in the post-2014 Research Excellence Framework. Updated July 2015. Retrieved May 27 2016.
13. http://www.hefce.ac.uk/media/HEFCE,2014/Content/Pubs/2014/201407/HEFCE2014_07_updated%20July%202015.pdf/
14. Publication, Policy and Reporting. Retrieved May 20, 2016, from <http://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/conditions-of-your-grant/publication-policy-and-reporting>
15. RCUK Policy on Open Access and Supporting Guidelines. (2013, April). Retrieved May 20, 2016, from <http://www.rcuk.ac.uk/documents/documents/rcukopenaccesspolicy-pdf/>
16. Research publication acknowledgement practice: Guidance for authors. Retrieved May 20, 2016, from <http://www.wellcome.ac.uk/managing-a-grant/end-of-a-grant/wtd037950.htm>
17. Tickell, A, Open access to research publications. Independent advice. (2016, February). Retrieved May 27, 2016, from <https://www.gov.uk/government/publications/open-access-to-research-independent-advice>
18. Starter Grants for Clinical Lecturers. Retrieved May 20, 2016, from www.acmedsci.ac.uk/download.php?f=file&i=28486