JISC Researcher Identifier Task and Finish Group.
Researcher Identifier Recommendations – Sector Validation

July 2012

Executive summary
Our study found widespread support for the implementation of researcher identifiers in the UK and strong but not unanimous support for ORCID as the most suitable candidate for achieving this.

There are some compelling reasons why a researcher identifier is appropriate at this time. For most (not all) researchers, there is a reasonable chance of drawing a boundary around their scholarly work and involvements and an understanding that there is a benefit in allowing researchers to choose to make certain elements of that information more publicly available.

We illustrate some misunderstandings and misconceptions which ORCID and those recommending UK wide implementation will need to address through clear public outreach.

Any researcher identifier will require a compelling demonstration of use case and interface, ORCID will need to demonstrate to individual researchers its power and potential and the benefits it can bring to them.

Publicity should be given, in particular to the minimum data required and the control which the researcher will have over the privacy of each associated element. ORCID will need to make it clear that the ID is a tool which in itself does not confer status.

We would welcome a clear statement on interoperability with ISNI but we do not consider that complete merger is a prerequisite.

We recognise the need to “seed” the ORCID database and in this context we recommend that, at least in the short term, the Names project should receive sufficient funding to share not only their data but also their expertise and experience. There may be a medium term role in cleaning UK data before bulk uploads.

We include a risk register with 21 mitigating actions for consideration.

In the light of the Task and Finish group’s stated desire for a contingency plan, we recommend that JISC encourage the experimental use of UUIDs as an underlying layer as suggested by CASRAI.

Many thanks are due to those who contributed their time and expertise to this report. Over a short period of time we received detailed responses from 142 people and conducted interviews with 27 stakeholders.
Responsibility for the content of the report is the authors’ alone.

Authors: Nicky Ferguson, Helen Lawton Smith, Richard Moore, Seb Schmoller
Contact: nicky [AT] clax.co.uk text/sms: 07770-262886 www.clax.co.uk
Contents

1 Background to this report........................................................................................................3

2 Introduction and background to the issue..............................................................................3

3 Methodology and tools used....................................................................................................4

   3.1 Exploration of possible costs and potential benefits.................................................. 4

   3.2 Questionnaire .............................................................................................................. 4

   3.3 Interviews .................................................................................................................... 4

   3.4 Validation workshop...................................................................................................... 4

4 Potential costs and benefits of implementing the recommendations.................................4

5 Insights from the questionnaire.............................................................................................9

   5.1 Respondent characteristics ....................................................................................... 9

   5.2 Response bias ............................................................................................................ 10

   5.3 Headline findings........................................................................................................ 10

6 Issues arising from questionnaire and interviews.................................................................13

   6.1 Competing or overlapping standards ......................................................................... 13

   6.2 Privacy and security concerns .................................................................................. 16

   6.3 Scope .......................................................................................................................... 19

   6.4 Contingency plan......................................................................................................... 20

   6.5 Non-commercial / Open source .............................................................................. 20

   6.6 Researcher ownership ............................................................................................. 21

   6.7 Seeding versus claiming ........................................................................................... 21

   6.8 Administration in the UK and the role of HESA ....................................................... 21

7 ORCID Risk Assessment and Mitigation...............................................................................22

Appendix 1 - JISC Researcher Identifier Initiative....................................................................28

Appendix 2 - VIAF and other IDs .......................................................................................... 29

Appendix 3 - questions and answers on ORCID................................................................. 31

Appendix 4 – detailed data from the questionnaire ............................................................... 34

Appendix 5 - Interviewees .................................................................................................... 38
1 Background to this report

This report was written in response to a need to validate the recommendations\(^1\) of the JISC-convened Task and Finish Group (TFG) on researcher identifiers. It builds on the detailed stakeholder interviews and analysis already undertaken by Nicky Ferguson for the Task and Finish Group for the stakeholder use cases\(^2\) (PDF) and identifier needs\(^3\) (PDF) reports. The full background to the group, its membership and details of the reports they commissioned appear in Appendix 1.

Although the report is written for JISC as the client, our brief states that the report should be written for use in the community and as dissemination of the relevant key issues.

For this reason, background information and pointers to further information have been included which will already be familiar to the Task and Finish Group but which we hope will be useful for a more general audience.

2 Introduction and background to the issue

Researchers not only need to identify themselves uniquely but they also need to attach their identities uniquely to a number of things: datasets, equipment, outputs, articles, media stories, citations, experimental notebooks etc. The number and diversity of the different systems and the different identifiers they currently use is very high, extremely time consuming and very frustrating. A researcher wants to unambiguously assert: it was me and no one else who did that at that place and that time and the results were xyz. The ultimate aim would be to have a digital academic passport of academic activity I did x experiments, published a articles, generated d datasets, etc. Researcher identifiers could be an important component for any system which records and shares observations of research work. It would also facilitate the collating and merging information from different sources which would be necessary for such a digital academic passport. In the future, it might also, for example, be possible to use researcher identifiers to record or discover who set up equipment or was present in a laboratory during an experiment.

Unique identifiers for individuals involved in research are a central component of many other processes. An identifier solution would enable, for instance:

- Disambiguation of authors; assigning publications to a specific individual would be possible automatically on both a national and global scale.
- Career tracking of researchers, from post-graduate research to emeritus professor, across institutions.
- A reduced burden on researchers by eliminating duplication of effort and enabling easier entry on forms for funding, appraisal, collaboration and publishing.
- Clarification of staff institutional affiliations; which can, for example, be used to confirm eligibility for access to international facilities.
- Creation and sharing of consistent, up to date and accurate CVs and records of achievements.

---

\(^1\) http://tinyurl.com/855exyt
\(^2\) http://ie-repository.jisc.ac.uk/568/1/report1-final.pdf
\(^3\) http://ie-repository.jisc.ac.uk/570/1/report2-final.pdf
3 Methodology and tools used

3.1 Exploration of possible costs and potential benefits

In addition to consulting within the UK research community in order to validate the TFG’s recommendations, JISC asked us to explore the business implications, costs and benefits of the potential adoption of ORCID. We asked Helen Lawton Smith\(^4\) to join the team to provide the expertise necessary for this and her analysis appears in Section 4.

3.2 Questionnaire

We created an online questionnaire which took respondents about 30 minutes to complete. Its contents were checked and validated with a number of academic contacts and with JISC, before it was launched on 24 May 2012. After a considerable social media effort to bring it to the attention of a variety of stakeholders, particularly researchers who had not previously considered the issues relating to the implementation of a unique researcher ID system, it closed after two weeks with 142 detailed responses from respondents who had fully completed the survey. We were pleased that a relatively high number took the trouble to consider and carefully reply to our quite demanding questions. We allowed free text responses at most points and these were most informative. A summary of the questionnaire results appears in Appendix 4 and the results of our analysis of the numerical data from the questionnaire are discussed in Section 5.

3.3 Interviews

We conducted interviews with stakeholders from a variety of backgrounds within and outside HE. The interviews generally lasted around 40 minutes and, with the respondents’ permission, most of them were recorded and transcribed. We do not discuss the contents of the interviews separately in this report, but of course they inform the whole work and are quoted or summarised where appropriate. A table of interviewees appears in Appendix 5. The topic guides with which we structured the interviews have been archived\(^5\).

3.4 Validation workshop

The interviews included a discussion with a group of stakeholders attending the JISC/CNI Workshop in Birmingham on 5\(^{th}\) July and we are particularly grateful to those present for helping to clarify important issues and concerns at short notice and at a late stage in the report writing.

4 Potential costs and benefits of implementing the recommendations

The headings for this section are taken directly from our brief as expressed in JISC’s invitation to tender.

1. **Assess the feasibility and potential costs of national scale membership and use of ORCID and how they might best be met.**
   - If the UK were to adopt an internationally recognised system such as ORCID there would be considerable national benefits – as outlined elsewhere in this report. With the international scope and ambition of ORCID, in addition to the intra-UK advantages, there will also be advantages for international

\(^4\) [http://www.geog.ox.ac.uk/staff/hlawtonsmith.html](http://www.geog.ox.ac.uk/staff/hlawtonsmith.html)

\(^5\) [http://www.webcitation.org/691rOX9J4](http://www.webcitation.org/691rOX9J4)
[Topic Guide 1](http://www.webcitation.org/691rOX9J4)
[Topic Guide 2](http://www.webcitation.org/691rRpgRN)
[Topic Guide 3](http://www.webcitation.org/691rUnLO6)
researchers coming to the UK and for UK researchers moving elsewhere within the ambit of the scheme.

- Benefits to researchers will be reduction of duplicate form-filling with different funding bodies, publishers and institutions currently requiring the same or similar information in slightly different formats. Researchers should also find easier access to authoritative information on their own and others’ publications, participation, outputs and achievements.

- Major benefits will arise to organisations within the UK because bodies such as research funders, institutions, publishers and sectoral bodies will be able to exchange data more easily.

- For this to be feasible there needs to be a considerable buy-in by the major funders and agencies. If these were all to make it a condition of interaction (as with the research councils move to Je-S6) then institutions will follow and then individuals will necessarily take part. In the UK, as in academia in general, the best drivers are often those of being left out or prevented from doing something necessary for your job unless you act.

- The minimum that is required is to get the buy-in of at least the organisations represented by members of the Task and Finish Group (and their equivalents in other countries important for research collaborations). Indeed, it may be worth piloting the process in a part of the UK with its own funding bodies in the first instance, perhaps Scotland through SFC, or through a single discipline.

- The involvement of HESA is essential. Although the questionnaire reflected that:
  
  a) HESA does not represent and would not claim or try to represent researchers
  b) many (most) researchers do not have direct contact with HESA and are not aware of their HESA staff ID
  c) individuals are often hostile to monitoring systems when their purpose is poorly understood and concerned about potential perceived misuse of data

- nevertheless HESA’s expertise in understanding how to aggregate data in a valid way, understanding the sensitivities and need for probity, and in knowing how to collate and publish results appropriately will be key to achieving a fair system. In addition, widespread adoption of ORCID is unlikely to occur unless HESA uses it (and unless the ID is required for any post 2014 REF exercise or its successors).

- There are issues about where the data is held. The UK government and sectoral bodies may be sceptical about a system where key information is stored in the US. For this and other important reasons, it is important to emphasise that the amount of data held centrally can be minimal, with UK based organisations using the central store for registry, claim, maintenance and assignment purposes but keeping key data under their own control with their own policies on sharing and permissions (informed by the individual researcher’s expressed permissions) see figure 1.

- There is likely to be resistance from some institutions (although we did not encounter any institutional opposition) and individuals (we did encounter some individual opposition). However, it should be noted that at 29th June (before the formal launch), 58 UK organisations had signed up as Participating Organizations, 16 of these being academic institutions/organisations and a further 6 UK learned societies.

- Costs will be those of converting a relatively small number of IT systems. On the basis that say 50 systems need to be converted and the rest follow

---

6 [https://je-s.rcuk.ac.uk/](https://je-s.rcuk.ac.uk/)
7 See [http://about.orcid.org/civicrm/profile?force=1&gid=10&country-Primary=1226&crmSID=3_u&crmRowCount=100&crmPID=2](http://about.orcid.org/civicrm/profile?force=1&gid=10&country-Primary=1226&crmSID=3_u&crmRowCount=100&crmPID=2) for latest information
because the same systems are used across institutions then perhaps an initial estimated cost of £1.5 million across the sector may be reasonable. In practice, HESA is better placed to estimate this figure through their existing contacts with relevant suppliers of software. In addition, if organisations wish to use the central system for bulk uploads of names and downloads of IDs then there will be subscription or licensing costs from ORCID. Although the costs of ORCID subscriptions have yet to be announced, they are likely to be annual costs, per large employer or research host, of between 5,000 and 10,000 US dollars depending on the size of the organisation. Use by individual researchers (attached or unattached) for claiming IDs and maintaining information, agreeing, challenging, refuting and setting permissions will be free.

figure 1 – sketch of processes and where they happen. P = permission

2. **Review minimum data elements to ensure that they will be fit for the range of purposes cited in the use case reports.**

There does not appear to be a definitive statement of the full set of data elements or fields. Name and email address are the minimum data elements or required fields, but may be kept private. Additional fields include other names, other email addresses, organization affiliation(s), degree, other identifiers, and research objects such as publication, patents, and awarded grants. Individuals may control privacy settings at the data element level. With the information currently available it seems that the data elements will be fit for the range of purposes cited in the use case reports. However, it is in the nature of developments that required changes may emerge should the UK proceed with the recommendations.

3. **Assess possible costs in providing this minimum data.**

If a majority of institutions were to participate, then the costs will be for each institution are likely to be:
• Costs of assigning ORCID identifiers. This will depend on the systems in use but could for instance be of the order of £5K, plus the time of the individuals which might be £15 per person\(^8\).

• Costs of converting existing records to use, as a secondary key, ORCID numbers. This is likely to be £5-10K per organisation\(^9\).

• On-going costs of maintaining the institutional data to allow for researchers coming from elsewhere, starting within the institution and communicating with others using ORCID aware systems. This will depend on the scale of adoption. Organisations that provide systems for universities in this area may make the changes providing there is widespread acceptance and provided that they are properly briefed by one of the key organisations that negotiates with them for the system as a whole. This means currently JISC or HESA. If the costs end up falling to the institutions on a one off basis then it is much harder to estimate.

• Software costs initially are potentially of order of magnitude £25K\(^10\).

• On-going costs might be around £10K per annum\(^11\).

These are ballpark figures and depend on degree of uptake: the more widespread the uptake the lower the overall costs per researcher will be. And furthermore, as we indicate in 7 below, we give no estimates for the savings in researcher time and other broader benefits that should flow from the widespread adoption of ORCID. We note that many institutions already have internal systems that perform the same or similar functions as a result, if ORCID replaced these systems the cost of maintenance would be budget neutral.

---

\(^8\) We base this on it taking an average of 30 minutes per person to create an ID and on average direct employment costs of £200/day. The figure of 30 minutes should be considered an average that includes initial set up and will vary across academics with differing volumes of activity. In our interviews we heard that existing equivalent systems, once established, can be operated for an average academic in under 5 minutes a week. Initial setup time will require a user to understand what needs to be provided and not provided, get to the relevant part of the system and enter personal details as outlined in 2 above. Some individuals will take time to look up their degrees and other identifiers. For some users, 30 minutes may prove to be an over estimate but for many, who will require explanation and rationale, it will be an under estimate. Once a record has been established and the end user is familiar with the operation of the interface, the average time taken to maintain the system will be proportional to the volume of records submitted, plus any identification and removal of erroneous data.

\(^9\) The processes which institutions use to make changes to their institutional information systems require formality. This includes design, validation, sign-off, testing and implementation. A £5k change corresponds to five person weeks and is crudely estimated as being the result of changes to five institutional systems with a minimum cost of one person week per change. These changes are not likely to be large in most cases once a system has been established and so the upper bound is likely to be no worse than twice the minimum case.

\(^10\) The initial software development and implementation costs for an organisation that has developed its own software will involve building and testing interfaces to, on average, five existing internal structures. Each of these, when fully tested and user trialled, will involve approximately five person weeks of effort overall, giving a cost of £5k per system based on £1k per week. (This assumes that software development is undertaken in-house).

\(^11\) Having acquired individual interfaces to the institution’s systems, in each year there are likely to be changes to the specification, these are intended to be minimised through the use of an encapsulated API. However, the larger activity cost is likely to be associated with supporting end users of the system. A support and minor modification cost of £2k per system i.e. two person weeks is modest but achievable across a set of systems provided system change is modest and is communicated well in advance.
4. Consider national or sector-wide subscription to commercial products to reduce costs and administrative burden.
   - As indicated above, national or sector-wide subscription to commercial products should be negotiated through HESA or a similar level organisation. There will be a variety of HR and finance systems to be modified.
   - It seems unlikely that a new third party will be able to establish a niche in this area.

5. Assess the potential for bulk uploads from institutions and other organisations to enable those that so wish to obtain IDs for their staff en masse.
   - This is one of the planned benefits of the ORCID membership but the system is not yet fully in place.
   - The problem with an initial upload of this type is that it disintermediates the individual researcher at an early stage. That makes it more likely that they will not keep their records up to date and may also be a disincentive to researcher participation.
   - Thus while it would be easy to invent a system involving only institutions and ORCID and funders, the real problem of maintaining it will not thereby be addressed. It was stressed in many of our interviews that researcher participation is absolutely key to the success of this venture. If researchers perceive it as another bureaucratic imposition then the aim of exchangeable and up-to-date checked information is unlikely to be achieved.

6. Establish the level of coverage of the HE community that would result in the greatest practical benefit
   - Clearly the maximum benefits flow if everyone plays the game. However, this may be unrealistic.
   - Considerable work is necessary on how to incentivise different groups and individuals: this point is expanded elsewhere in the report.
   - There are various stages in a research worker's career and incentives are needed not only for the new and mid-career researcher but also the experienced and senior researchers.
   - If the researcher perceives that participation will enhance visibility and career prospects then there is likely to be significant take up. If the system is perceived as being merely an aid to institutional processes then there may be an element of "paucity of commons", where it seems to be best for any individual researcher if everyone else's data is there but not their own. Such subtleties in perception of value of the system are important.

7. Estimate the scale of any potential benefits to UK HE of national scale membership and use of ORCID.
   - Our research to date has not been able to quantify this benefit other than to say that it is perceived by the respondents to the survey and the interviewees as substantial. The best estimate might be summarised by one of our interviewees with experience working at a senior level in UK university and research council contexts:

   I know that there are multiple people who spend an awful lot of their time just trying to track down publications generated out of the organisation [research council]. And those people could be much better deployed doing other things. And as soon as you're talking about multiple people, you're talking about hundreds of thousands of pounds a year, that potentially... not removed all together, but can be significantly reduced. And that's one UK organisation. So if you can save half a person say at an organisation like [a research council], I think that's pretty conservative. And you can do something similar at each university in the UK, then you only need about ten institutions to cover the entire running costs of ORCID for a year. So it seems to me that modest membership, modest sort of infrastructure support payments can cover the costs.
Outside of HE, but of direct relevance to it, it is thought that publishers spend millions annually on disambiguation and deduplication, costs which presumably are passed on to HE as customers in one way or another. Publishers’ enthusiasm to support ORCID might indicate that this theory has some validity.

8. Assess timescales for costs/benefits

- The short-term costs will outweigh the benefits as it will take time for the system to become operational and then achieve critical mass. In addition, the lead time involved in changing such things as the HESA Staffid can be considerable because of the consultation involved and the need to brief software suppliers, who then need to make changes.
- It seems unlikely that any overall benefits will be seen in the first three years.
- However, in the medium 3-5 year period, real benefits would emerge as part of a national/international system.
- Full benefits should be achieved from five years onwards, assuming widespread uptake.

5 Insights from the questionnaire

We opened a non-anonymous online questionnaire using Survey Monkey between 24 May and 7 June 2012. Data summarising responses to the questionnaire is in Appendix 4. The questionnaire had four main sections covering:

- the purpose of ORCID;
- demographic and role-related information about respondents;
- Likert-scale questions relating to different aspects of the goal of widespread uptake of ORCID;
- Likert-scale questions relating to the process of reaching the goal of widespread uptake of ORCID.

Each block of Likert-scale questions was followed by an optional text-box.

At the end of the survey respondents were required to indicate the basis on which any comments they provided could be included in our report.

616 people opened the survey, of whom 222 (36%) started it by completing the first question. Of these 141 (63.5% of starters) completed the survey by responding to the permissions question at the end of the survey.

5.1 Respondent characteristics

Respondents were:

- overwhelmingly based in the UK (95.6%) and responding as an individual (95%) rather than on behalf of an organisation;
- mainly based in UK HE (81%);
- mainly researchers (70.5%), but with substantial representation from those managing researchers (15%), managing research related services (13%), managing data about researchers and research (14%).

There was rather smaller representation from those running relevant software systems (8.5%), making relevant software systems (5%), or employed in scholarly publishing (3.5%).

The disciplinary spread was reasonably wide, with economic and social sciences being the largest disciplinary category (22%) and management and finance being very much the smallest (1.4%).

84.5% of respondents were based in organisations employing more than 100 staff and 67% in organisations employing more than 1000 staff. Respondents were evenly spread across

---

12 We have archived a copy of the full text of the questionnaire at http://www.webcitation.org/69091lU4Z
four main ten-year age bands between 26 and 65, and between men (54%) and women (44.5%).

5.2 Response bias
The sample is widely enough drawn for the views expressed to provide useful insights. But 142 respondents to an open survey cannot be viewed as a representative sample of UK researchers; furthermore we believe that it is safe to assume that respondents reached using social media and through JISC and other online networks would tend to be more supportive of ORCID than would be the research population as a whole. For this reason we have been conservative in our interpretation of the data by defining validation/agreement using the following arbitrary boundaries, which are colour-coded in Appendix 4:

5.3 Headline findings
We probed the data for differences of view between different categories of respondent:

- Organisational representatives (n=7)
- Women respondents (n=63)
- Individual researchers (n=100)
- Managers of data about researchers and research (n=20)
- Respondents under 36 years old
- Non-STEM respondents (n=62)
- STEM respondents (n=61)

These differences are represented in Appendix 4 – and below – by “carpets” with colours representing the boundary between different degrees of validation/agreement.

5.3.1 The goal (questions 14 – 18)
Q14. Overall there was very strong, strong or mild agreement that each of the following actors "would benefit considerably from the widespread adoption of ORCID or an equivalent solution":

<table>
<thead>
<tr>
<th>Actor</th>
<th>Likert average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researchers</td>
<td>4.16</td>
</tr>
<tr>
<td>Organisations that undertake or host research</td>
<td>4.30</td>
</tr>
<tr>
<td>Publishers</td>
<td>4.01</td>
</tr>
<tr>
<td>Research funders</td>
<td>4.33</td>
</tr>
<tr>
<td>Scholarly societies</td>
<td>3.79</td>
</tr>
<tr>
<td>Statistical agencies</td>
<td>4.28</td>
</tr>
</tbody>
</table>

Unsurprisingly, organisational representatives and managers of data about research and researchers were somewhat more in agreement with the question 14 statement than the sample as a whole, as were STEM respondents. Conversely Non-STEM respondents were somewhat less in agreement with the question 14 statement, as was the case with respondents aged less than 36. There was almost no gender difference in responses.
Q15. The questionnaire suggested that “the widespread uptake of ORCID or an equivalent solution is intended to bring a number of benefits”. Respondents overall were either in strong or mild agreement as to the importance of each of a range of nine benefits.

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Improvements to the way in which research is administered</td>
<td>3.85</td>
</tr>
<tr>
<td>2. Simplified sign-on procedures to research-related systems</td>
<td>4.07</td>
</tr>
<tr>
<td>3. Reduced form-filling during the process of applying for grants</td>
<td>4.24</td>
</tr>
<tr>
<td>4. Better and more complete tracking of individual researchers’ careers</td>
<td>4.08</td>
</tr>
<tr>
<td>5. Improvements in the way research collaborations are recorded</td>
<td>3.93</td>
</tr>
<tr>
<td>6. Improvements in the way in which publications, grants, research projects, and researchers are mapped and linked</td>
<td>4.21</td>
</tr>
<tr>
<td>7. Better interoperation of local and national systems that encode and store data about researchers</td>
<td>3.85</td>
</tr>
<tr>
<td>8. Easier creation of authoritative lists of publications, citations and CVs</td>
<td>4.20</td>
</tr>
<tr>
<td>9. Smoother exchange of data about researchers between institutions during the preparation of collaborative bids, or when staff move between employers</td>
<td>3.91</td>
</tr>
</tbody>
</table>

Managers of data about researchers were more in agreement than respondents as a whole that ORCID would bring benefits across all nine of the benefits suggested. Interestingly, non-STEM respondents were in very strong agreement concerning the importance of 2. Simplified sign-on procedures to research-related systems and 3. Reduced form-filling during the process of applying for grants. There was inconclusive agreement amongst younger respondents (aged less than 36) concerning the importance of 1. Improvements to the way in which research is administered and 7. Better interoperation of local and national systems that encode and store data about researchers. There was almost no gender difference in responses.

Q16. Respondents overall were only mildly in agreement with the suggestion that “the Higher Educational Statistical Agency (HESA) or an entity like it should be the kind of trusted body to administer and integrate the use of researcher identifiers with currently used staff and postgraduate IDs in UK; and to advise institutions on data sharing issues”, whereas managers of data about research and researchers were strongly in agreement with this suggestion. In contrast, the extent of agreement amongst STEM respondents was inconclusive.

Q17. Respondents overall were in strong agreement that “any solution must be capable of covering all participants in UK research, whatever their contribution or standing, including those not in traditional research roles”. STEM respondents were in very strong agreement with his statement.

Q18. All categories of respondent were in strong agreement that “alongside ORCID (or any equivalent solution), if it becomes operational, there will need to be a suitable UK wide contingency plan to ensure the continuity of ORCID’s functions if for any reason ORCID ceases or fails”, with respondents aged below 36 very strongly in agreement.

5.3.2 The process of reaching the goal (questions 20 – 22)

Q20. All categories of respondent were in strong agreement that “prior to the ORCID (or any equivalent solution) being adopted by my organisation a business case to justify it will be needed”, with organisational representatives very strongly in agreement with this statement. Women respondents were noticeably more strongly in agreement with this statement than men.
Q21. We asked respondents to rate the importance of a range of seven issues within a business case for the adoption of ORCID. Respondents overall rated all seven issues as important or very important:

1. A clear summary of the practical benefits 4.72
2. Interoperability between different identity standards and between different systems that encode and store research IDs 4.31
3. Oneoff implementation costs of getting ORCID or its equivalent established within an organisation 4.21
4. Ongoing local costs of maintaining ORCID or its equivalent within an organisation 4.38
5. A clear statement from ORCID on interoperability and boundary issues with potentially overlapping standards such as ISNI and VIAF 4.15
6. Privacy and control issues 4.57
7. An assessment of the risks and benefits that might arise from implementation 4.31

In line with their responses to Q20, women respondents rated the importance to their organisation of each of all seven issues more highly than men.

Q22. We asked respondents to rate from their own perspective the extent to which the business case should be written with the eleven different organisations/categories in mind. Here the difference in importance of the different categories varied widely:

<table>
<thead>
<tr>
<th>Organisation/Category</th>
<th>Likert Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public sector organisations, including universities who employ researchers</td>
<td>4.74</td>
</tr>
<tr>
<td>Charitable sector organisations who employ researchers</td>
<td>4.22</td>
</tr>
<tr>
<td>Private sector organisations who employ researchers</td>
<td>3.93</td>
</tr>
<tr>
<td>Publishers of research</td>
<td>3.99</td>
</tr>
<tr>
<td>Research funders</td>
<td>4.33</td>
</tr>
<tr>
<td>Individual researchers</td>
<td>4.40</td>
</tr>
<tr>
<td>Professional and scholarly associations who have researchers as members</td>
<td>3.78</td>
</tr>
<tr>
<td>Government departments and agencies</td>
<td>3.55</td>
</tr>
<tr>
<td>Creators and suppliers of software systems such as HR systems that encode and store IDs</td>
<td>3.48</td>
</tr>
<tr>
<td>Politicians and planners</td>
<td>2.88</td>
</tr>
<tr>
<td>Social science researchers who are interested in, for example, the mobility of researchers</td>
<td>3.09</td>
</tr>
</tbody>
</table>

A plausible explanation for these differences is that the further a category was from the ken of respondents the less importance was attached to the category. This is supported by the fact that the organisational representatives category of respondents and the managers of data about research and researchers category gave much greater importance for the business case to “Creators and suppliers of software systems such as HR systems that encode and store IDs”:

9. Creators and suppliers of software systems such as HR systems that encode and store IDs

<table>
<thead>
<tr>
<th>Organisation/Category</th>
<th>Likert Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creators and suppliers of software systems such as HR systems that encode and store IDs</td>
<td>4.40</td>
</tr>
</tbody>
</table>

13 A similar effect can be seen in the response of economic and social scientists for “Social science researchers who are interested in, for example, the mobility of researchers”, where the Likert score is 3.40 as compared with 3.09 for respondents as a whole.
6 Issues arising from questionnaire and interviews

6.1 Competing or overlapping standards

6.1.1 ORCID, ISNI, VIAF and UUID

Each of us will have many identifying IDs – passport numbers, National Insurance numbers, HR record numbers and many other numbers which may be used for our professional, social and financial lives (and deaths). Do we need another one? If so which? Clearly it is unlikely (as well as being very undesirable), that any one of these identifiers will achieve national and international recognition such that it becomes our sole ID for all aspects of our lives\(^\text{14}\). The advantages for researchers (and their employers, publishers and funders) of a unique identifier which is nationally and internationally recognised have been spelt out in previous work and above. The Task and Finish Group’s recommendation is that ORCID\(^\text{15}\) (the Open Researcher & Contributor ID) offers the best likely solution to the sector’s needs. We explored opinion on this and other possible candidates.

ISNI – the International Standard Name Identifier - is an ISO Standard (ISO 27729) whose scope is the identification of Public Identities of parties: that is, the identities used publicly by parties involved throughout the media content industries in the creation, production, management, and content distribution chains. The ISNI system uniquely identifies Public Identities across multiple fields of creative activity. ISNIs are assigned to the Public Identities of Parties that participate in the creation, production, management or distribution of cultural goods in the digital environment. Those Parties can be natural persons (a human being like a book author), legal entities (like a Record Label) or even fictional characters (like Peter Pan).

VIAF - the Virtual International Authority File, implemented and hosted by OCLC, is a joint project of several national libraries (including the Bibliothèque nationale de France\(^\text{16}\) and the British Library) plus selected regional and trans-national library agencies. The project’s goal is to lower the cost and increase the utility of library authority files by matching and linking widely-used authority files and making that information available on the Web.

Although they remain separate, ISNI and VIAF will be working so closely together that it seems likely that to an end user, or creator, they will appear to be completely interoperable. As Hickey commented:

> ISNI will be using a copy of VIAF, along with other files, as its base file. As ISNIs are assigned to entities that have VIAF IDs, that information will be fed back to VIAF, so there will be links between the two files. Also, as ISNI processing discovers errors in the VIAF file (such as two names that should be merged) ISNI will inform VIAF.

ISNI and VIAF have their roots in a top down model of authority files. ORCID, on the other hand, is keen to promote the concept of authors “claiming” IDs, so it looks more like a bottom up model, although it does have plans to allow member institutions to do bulk uploads and assignments.

There are people from ISNI and VIAF closely involved with ORCID and therefore strong informal links between the organisations. However, discussions are ongoing about the levels of interoperation and cooperation. We spoke to Andrew MacEwan from the British Library. The BL have declined to fully endorse the recommendations of the Task and Finish Group because of their commitment to ISNI and we wanted to investigate why. MacEwan, who represents the BL and the conference of the European National Librarians on the board of ISNI and is actively engaged with ORCID at the level of the ISNI board says

\(^{14}\) See Appendix 2 for an informative discussion by Thom Hickey of OCLC on why we need multiple identifiers and on the relationship between VIAF, ISNI and ORCID

\(^{15}\) [http://www.orcid.org/](http://www.orcid.org/)

ORCID is a great thing because I think they have a lot of strengths and they’d have engagement with user organisations, institutions and the actual researchers [...] the ORCID and JISC mainspring is the direct interaction with researchers. [...] we’re very supportive of what they’re doing, we’d just like it to be one system if we can get it to work. [...] We’ve achieved the making sure that there’s no duplicates, i.e. we won’t independently assign one identical number to one researcher whilst ORCID would say assign the same number to another. Because we’ve given them a block of ISNIs, so they can’t ever assign the same number that we’ve assigned to someone else. [...] What we want is something more than that, which is if they assign an ORCID, we’ve actually drawn that into the ISNI ecosystem as well. So that it becomes the same number. And the reason we want that as the model is that we think it is better to have a single number that’s used and can be shared and linked across all the different constituencies and partners that want to share an ID. At some point we will want to make assertions that this ORCID is the same as that ISNI – sharing an identifier is another way of making that assertion and building it into the infrastructure early on. ISNI is focussed on two things: (1) linking contributed data to fix ISNI identities; (2) diffusing the ISNIs back out to contributors’ databases. So ISNIs are diffused into records for works, books, articles, datasets, etc. This will require mechanisms for managing deprecation, merging and splitting of ISNIs and appropriate reciprocal reporting mechanisms between Registration Agencies, VIAF libraries, end users and the ISNI assignment system to maintain accuracy of diffusion. Maintaining a mapped relationship to another identifier doing the same job adds another layer of work to the effort of maintenance.

Cameron Neylon is not convinced that a single number is essential, noting that, in the semantic web, a “same as” statement would accomplish a lot, providing that the work on quality, de-duplication, disambiguation etc is still happening in the appropriate places.

> it’s useful for them to be interoperable. And there’s value in pursuing the interoperability [...] But [...] we’re talking about the linked-data world, [...] to the extent that any of these systems work and provide any value at all, doing that same as assertion will give you the additional value and that’s fine; you don’t need to worry about whether there are a couple of identifiers.

From the end user point of view, it seems to us that there are two possibilities – **either**

a) I go into ORCID and claim an identity, then I’ll be told "Actually we think you’ve already got an ISNI because of xyz publications – do you want to use that for your ORCID or do you deny that you are that person or do you not necessarily deny you are that person but you want a complete new ORCID? (unusual but definitely needs to be possible for a number of reasons)

**or**

b) I go into ORCID and claim an identity, am assigned an ORCID and at some stage in the future the ORCID is identified to be probably “same as” ISNI number and the researcher is given the chance to confirm or deny that at that point.

Either possibility is probably acceptable to end users. It would certainly be useful for JISC to make it known that interoperability between ORCID and ISNI is an important factor in any final decision on adoption of a researcher identifier standard; but it is not our opinion that complete merger of ISNI and ORCID is a prerequisite.

**UUIDs (universally unique identifiers)** are mainly used in software creation but are suggested by some as the best candidate for researcher identifiers. **An advantage is that:**

> anyone can create a UUID and use it to identify something with reasonable confidence that the same identifier will never be unintentionally created by anyone to identify something else. Information labeled with UUIDs can therefore be later combined into a single database without needing to resolve identifier (ID) conflicts\[17].

---

This means that it is perfectly reasonable to generate UUIDs for things completely independently of any central organising body, and so makes them very cheap and long-lasting.\textsuperscript{18}

But the drawback is in getting them used

\textit{People do not like them however – subjectively – they do not like them as part of visible URLs, they do not like them as identifiers to wield, and they do not like identifiers for themselves that they cannot remember by rote.}\textsuperscript{19}

Keith Jeffreys of STFC, a proponent of UUIDs, says that this misunderstands the way they would be applied – they would be invisible to researchers and used by software clients on their desktops or in institutional systems as part of the daily workflow, after the researcher has authenticated in the normal way.

It is hard for us to see how such a standard would be widely used without the “killer apps” which would get users over the dislike and mistrust mentioned above. However, it is undoubtedly true that any ID will require a compelling demonstration of use case and interface before it catches on and gathers enough critical mass or social media coverage to persuade individual researchers to claim one. We believe that critical mass (of services as well as users) is more likely to gather around ORCID.

We spoke to several people who raised the issue of future planning and concern about a single point of failure – in a number of different contexts. This was also raised by the Task and Finish Group in its recommendation 3

\textit{There will need to be a contingency plan to build a national solution that would provide the functionality currently predicted for ORCID.}

Work going on at the University of Oxford is interesting in this regard in that the university’s digital libraries are looking at using UUIDs for a number of reasons:

- Cheap to issue
- Decentralised system means not relinquishing control
- Not reliant on a registry whose funding model may or may not be secure for the future
- No questions about whether or not a person is a researcher, or indeed a person, since UUIDs can be issued for anything

However, it is recognised that a unique researcher identifier (with the services and systems that would be built around it) would enhance efficiency, enable more and fuller citations, and aid de-duplication and disambiguation. Plans are at an early stage yet, but it seems likely that Oxford will use UUIDs to identify people but will use “same as” statements to link to ORCIDs and ISNIs where they are available. It remains to be seen if this will be more or less efficient than, say, using ORCID as the primary identifier for researchers and the services they need and generate. But, in the light of the Task and Finish Group’s reasonable desire for a contingency plan, it may be worth JISC encouraging the trial use of UUIDs in this way.

This is particularly suggested and supported by the work being done by The Consortia Advancing Standards in Research Administration Information (CASRAI)\textsuperscript{20}. JISC is leading efforts to develop a UK chapter of this international, community-driven, non-profit standards organisation\textsuperscript{21}. David Baker, CASRAI Executive Director, puts it like this

\begin{itemize}
  \item \textsuperscript{18} \url{http://benosteen.wordpress.com/2011/09/13/orcid-some-questions-and-answers/}
  \item \textsuperscript{19} ibid
  \item \textsuperscript{20} \url{http://casrai.org/about}
  \item \textsuperscript{21} \url{http://www.jisc.ac.uk/news/stories/2012/04/casrai.aspx}
\end{itemize}
6.1.2 Why a researcher identifier and not just a person ID?

It is a reasonable question. There are good reasons why some individuals and certainly some organisations would prefer to have an ID with a wider scope (why not an academic ID, for example so that a student would be able to keep the same ID through school, undergraduate studies and into a research career – but then why restrict it to academia, many will go to work in business, government or the third sector?). It seems to us that there are some compelling reasons why a researcher identifier is appropriate at this time. For most (not all) researchers, there is a reasonable chance of drawing a boundary around the scholarly work and involvements and an understanding that there is a benefit in making that information more publicly available than other personal or social information:

>a big part of my life is being a researcher, but it's not the whole of my life ... the value of a research[er] identifier is it lets me choose to express myself as a researcher, and to associate the work that I do, that I see as being part of my scholarly persona

Because the ORCID is focussed on researchers, there is the potential and the promise of a set of services and systems (e.g. Mendeley, ScholarOne and Nature have apparently been building to the ORCID API) which will be able to engage with ORCID in a way which is not yet evident with the other standards mentioned above. The potential application of ORCID across international boundaries and its focus on Research output is sufficiently different from national ID systems with no international reach such as the UK’s Unique Learner Number\textsuperscript{22} to make it sufficiently valuable to the researchers themselves.

6.2 Privacy and security concerns

6.2.1 Misunderstandings and misapprehensions

Misunderstandings and misapprehensions are significant because they need to be considered in the way recommendations are framed and future decisions are communicated.

*There is already a ResearchID for researchers, don't see why JISC needs to invent its own system for them.*

*Why ORCID is already an unique id system ResearcherID?*

Michael Taylor, from Elsevier and active in the development of ORCID, reinforces these messages noting there is a lot of confusion about the generic term “researcher ID” and the proprietary system “ResearcherID”, owned by Thomson Reuters. Although we did provide sufficient background information with the questionnaire to answer these queries, it is clear

---

\textsuperscript{22} The UK unique learner Number (ULN) tracks an individual’s academic performance and is described at [http://www.learningrecordsservice.org.uk/products/uln/](http://www.learningrecordsservice.org.uk/products/uln/)
that work will need to be done to explain or clarify the situation. The fact that Thomson Reuters made the code for ResearcherID available to the ORCID project to act as a code base (without retaining rights on future code developments) is a nuance which has not been appreciated in parts of the community.

**I think the problem for the researcher will be in acquiring an identity**

An ORCID will be freely available to an individual applying for or, more accurately, "claiming" one. The possession of an ORCID will not (or should not) confer any status. The process of accrediting and validating an individual as a researcher will rest with the appropriate research bodies, whether these be institutions, funding organisations, publishers or learned societies.

**I am unsure of the benefits to scholarly societies and statistical agencies if the IDs do not encode semantic information as suggested.**

There is general agreement among experts that to encode semantic information would be unwise and prejudice the longevity and security of the ORCID. This issue was dealt with in previous reports but a simple explanation needs to be available to those interested.

**I think ORCID might be a bit like a National Insurance number which is of more use to other agencies than the owner of the number - but in the end you learn your NI number as other people ask for it so much!**

The idea behind ORCID is that it should be researcher driven and certainly our respondents generally agree that if it is not attractive to and/or widely used by researchers then it will not gain widespread acceptance. Care should be taken that it is not perceived as just another number to learn with only bureaucratic purposes.

**Tracking individual’s affiliations/grants will only be of use if people have always had UK research involvement, what about people who leave the system during part of their career and go "off radar"? A lot of this sounds like a cumbersome central database that won't give an overview of an individual researcher’s career. Most researchers will have work activities beyond the ORCID system (e.g. Impact with outside schools etc) so creating a CV relies on each person’s own record keeping not some centralised database.**

ORCID has been suggested as the most suitable candidate for a researcher identifier precisely because it has international participation and will not be parochial in scope. Independent and unattached researchers will be able to use it and, while it will not be, nor is intended to be, a complete CV tool, a variety of input mechanisms should make it possible to bring together research activities, including those of sporadic researchers or those whose work takes them into non-research areas for however long.

**in order for a system like this to be practically workable there must be an 'opt out' for those not comfortable with giving over all their details to an agency.**

The minimum data elements required by the system are name and email address – and both those elements may be set to private (i.e. unavailable to enquirers) or “share with authorised contacts” or “public”. A researcher may choose only to give the minimum data elements to ORCID and to give other particular details to whichever research organisations are trusted.

### 6.2.2 Privacy and security - reservations, issues, objections and misgivings

Here we focus on what we consider to be genuine issues which need to be addressed and on whether they need addressing before a decision on ORCID adoption.

Even where researchers claim IDs themselves, there could be privacy concerns where those IDs are then used by organisations such as employers and grant awarding bodies. Where bulk upload and registration is taking place the concerns are justifiably greater. Perhaps procedure errors are a greater concern here than deliberate hacking or data theft. One might imagine an institution performing a bulk upload for its new postgraduate researchers, then using those numbers to link to personal information of an individual. Then, in
exchanging information with a publisher or research council, mistakenly giving personal information perhaps medical details, home address, which might sometime later be harvested by a web service matching experts with media requests for interviews and obtained by a hacker who publishes the entire list on a Russian website. Clearly several very significant things need to go very wrong in different places and processes for this imaginary scenario. First I (as the researcher) should be giving permission for any sharing of information – an example of what should be happening at the assignation centre is shown in figure 2. Secondly, I may give my institution a variety of permissions - e.g. for my publications record and repository entries to be shared with any publisher, for my home contact details not to be shared at all, for my salary grade point to be shared only with potential funders for the purpose of funding applications and for other basic information (name, dept, qualifications) to be publicly available. Thirdly this nightmare scenario requires system security breaches on behalf of both the imaginary institution and the imaginary publisher or research council and a lapse of attention in at least three places. But it could happen. So all possible steps should be taken to mitigate any such risk. In particular, any secondary data which is held centrally as part of the ORCID should be subject to researcher authorisation (initially on a field by field and subsequently on an organisation or case by case basis):

![ORCID Authorization](figure2.png)

**figure 2 - An illustration of the possible permission process proposed by ORCID**

Systems using an ID and containing secondary information will need to be registered with the Data Protection Service in the UK separately. Such systems will need to be separately evaluated and have a security policy that is commensurate with the risk of the data they protect. We allude to this in our risk table below in mitigation M18

Publicity should be given to a clear statement that the minimum data required to be made public is the number alone, while the minimum required to be held (but not necessarily shared with anyone) is just the ID, an email address and the (given-at-claim-time) name. Of course, I may claim to be called Jane Smith but other organisations holding bibliographic and reproducing and performing rights information would require quite extensive authorisation against their own system before accepting changes or amendments to their records about (possibly another) Jane Smith. So the ORCID is described is described as an Allow then Deny Later system

One respondent posed the question to us “How is this different from a compulsory ID card?” The key difference is the process of authorisation and validation. So before any UK-wide adoption it needs to be made clear to the community that:

---

23 [http://tinyurl.com/6rh376s](http://tinyurl.com/6rh376s)
a) The researcher can claim, withdraw, comment and dispute all information held at the registration authorities.

b) Where the ID is used by an organisation and further data added for organisational purposes then the entire record should be available to the researcher to validate, unless a clear case is made for confidentiality (referee’s comments perhaps). Such rights in the UK are already underpinned by the data protection act.

c) Sharing with other bodies, or publicly, should be authorised by the researcher and should include the capability of giving authorisation, on a field by field basis (see figure 2 above).

It’s easy in such discussions to overlook the fact that much personal information is already shared in a haphazard and sometimes unintentional (on the part of the individual) way by other services. Gmail, for example, will pick up and display the profile (often including other contact details, interests, memberships etc) of anyone in your contacts list. This is not to say that commercial players like Google, Facebook, Twitter and LinkedIn provide an example to follow (the recent concern about changes to terms and conditions suggest this is not the case). But it is worth pointing out that there are many people (doubtless including some researchers) who intentionally or unintentionally share more than their scholarly achievements.

The combination of the data subject regularly seeing, or being notified of, the data associated with an ID by together with a vested interest in the data being correct will, we anticipate, provide a feedback mechanism that rapidly identifies and corrects most data associated with living, active researchers.

6.3 Scope

Most respondents agree that the scope (i.e. “who is a researcher”) should be wide. Initially at least a balance will be necessary but certainly unattached researchers, postgrads, technicians will and should all be included. As we understand it, anyone will be able to claim an ID – the accreditation mechanisms will happen at other places (e.g. research councils will use their internal systems to decide whether Jane Smith [http://orcid.org/0137-1963-7688-2319] is an accredited researcher according to them). Many respondents expressed concern that ORCIDs should be available to researchers outside institutional or conventional settings:

*those outside formal institutions, if they wish to be included - think of Peter Mitchell FRS ([http://en.wikipedia.org/wiki/Peter_D._Mitchell](http://en.wikipedia.org/wiki/Peter_D._Mitchell)) who spent much of his career as an independent privately funded researcher working from a house in Bodmin.*

*What if people don't want to give you their details - will this stop them from working in research?*

*As a one-person research and consultancy company - but still active in scholarly publication and with over 250 lifetime units of output, it is important that any new system is capable of including people like me with a low entry cost.*

*Given the fragility of the job market, and the way in which researchers move in and out of it (especially women), I think it's crucial that the system doesn't impose institutional value judgements on the researchers. It should therefore be able to cover all participants.*

*I would be concerned about what criteria are used to assess whether someone is entitled to a number. Would it be based on publication? If so, what sort of publication? or would it be based on membership of a particular institution or learned society? Would it be extended to 'service user/peer researchers'? Ultimately, I fear there is a risk that researchers working outside academia might be unfairly excluded.*

---

24 If you have a gmail account go to Contacts, More, Print, All contacts – it may surprise you.
Note that retired academics mostly continue their work, but are rarely supported by grants or by universities (beyond minimal provision of internet access and library facilities). We tend to get left off the records.

Be careful not to tightly couple a ‘Researcher’ with an official employment position. There are amateurist scientists who are not attached to an institution but who nevertheless contribute regularly to the creation of knowledge.

recognise the wealth of researchers beyond academia

It was nearly unanimously recognised that a national identifier would be a very partial solution and that an international scope would certainly be preferable and probably necessary if a unique identifier is to be recommended or used.

6.4 Contingency plan

The Task and Finish Group’s recommendation 3 included:

There will need to be a contingency plan to build a national solution that would provide the functionality currently predicted for ORCID. This will need to be in place and fundable, to ensure that it can be rolled out as quickly as possible should ORCID fail or diverge from UK needs.

Respondents to the questionnaire were asked specifically about this and a variety of responses were made:

similar to (for example) DOIs, even if the central resolver service goes away for some reason in the future, researcher identifiers already in circulation by that time in the UK will still have value on their own for cross-linking and integration purposes.

Ensure ORCID does not fail - this is safer than starting some national substitute that is bound to fail.

Current ORCID statement is that it is open source and will be run and maintained by a disinterested body; this should remain the case. And if ORCID or any other similar system should be adopted, it must be ensured that it is open source and also unconstrained in other ways for easy replication should host body fail (e.g. commitment to release associated domain names for resolvers, etc)

This isn’t just a UK issue so not clear a UK contingency would be particularly useful

the UK can go a long way to ensuring that ORCID does not fail - and our EU partners should be involved in this too

The contingency plan that we consider most likely to be effective is articulated by the work being done by The Consortia Advancing Standards in Research Administration Information (CASRAI). We mention at the end of section 6.1.1 the use of UUIDs and the suggestion made by David Baker of CASRAI. We recommend pursuing this as the best future strategy for a contingency plan and encouraging and closely watching existing UK efforts to use UUIDs as an infrastructure “underneath” a variety of identifiers including researcher identifiers.

6.5 Non-commercial / Open source

There were several expressions of concern amongst questionnaire respondents and during interviews that ORCID should be “non-commercial” or “open source”. Exactly what is meant by this is not always clear from a comments field in a questionnaire. The ORCID organisation itself is registered as non-profit and this is important (of course ORCID will be used by commercial organisations and for approved commercial applications – and this is not only permitted but highly desirable, but there will need to be an approval process or restrictions placed on this – clearly mass email of researchers from providers of pension provision, for example, would not be acceptable). It is guaranteed that an individual will be
able to claim an ORCID without charge. It is not completely clear, but we assume it is the case, that individuals and small organisations will also be able to access a free look-up service. We understand that the ORCID code will be open source although there have been some complaints about delays in the release of the code. We also understand that there are no restrictions placed by Thomson Reuters (who owned the original code) on the future distribution of the code. Indeed we were advised that ORCID has invested considerable resources in cleaning the original Thomson Reuters code to ensure that it is now completely free of any connection to Thomson Reuters.

6.6 Researcher ownership

Several respondents comment on how researcher participation / enthusiasm / ownership is a key feature. Ordinary researchers are unlikely to find themselves on the governing or oversight organisation of such an organisation but they need to feel that they are represented on the board and indeed that researchers’ interests are steering the governing or oversight organisation. Claim, assert, deny, amend are all key features to emphasise in publicity – if researchers feel it is a top down initiative led by institutions, funders, publishers, government agencies or a combination of all four, then they are unlikely to participate in numbers.

6.7 Seeding versus claiming

On the other hand, it will be important to create a mass of ORCIDs and this is likely to happen by using existing data sources including the upload/download by institutions. In this context it is important that existing data sources such as the UK Names project are used. Names have done much work on disambiguation and cleaning data from a wide variety of UK academic sources. At least in the short term, it will be important for Names to receive sufficient funding to ensure that this work is not lost, that the names data can be used to seed the ORCID database and that the expertise and experience of the Names project can be applied to ensuring that ORCID has an initial tranche of substantial, secure and accurate data and inherits advice on tried and tested procedures for dealing with problems in assigning and maintaining IDs for the UK academic community.

6.8 Administration in the UK and the role of HESA

We asked respondents in both interviews and questionnaires about the role of HESA in the potential UK adoption of ORCID. There was a wide variety of responses which probably reflected poor awareness of HESA’s current role and a variety of prejudices about the importance or desirability of gathering statistical data. Nevertheless we were able to reach some conclusions:

- If ORCIDs are to be widely used in the UK academic community then it would make sense, indeed be essential, for institutions to include ORCIDs in their HESA returns and in the Research Excellence Framework (REF) following the one which will be completed in 2014. HESA themselves should request this and HESA’s statutory customers should require it.
- It would make sense for HESA to be using the same format of number for staff and student IDs. This does not necessarily fit with the ORCID intention to cover only researchers. HESA must decide on its response to this – if the strategy outlined above in “Contingency plan” is implemented it could be that UUIDs underlie ORCIDs, Staff IDs and Student IDs.
- Because of its current role in requesting from institutions data sorted by a person ID, HESA should have a role to play in advising and coordinating the response of the UK HE community to a decision to encourage the widespread adoption of ORCIDs. But it needs

---

25 ORCID principle #8 states that all software developed by ORCID will be publicly released under an Open Source Software license approved by the Open Source Initiative. For the software it adopts, ORCID will prefer Open Source. [http://goo.gl/IVMmS](http://goo.gl/IVMmS)
to be borne in mind that ORCIDs will have a much wider applications than HE. UK researchers work in many situations outwith the purview of HESA: research institutes, commerce and business, government (local and national), other research organisation such as Wellcome, Rowntree etc, charities and trusts and on their own. HESA should not be expected to represent or advise such a diverse community.

- It would not be appropriate for HESA to represent researchers or the research community in the governance of ORCID. But we would expect ORCID to welcome a HESA representative to advise on UK procedural and statistical requirements and practices.

7 ORCID Risk Assessment and Mitigation

Our primary brief was to validate the recommendations of the Task and Finish group. We make some recommendations ourselves in the Executive Summary (cover page) but we conclude by presenting a table of risks and mitigations to assist the Task and Finish group, and the HE and academic research community in general to push forward with this initiative.

Our assumption in generating the risk and mitigation framework below is that the core service will have been adequately designed and have appropriate security mechanisms in place, been constructed with appropriate levels of end-user functionality in mind and have an information architecture and information space capable of fulfilling the identified core objectives.

With these assumptions in mind, the table below provides a risk assessment that looks at the wider cultural and technical adoption risks which would prevent such a system being adopted. It provides a series of mitigating actions as a means of reducing these risks. The table also provides a risk description, a risk score based on the probability of a risk occurring and the impact should the risk transpire.

There are 21 mitigating actions for the Task and Finish group to consider.

Risk scores have been estimated before mitigation and have been assessed on the basis of their probability (P) on a score of 1-5 and their impact (I) should the risk occur again on a scale of 1-5. The Risk score (S) is the product of Probability x Impact giving an overall risk scale of 1-25.
\[ P = \text{probability} \quad \text{– the likelihood that the risk will occur on a scale of 1-5 were 1 is low probability} \]

\[ I = \text{Impact} \quad \text{– if the risk occurs then what would the impact be on a scale of 1-5 where 1 is low impact} \]

\[ S = \text{the product of } P \times I \text{ giving a joint score ranging from 1 – 25 where 1 is a low risk and 25 is high risk} \]

<table>
<thead>
<tr>
<th>Risk</th>
<th>P</th>
<th>I</th>
<th>S</th>
<th>Mitigation and comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low uptake or adoption by end users:</strong>&lt;br&gt; If support from end users of the service is low as a result of lack of awareness or apathy or a groundswell of opinion that ORCID does not represent the best interests of researchers or research bodies, then the service is unlikely to gain the critical mass necessary to achieve mass adoption.</td>
<td>3</td>
<td>5</td>
<td>15</td>
<td><strong>M1</strong> Establish a strong governance and oversight group with representation from researchers who are held in high esteem. Such a group will need to ensure authentic participation from all stakeholders. <strong>M2</strong> The programme for implementing the service should have a significant communications strand that focuses on benefits to end users and addresses some of the misconceptions identified in the study. <strong>M3</strong> Establish and sustain advocacy for the idea and the implementation using social media and other communication channels. See also <strong>M21</strong> below</td>
</tr>
<tr>
<td><strong>Poor disambiguation with incorrect attribution of material:</strong>&lt;br&gt; A relatively small amount of material by a few key authors is poorly or wrongly attributed; this has the potential to undermine confidence in the service. Allocation of researcher identifiers to “back catalogue” material might be a typical source of such errors.</td>
<td>3</td>
<td>3</td>
<td>9</td>
<td><strong>M4</strong> Daily or weekly email digests of any changes made to a researcher’s data, with the ability for researchers to flag correct/incorrect records is highly recommended. We suggest digests rather than individual change records as this will avoid email overload and dissatisfaction with the service. <strong>M5</strong> A data quality function, which includes a low key dispute and conflict resolution service, would assist in addressing errors in a timely manner, something that is important if the error is not to propagate across the information space. Interoperation with ISNI would be an important mitigating factor in the “back catalogue” issue.</td>
</tr>
<tr>
<td><strong>Multiple researchers using the same ID:</strong></td>
<td>2 4 8</td>
<td><strong>Mitigation M4</strong> described above if used will help identify and resolve such issues at an early stage. <strong>M6</strong> Establishing a public facing “errata” service to provide an audit trail of resolved errors and visibility of changed records will reduce service desk calls and act as an authoritative reference.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>------</td>
<td>--------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
|  The risk of multiple researchers having the same ID is reduced when the researchers are active however it is entirely possible that two current researchers could be allocated (or, more likely, mistakenly recorded as having) the same ID either as an administrative or programmatic error within local systems used within an institution.  

We can expect such errors to occur in greater number should the scheme be extended to cover deceased or inactive researchers.  

Researcher identifiers will be used in multiple systems including hard copy, as a consequence retrospective disambiguation between existing researcher identifiers will be necessary. | | |
| **Individual researchers with multiple IDs:** | 3 4 12 | Mitigation M4 and M6 are equally applicable to reducing this risk.  

We would expect that most individual researchers would wish to consolidate their publication record. It is likely that for a variety of good reasons it may on occasion be beneficial for a researcher to have multiple IDs. Data quality reviews should consider such a likelihood.  

While problems will certainly occur, identifying and addressing such issues in a timely manner combined with cross referenced checking mechanisms will reduce its impact. |
<table>
<thead>
<tr>
<th><strong>Inability to populate a back catalogue:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>We can expect active researchers, research bodies and academic publishers to have a vested interest in establishing a comprehensive record of their publications. There is a potential conflict should a researcher not wish to have an ID while the research body or publisher wishes for its material to be recognised.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Technical failure(s):</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>All services are subject to technical failure. Distributed systems with application programmatic interfaces (APIs) are particularly susceptible to failures associated with change management.</td>
</tr>
<tr>
<td>It is the nature of end users that they do not distinguish between a technical failure in the core system and those provided by third party systems that consume such services.</td>
</tr>
<tr>
<td>Modern software services are expected to have 99.99% availability, failure to achieve this level of availability reduces end user confidence in the system.</td>
</tr>
</tbody>
</table>

| **M7** if the system is to be embedded systematically then research funding bodies should be encouraged to require researcher identifiers to be used. Note that the minimum level of information associated with an ID is very low and only the number (not name or other properties) must be publicly visible |
| **M8** Preserving the right of a living individual not to have a researcher identifier allocated to them would feel to be a pragmatic solution to addressing adoption issues. End user adoption has to be a critical success factor (CSF). Cultural issues that prejudice end user acceptance should be recognised and mitigated against as a priority by the governance body |
| **M9** Having a strong communication strategy that advertises system and service change well in advance (a minimum of 12 months is recommended) to allow third parties to modify dependant systems |
| **M10** Well specified APIs that are as far as possible backwards compatible should be considered |
| **M11** Public facing instrumentation that displays the status of the system and service is a way to allow end users to self-diagnose issues and, when combined with a robust set of test services and data that are benchmarked, allow developers and end users to construct robust 3rd party interfaces |
| **M12** Strong service management that models itself on ITIL processes with effective configuration and change control management combined with systems instrumentation |
**Failure(s) in governance:**
Examples of governance failure might include: Significant outages, data loss or corruption, weak untested APIs, widening the scope and range of applications that a researcher identifier applied to, weak change or configuration control etc. Any one of these could result in significant reputational damage to the service.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M13 Robust governance that includes representation from stakeholders is highly desirable. Such a group may need to have suppliers as members or a sub group if third party applications that depend on a researcher identifier are likely or intended.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M14 Establishment of an end user panel to review system and service changes is desirable.</td>
</tr>
</tbody>
</table>

**Managing data coherence and change in third party applications:**
Incorrect researcher identifiers generated, will be propagated across third party “satellite” systems some of which are likely to have been constructed on the basis that the researcher identifier is unique for any given individual and will not be subject to change.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M15 The communication strand that is associated with any such rollout should make it clear that the researcher identifier should only be used for the approved purposes. In addition “satellite” systems should never use the researcher identifier as a Primary Key and should always have their own internal unique index.</td>
</tr>
</tbody>
</table>

**IDs brought into disrepute as a result of inappropriate use:**
Interviewees commented that, if it were reliable, an ID could have multiple applications in any institution. Examples included booking resources, recording and tracking career history of an institution’s graduates and researchers, identifying previous applications for funding etc.

There is a risk that if such applications expand significantly beyond the original intended scope the architectural security model that underpins the service design will not be sufficiently robust to prevent identity and resource theft.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mitigation M15 and M13 should partially cover this risk. A unique index will be seen by many to have multiple applications if adoption levels are high. Governance mechanisms should be established to ensure that its formal use is restricted to the approved applications.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In addition corporate bodies and individuals should be made aware of the possible risks should personal or confidential information be associated with an ID.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Some areas of research are particularly sensitive (researchers involved with animal testing are an obvious example). It is outside the scope of this report to assess or identify the risk associated with such an edge case but appropriate mitigation needs to be in place to prevent personal data held at local level and associated with an ID from leaking.</td>
</tr>
</tbody>
</table>
Incomplete publication record:
We can anticipate that there will be significant gaps and omissions of publications associated with an individual’s researcher identifier. If the system attains a critical mass then we can expect the majority of individual researchers to ensure that the record is well populated if not comprehensive.

Until that point is reached there is a risk that any decision support systems that use attributed researcher identifier associated publications as their main source of validating an academic record will be criticised.

Deliberate subversion of the system:
All systems are subject to some degree of deliberate subversion, we can anticipate that this will take the form of creating fake IDs, falsely claiming IPR on publications, attributing publications incorrectly to a researcher identifier etc.

Critical and systemic failure of system for technical, cultural or financial reasons:
It is possible that a system fails for any of the reasons described above. Once established, such systems typically underpin a significant number of administrative processes and become a part of the administrative fabric of a sector or organisation.

Difficult or time consuming to use:
In the interviews and in our study, an observation commonly articulated was that the system should be both simple and quick to use. Failing to keep the system intuitive to use, responsive and once established, relatively quick to modify was identified as a CSF. Failure to meet these functional and non-functional requirements is likely to result in the service being abandoned by its end user population.

M16 Once the service has been established specifying that academic publications will be cross checked against the researcher identifier system could be expected to positively promote adoption of such a system.

M17 Appropriate controls and delegated responsibility for allocating researcher identifiers is strongly recommended.

M18 While we can anticipate that the core systems will have appropriate security mechanisms, satellite systems that contain the researcher identifier and associated personal data will need to have appropriate validation and verification mechanisms at a level commensurate with the resources and data they protect.

M19 It was suggested in the interviews that applying a shadow open-source ID such as that provided by UUID could provide an appropriate level of insurance.

M20 Our assumption is that the application itself will have been designed with this risk in mind. The risk therefore is associated with the non-functional infrastructure requirement to provide an architecture that can scale and is designed to provide adequate performance and capacity under maximum load.

M21 Establishing a set of compelling and persuasive use cases and/or demonstrations.
Mitigation M11 (instrumentation) has a significant role to play in ensuring that the service remains within any service level agreement.
Appendix 1 - JISC Researcher Identifier Initiative

Unique identifiers for individuals involved in research are a central component of many other processes. An identifier solution would enable, for instance:

- Disambiguation of authors; assigning publications to a specific individual would be possible automatically on both a national and global scale.
- Career tracking of researchers, from post-graduate research to emeritus professor, across institutions.
- A reduced burden on researchers by enabling easier entry on forms for funding and appraisal.
- Clarification of staff institutional affiliations; which can, for example, be used to confirm eligibility for access to international facilities.

In light of this, JISC convened a ‘task and finish’ group, consisting of representatives from key organisations and bodies from the UK HE sector, alongside experts and researchers, to draft a set of recommendations for the UK in how to implement an identifier solution that will meet the needs of the sector now and in the future. Organisations represented on the group included:

- The Association of Research Managers and Administrators (ARMA)
- The Higher Education Funding Council for England (HEFCE)
- The Higher Education Statistics Agency (HESA)
- The Medical Research Council (MRC)
- The Natural Environment Research Council (NERC)
- The Universities and Colleges Information Systems Association (UCISA)
- The Wellcome Trust

Researcher viewpoints were offered by Dr Simon Coles (Head of the National Crystallography Service, University of Southampton) and Dr Cameron Neylon (Senior Scientist, Science and Technology Facilities Council). The group was supported by staff from the Innovation Support Centre at UKOLN, University of Bath and JISC.

The group met in May 2011, and discussed the information they would need to arrive at a set of recommendations. This led to the creation of a series of eight reports which were used to inform and guide discussion of the challenges, issues and opportunities facing the sector in developing a unique identifier solution for research activity (see below for reports).

After six months of information gathering and exploration, the group met on 25th November 2011 and set out a proposal for ten recommendations to the sector and a consultation process for ensuring that the recommendations represent a broader consensus. This consultation will be undertaken between now and the end of May 2012 and will be used to direct the sector wide adoption of an identifier solution.

Reports:


Appendix 2 - VIAF and other IDs

[This is reproduced from a blog entry by Thom Hickey (OCLC's chief scientist). The authoritative version can be read together with any comments, updates and clarifications at: http://outgoing.typepad.com/outgoing/2011/07/viaf-and-other-ids.html]

One of the things VIAF does is to provide a uniform identifier for entities named in library authority files. Currently, VIAF covers personal, corporate and jurisdictional names, but other names such as those for works, expressions, families, trademarks, and non-jurisdictional geographic names are all within the scope of VIAF.

Of course named entities in libraries overlap with those used other places. This is especially obvious for personal names, where there are many ways to identify people. In fact, OCLC is involved in at least two new ways of identifying people, the International Standard Name Identifier (ISNI) and the Open Researcher & Contributor ID (ORCID) and how VIAF relates to these is a common question. As VIAF has been around for several years and ISNI and ORCID have yet to go public, the relationship is evolving, but I'll try to describe at least how I see it.

But first, why should there be multiple identifier schemes? Couldn't we all just agree on a single one and have everyone use it? Well, no we couldn't. Even within a single organization that can be difficult, and there are many not just organizations, but many communities interested in identifying entities such as people. Using people for the example, here is a quick list of reasons why we can't just all share one identifier scheme:

- Historical reasons: So many groups have their own lists of people that most new naming schemes are made by merging existing lists. Existing policies are already in place that have to be respected but sometimes conflict across communities.
- Different ideas of what we are identifying: Are pseudonyms allowed? How about collective personal authors? Dead people? Deities? Imaginary characters? Are we describing public names or actual people?
- Different information: Information is needed to differentiate the names. Do we have dates (e.g. library authority files) or institutional affiliations (e.g. article authors) or other information? Merging files without consistent common information is fraught with difficulty.
- Who creates the names: Does the system need real-time updating? Can anyone add a name at any time? What about editing, splitting and merging names?
- Different range of entities: Creators or subjects? Current or historical? ORCID and ISNI are much more interested in live authors than historical figures.
- Different priorities and control: Who's in charge? Every system will have more requirements than can be implemented any time soon. Setting priorities for what
comes next can be critical in gaining acceptance of a scheme. Even cohesive
groups have difficulty setting priorities; across more disparate groups it becomes
impossible.

Experience has shown that no single scheme can accommodate the myriad array of
requirements that different communities have. VIAF tries to satisfy the global library
community. While that is a challenge, at least the participants come with similar
expectations and data. Who is supporting a scheme and what they hope to accomplish are
important clues as to what a scheme will provide.

ISNI is part of the ISO's international standard identifier program and is run by the ISNI
International Agency. It's first supporters are mainly rights organizations with a strong
element of library participation through the Conference of European National Librarians. In
fact, the entities it plans to identify (especially entities with current rights, such as authors,
performers and publishers) overlap quite highly with library authority files. OCLC is
currently implementing ISNI and will be administering the system. Since OCLC has also
implemented VIAF, it is not too surprising that VIAF is working closely with ISNI, although
the participation of the French National Library and the British Library in ISNI was just as
important in making this happen. ISNI will be using a copy of VIAF, along with other files,
as its base file. As ISNIs are assigned to entities that have VIAF IDs, that information will
be fed back to VIAF, so there will be links between the two files. Also, as ISNI processing
discovers errors in the VIAF file (such as two names that should be merged) ISNI will inform
VIAF.

ORCID grew out of Thompson-Reuters' ResearcherID system. It has the support of a
number of major publishers, scholarly organizations and universities and is most interested
in current authors, especially of scholarly articles and books. A not-for-profit organization
(ORCID, Inc.) has been formed to run it. Since many authors of articles do not produce
material that libraries control, the overlap with library authority files is limited. Also, partly
due to its roots in the ResearcherID system, it is very interested in author-created and
claimed IDs, a concept not yet widely accepted by libraries. Although VIAF is interested in
ORCID (e.g. I am OCLC's representative on the ORCID Board), it is not yet obvious how the
two systems will interact.

In my mind, ORCID and ISNI have quite a bit in common. They are both new organizations
organizing large and overlapping parts of the identity problem for published information.
The two organizations talk, but so far no easy way of working together has been found.
Since both are in the midst of implementation of their initial systems (ISNI expected in the
fourth quarter of 2011, ORCID some time later), closer cooperation right now is difficult, but
I expect that longer term, ways will be found for the two systems to work together. All
three systems will get mutual benefits from linking, so I predict that eventually they will be
linked.

That VIAF and ISNI will be closely related is good news for libraries, since it will make our
identifiers much easier to link to ISNIs, which can be expected to be encountered in many
places in the future. Providing bridges between systems is a major role for VIAF. While
VIAF is not only about libraries (the Getty Union List of Artist Names is included), ISNI will
be our first formal relationship outside the cultural-heritage sphere. We hope it will not be
the last.
Appendix 3 - questions and answers on ORCID

Two sets of questions from Nicky Ferguson and answers, from Ben O'Steen and Martin Fenner, on ORCID.

Comments can be made on Ben's blog at UKOLN:  
http://technicalfoundations.ukoln.info/blog/orcid-some-questions-and-answers

and at the ORCID FAQ page:  http://www.orcid.org/faq#t41n521

1. ISNI, ORCID, VIAF etc ... will they each or should they be a subset of UUID, in a world where there is a need for identifiers for all sorts of things from lab notebooks to datasets to institutions, as well as researchers?

ORCID and VIAF have both plumped for a ‘short’ number and a verbal prefix (e.g. VIAF ID: 747462). It is intended (eventually) that the profile corresponding to a given ORCID should be able to be found from an ORCID site, and not necessarily the ORCID site.

You can currently construct URLs for both where that ID number is used as a suffix to do a lookup on that researcher/author/etc, with effort and consideration being made so that the URL prefix will not change in the near future. It is naive to think that any URL prefix that will never, ever change but keeping the URL usable for as long as humanly possible is given serious thought.

With UUIDs, you will have to do something identical as there is no DNS lookup *system* for them but a handful of individual sites that record links as it suits them. Due to the UUID range being so large, the key advantage of the scheme is that given a suitably random manner to generate them, collisions between UUIDs made on separate systems are incredibly rare. I’m not sure that anyone has recorded a collision yet, (disregarding those due to poorly configured entropy pools on virtual machines) This means that it is perfectly reasonable to generate UUIDs for things completely independently of any central organising body, and so makes them very cheap and long-lasting.

People do not like them however – subjectively – they do not like them as part of visible URLs, they do not like them as identifiers to wield, and they do not like identifiers for themselves that they cannot remember by rote.

2. Who decides who is a researcher? In the UK some universities call all their members of staff “teacher/researchers”, others make a clear distinction. What about schoolchildren who jointly author a paper? What about researchers in charities or industry who may never author a paper. What about peer-reviewers and research “users”?

ORCID currently is an “Allow then Deny Later” system. The main ‘ORCID’ site will be a self-signup website (with an initially limited ability for proxies to sign up and create and amend profiles for others) and the ‘researcher-iness’ of profiles will not be policed as there is no need to, unless the profile claims something untruthful.

The core of the system is based on trust – if a person claims an institutional affiliation, that will be marked as untrusted until that institution verifies this. If an institution or research group doesn’t verify the data, care is being taken that this is displayed as clearly as possible.

There is no need to police people, only to police the claims they make about themselves and the works they claim to have a hand in publishing.

3. Even institutions which pride themselves on their research may only have 20-30% of their staff who are researchers, how do you sell a business case to them that they should alter their systems to accommodate an identifier for only a minority of the staff on their finance/HR/security systems?

Again, the ORCID system (and to an extent the VIAF system) is geared to help the researcher – at a basic level, keeping a note of the ID which a researcher has is all that is required to begin to benefit from it. I think that due to the well understood pace at which change occurs within the administrative systems of an institution, the first meeting at which a business case for change might need to be presented will occur many, many months after the researchers have adopted the system for themselves as just part of the academic toolset. And if the researchers do not find it useful, then it will disappear like so many of the previous ID systems.
4. Similar question about researchers themselves – they have been disappointingly reluctant to deposit their papers in repositories and to use grant numbers in their publications, even when “mandated” – who will design the compelling interfaces which will encourage them to use ORCID … in the academic community we don’t have a great track record at designing compelling interfaces?

It is not an academic community that is designing the interface for one – it has already been outsourced to a small team of local designers and developers that Crossref have had good working relationships with so there is hope there. The key will be whether or not the system will save time for the researcher and make certain tasks that they already do easier.

The API for the ORCID service is very much the focus at the moment and certain use-cases have been thought through, such as encouraging publishers and journal submission processes to use the ID system, rather than get the researcher (or PA/postgrad by proxy) to fill in all their information again, as well as bootstrapping the ORCID database with information already within existing bibliographic databases so that many profiles need only be claimed and verified, rather than generated anew.

I do not mean to knock the institutional repository scene unduly (having been an institutional repo person myself) but I have yet to see more than a few repositories strive to make the researcher’s lives easier and better. It is worth noting that those repositories are the one’s that are thriving.

5. What role would a national registry need to play to map ORCID (or a.n. other identifier) with key information?

In short, include something semantically similar to ‘rdf:seeAlso’ within the database/triplestore/profile for the national registry’s version of the same person. Many of the codebase changes occurring at this time are so that the informational claims within other whitelisted registries can be automatically shown and interpreted within the ORCID store, moving towards a multi-trust system.

6. I understand that the idea is that the researchers themselves would control the registration and updating processes – but institutions, funders and government agencies will surely want to maintain their own registries/database using the ID … yes? Is the mechanism for change control of personal information thought out?

As mentioned above, the changes occurring and being implemented are to effect a solid multi-trust control system, which will allow for the kind of distributed profiles you mention to be accepted. However, the systems have to provide data such that a machine can use it, and that may be the sticking point for a few of these systems.

7. ORCID applies only to staff with peer-reviewed publications, and cannot be used for PhD students.

Every person who is doing scholarly work can obtain an ORCID identifier, and this of course also includes PhD students. You don't have to have peer-reviewed publications to obtain an ORCID identifier.

8. ORCID is only for authors, not other researchers

No, although ORCID is referred to as an author identification mechanism in various places on the web, in fact anyone who considers themselves a researcher whether they are a school student or advertising executive or an official in government will be able to get themselves an ID - for free (see 1)

9. Researchers were keen on ORCID until they heard they would have to pay to register.

Individual researchers will never pay to register or use an ORCID identifier and this is clearly stated in the ORCID principles: 5. Researchers will be able to create, edit, and maintain an ORCID ID and profile free of charge.

10. So how is ORCID an open standard then if organisations have to pay to use it? Is it open to other developers to build on it?

ORCID is open because the organization, the data and the software are open: a) ORCID is open to any organization with an interest in scholarly communication, b) all profile data contributed to ORCID by
researchers or claimed by them will be made openly available under a CC Zero waiver, c) all software
developed by ORCID will be released under an Open Source license.

11. *They way the UK organisations are looking at it is - if we have to change our systems
to accommodate a new person-ID then we should have one that can apply to all academic staff (UK and international) whether or not they are researchers ... we have to report on and manage collaborations between all our staff and many institutions do not differentiate anyway between teaching and research staff. So can we ask teachers, lab assistants and undergraduates all to get ORCIDs - i.e. would ORCID look favourably on being used far more widely than just by researchers? Could the system stand everyone involved in UK academia having an ORCID?*

ORCID identifiers are intended for people doing scholarly work. Although this definition will be applied very broadly (e.g. also including teaching) and there will be no mechanism by ORCID to verify this, ORCID identifiers are not intended to be used far more widely, e.g. as an identifier for every person working at an academic institution.

12. *Despite all protestations to being a not-for-profit neutral organisation ORCID is in fact controlled by Thomson-Reuters and the ORCID ID is really the TR ResearcherID. It will not be accepted either by researchers or other publishers as it is and will remain proprietary. Thomson Reuters sits on the ORCID Board of Directors and is involved in the Business and Technical Working Groups. Decisions by the Board are made by majority vote and the majority of ORCID Board members are from academic and non-profit organizations. The ResearcherID software was licensed from Thomson Reuters to ORCID, and the ORCID Board decides how to use and further develop the software, including releasing it as Open Source software.*
Appendix 4 – detailed data from the questionnaire

<table>
<thead>
<tr>
<th>RESPONDENT CHARACTERISTICS</th>
<th>10. If you work for an organisation, which of the following categories best describe your role?</th>
<th>11. Responding as an individual or on behalf of an organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q5. In which country are you mainly based?</td>
<td>Organisation that undertakes or hosts research 121 85.2%</td>
<td>Individual 135 95.1%</td>
</tr>
<tr>
<td>UK</td>
<td>Research funder 8 5.6%</td>
<td>Organisation 7 4.9%</td>
</tr>
<tr>
<td>USA</td>
<td>Scholarly society 5 3.5%</td>
<td></td>
</tr>
<tr>
<td>Iceland</td>
<td>Statistical agency 1 0.7%</td>
<td></td>
</tr>
<tr>
<td>Argentina</td>
<td>Publisher 2 1.4%</td>
<td></td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>Other 6 4.2%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q6. What is your gender?</th>
<th>12. Which of the following best describes your role? [Please select one or, at the very most, three.]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Researcher 100 70.4%</td>
</tr>
<tr>
<td>Female</td>
<td>Manager of research-related services 18 12.7%</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>Judge of applications for research funding 5 3.5%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q7. The age profile of respondents</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>25 or younger</td>
<td>Commission of research 1 0.7%</td>
</tr>
<tr>
<td>26-35</td>
<td>Manager of data about researchers and research 20 14.1%</td>
</tr>
<tr>
<td>36-45</td>
<td>Responsible for running the software systems that would need to 12 8.5%</td>
</tr>
<tr>
<td>46-55</td>
<td>Responsible for making the software systems that would need to 7 4.9%</td>
</tr>
<tr>
<td>56-65</td>
<td>Employed in scholarly publishing 5 3.5%</td>
</tr>
<tr>
<td>over 65</td>
<td>Other 12 8.5%</td>
</tr>
<tr>
<td>Not applicable</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>8. The size profile of respondents’ organisations</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;1000 staff</td>
<td>Research institute (charity), Research council 10 7.0%</td>
</tr>
<tr>
<td>100 to 999 staff</td>
<td>Arts, Humanities, Languages and Music 11 7.7%</td>
</tr>
<tr>
<td>10 to 99 staff</td>
<td>Biotechnology and Biological Sciences 19 13.4%</td>
</tr>
<tr>
<td>1 to 9 staff</td>
<td>Engineering and Physical Sciences 11 7.7%</td>
</tr>
<tr>
<td>Not applicable</td>
<td>Education 18 12.7%</td>
</tr>
<tr>
<td>** Research institute (charity), Research council</td>
<td>Economic and Social Sciences 31 21.8%</td>
</tr>
<tr>
<td>** Research institute (charity), Research council</td>
<td>Medical 8 5.6%</td>
</tr>
<tr>
<td>** Research institute (charity), Research council</td>
<td>Mathematics, Computer Science and Informatics 16 11.3%</td>
</tr>
<tr>
<td>** Research institute (charity), Research council</td>
<td>Management and Finance 2 1.4%</td>
</tr>
<tr>
<td>** Research institute (charity), Research council</td>
<td>Environmental and Earth Science 7 4.9%</td>
</tr>
<tr>
<td>** Research institute (charity), Research council</td>
<td>N/A 19 13.4%</td>
</tr>
</tbody>
</table>

13. If you have a disciplinary focus, in which one of the following broad categories is it mainly concentrated?

- Arts, Humanities, Languages and Music 11 7.7%
- Biotechnology and Biological Sciences 19 13.4%
- Engineering and Physical Sciences 11 7.7%
- Economic and Social Sciences 31 21.8%
- Medical 8 5.6%
- Mathematics, Computer Science and Informatics 16 11.3%
- Management and Finance 2 1.4%
- Education 18 12.7%
- Environmental and Earth Science 7 4.9%
- N/A 19 13.4%
14. The following “actors” in the UK research community will benefit considerably from the widespread adoption of ORCID or an equivalent solution.

<table>
<thead>
<tr>
<th>Actor</th>
<th>Score</th>
<th>% D/K</th>
<th>Score</th>
<th>% D/K</th>
<th>Score</th>
<th>% D/K</th>
<th>Score</th>
<th>% D/K</th>
<th>Score</th>
<th>% D/K</th>
<th>Score</th>
<th>% D/K</th>
<th>Score</th>
<th>% D/K</th>
<th>Score</th>
<th>% D/K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Researchers</td>
<td>4.16</td>
<td>5%</td>
<td>4.22</td>
<td>5%</td>
<td>4.03</td>
<td>6%</td>
<td>4.42</td>
<td>5%</td>
<td>4.07</td>
<td>4%</td>
<td>4.07</td>
<td>5%</td>
<td>4.16</td>
<td>7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Organisations that undertake or host research</td>
<td>4.30</td>
<td>5%</td>
<td>4.25</td>
<td>6%</td>
<td>4.18</td>
<td>7%</td>
<td>4.80</td>
<td>0%</td>
<td>4.10</td>
<td>7%</td>
<td>4.09</td>
<td>6%</td>
<td>4.36</td>
<td>5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Publishers</td>
<td>4.01</td>
<td>12%</td>
<td>3.92</td>
<td>16%</td>
<td>3.89</td>
<td>13%</td>
<td>4.26</td>
<td>5%</td>
<td>3.88</td>
<td>11%</td>
<td>3.79</td>
<td>15%</td>
<td>4.07</td>
<td>10%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Research funders</td>
<td>4.33</td>
<td>7%</td>
<td>4.31</td>
<td>6%</td>
<td>4.23</td>
<td>9%</td>
<td>4.63</td>
<td>5%</td>
<td>4.17</td>
<td>9%</td>
<td>4.18</td>
<td>10%</td>
<td>4.34</td>
<td>5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Scholarly societies</td>
<td>3.79</td>
<td>11%</td>
<td>3.83</td>
<td>14%</td>
<td>3.67</td>
<td>12%</td>
<td>4.06</td>
<td>10%</td>
<td>3.66</td>
<td>16%</td>
<td>3.72</td>
<td>15%</td>
<td>3.68</td>
<td>8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Statistical agencies</td>
<td>4.29</td>
<td>19%</td>
<td>4.30</td>
<td>25%</td>
<td>4.22</td>
<td>23%</td>
<td>4.33</td>
<td>10%</td>
<td>4.29</td>
<td>20%</td>
<td>4.04</td>
<td>23%</td>
<td>4.37</td>
<td>20%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

15. The widespread uptake of ORCID or an equivalent solution is intended to bring a number of benefits. From your own perspective please indicate the importance of the following:

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Score</th>
<th>% D/K</th>
<th>Score</th>
<th>% D/K</th>
<th>Score</th>
<th>% D/K</th>
<th>Score</th>
<th>% D/K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Improvements to the way in which research is administered</td>
<td>3.85</td>
<td>8%</td>
<td>3.86</td>
<td>8%</td>
<td>3.63</td>
<td>9%</td>
<td>4.55</td>
<td>0%</td>
</tr>
<tr>
<td>2. Simplified sign-on procedures to research-related systems</td>
<td>4.07</td>
<td>3%</td>
<td>4.13</td>
<td>5%</td>
<td>4.08</td>
<td>4%</td>
<td>3.95</td>
<td>0%</td>
</tr>
<tr>
<td>3. Reduced form-filling during the process of applying for grants</td>
<td>4.24</td>
<td>4%</td>
<td>4.28</td>
<td>5%</td>
<td>4.28</td>
<td>2%</td>
<td>4.22</td>
<td>10%</td>
</tr>
<tr>
<td>4. Better and more complete tracking of individual researchers’ careers</td>
<td>4.08</td>
<td>3%</td>
<td>4.08</td>
<td>3%</td>
<td>4.02</td>
<td>2%</td>
<td>4.26</td>
<td>5%</td>
</tr>
<tr>
<td>5. Improvements in the way research collaborations are recorded</td>
<td>3.93</td>
<td>4%</td>
<td>3.95</td>
<td>5%</td>
<td>3.83</td>
<td>4%</td>
<td>4.20</td>
<td>0%</td>
</tr>
<tr>
<td>6. Improvements in the way in which publications, grants, research projects, and researchers are mapped and linked</td>
<td>4.21</td>
<td>3%</td>
<td>4.20</td>
<td>3%</td>
<td>4.04</td>
<td>4%</td>
<td>4.60</td>
<td>0%</td>
</tr>
<tr>
<td>7. Better interoperation of local and national systems that encode and store data about researchers</td>
<td>3.85</td>
<td>6%</td>
<td>3.86</td>
<td>8%</td>
<td>3.59</td>
<td>7%</td>
<td>4.60</td>
<td>0%</td>
</tr>
<tr>
<td>8. Easier creation of authoritative lists of publications, citations and CVs</td>
<td>4.20</td>
<td>4%</td>
<td>4.25</td>
<td>5%</td>
<td>4.08</td>
<td>4%</td>
<td>4.55</td>
<td>0%</td>
</tr>
<tr>
<td>9. Smoother exchange of data about researchers between institutions during the preparation of collaborative bids, or when staff move between employers</td>
<td>3.91</td>
<td>2%</td>
<td>4.03</td>
<td>2%</td>
<td>3.80</td>
<td>2%</td>
<td>4.10</td>
<td>0%</td>
</tr>
</tbody>
</table>
16. The Higher Educational Statistical Agency (HESA) or an entity like it should be the kind of trusted body to administer and integrate the use of researcher IDs with currently used staff and postgraduate IDs in UK; and to advise institutions on data sharing issues.

17. Any solution must be capable of covering all participants in UK research, whatever their contribution or standing, including those not in traditional research roles.

18. Alongside ORCID (or equivalent), if it becomes operational, there will need to be a suitable UK wide contingency plan to ensure the continuity of ORCID’s functions if for any reason ORCID ceases or fails.

20. Prior to the ORCID (or any equivalent solution) being adopted by my organisation a business case to justify it will be needed.
### 21. Assuming that a business case is being developed please indicate the importance of each of the following issues within the case.

<table>
<thead>
<tr>
<th>Issue</th>
<th>All (Score)</th>
<th>Women (Score)</th>
<th>Ind Res (Score)</th>
<th>MDRR (Score)</th>
<th>20 &lt;36yrs (Score)</th>
<th>NonStem (Score)</th>
<th>Stem (Score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. A clear summary of the practical benefits</td>
<td>4.72 8%</td>
<td>4.88 10%</td>
<td>4.69 10%</td>
<td>4.85 0%</td>
<td>4.63 9%</td>
<td>4.67 6%</td>
<td>4.75 13%</td>
</tr>
<tr>
<td>2. Interoperability between different identity standards and different systems that encode and store research IDs</td>
<td>4.31 15%</td>
<td>4.51 19%</td>
<td>4.17 17%</td>
<td>4.65 0%</td>
<td>4.17 22%</td>
<td>4.22 19%</td>
<td>4.28 15%</td>
</tr>
<tr>
<td>3. Ongoing local costs of maintaining ORCID or its equivalent established within an organisation</td>
<td>4.38 15%</td>
<td>4.60 16%</td>
<td>4.35 18%</td>
<td>4.20 0%</td>
<td>4.49 18%</td>
<td>4.48 16%</td>
<td>4.29 20%</td>
</tr>
<tr>
<td>4. Privacy and control issues</td>
<td>4.15 23%</td>
<td>4.40 29%</td>
<td>4.09 30%</td>
<td>4.53 5%</td>
<td>4.03 27%</td>
<td>4.07 27%</td>
<td>4.22 26%</td>
</tr>
<tr>
<td>5. A clear statement from ORCID on interoperability and boundary issues with potentially overlapping standards such as ISNI and VIAF</td>
<td>4.57 9%</td>
<td>4.68 11%</td>
<td>4.57 12%</td>
<td>4.65 0%</td>
<td>4.40 11%</td>
<td>4.55 6%</td>
<td>4.50 15%</td>
</tr>
<tr>
<td>6. An assessment of the risks and benefits that might arise from implementation</td>
<td>4.31 11%</td>
<td>4.42 13%</td>
<td>4.21 15%</td>
<td>4.35 0%</td>
<td>4.38 11%</td>
<td>4.29 6%</td>
<td>4.20 20%</td>
</tr>
</tbody>
</table>

### 22. Assuming that a business case is being developed please indicate from your own perspective the extent to which you think the business case should be written with the following organisations/categories in mind.

<table>
<thead>
<tr>
<th>Organisation/Categories</th>
<th>All (Score)</th>
<th>Women (Score)</th>
<th>Ind Res (Score)</th>
<th>MDRR (Score)</th>
<th>20 &lt;36yrs (Score)</th>
<th>NonStem (Score)</th>
<th>Stem (Score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Public sector organisations, including universities who employ researchers</td>
<td>4.74 7%</td>
<td>4.82 5%</td>
<td>4.71 7%</td>
<td>4.95 0%</td>
<td>4.83 11%</td>
<td>4.67 3%</td>
<td>4.78 10%</td>
</tr>
<tr>
<td>2. Charitable sector organisations who employ researchers</td>
<td>4.22 15%</td>
<td>4.13 16%</td>
<td>4.21 11%</td>
<td>4.24 15%</td>
<td>4.22 18%</td>
<td>4.14 10%</td>
<td>4.28 13%</td>
</tr>
<tr>
<td>3. Private sector organisations who employ researchers</td>
<td>3.93 13%</td>
<td>3.87 14%</td>
<td>3.87 11%</td>
<td>4.00 5%</td>
<td>3.87 13%</td>
<td>3.82 10%</td>
<td>3.94 13%</td>
</tr>
<tr>
<td>4. Publishers of research</td>
<td>3.99 12%</td>
<td>4.07 14%</td>
<td>3.84 11%</td>
<td>4.42 5%</td>
<td>3.92 13%</td>
<td>3.93 11%</td>
<td>3.96 15%</td>
</tr>
<tr>
<td>5. Research funders</td>
<td>4.33 8%</td>
<td>4.44 6%</td>
<td>4.19 7%</td>
<td>4.68 5%</td>
<td>4.28 11%</td>
<td>4.32 3%</td>
<td>4.22 11%</td>
</tr>
<tr>
<td>6. Individual researchers</td>
<td>4.40 7%</td>
<td>4.43 5%</td>
<td>4.39 7%</td>
<td>4.30 0%</td>
<td>4.45 11%</td>
<td>4.43 3%</td>
<td>4.27 10%</td>
</tr>
<tr>
<td>7. Professional and scholarly associations who have researchers as members</td>
<td>3.78 13%</td>
<td>3.80 13%</td>
<td>3.67 10%</td>
<td>4.00 5%</td>
<td>3.62 13%</td>
<td>3.89 10%</td>
<td>3.60 13%</td>
</tr>
<tr>
<td>8. Government departments and agencies</td>
<td>3.55 12%</td>
<td>3.57 11%</td>
<td>3.46 9%</td>
<td>3.50 10%</td>
<td>3.49 13%</td>
<td>3.57 6%</td>
<td>3.55 13%</td>
</tr>
<tr>
<td>9. Creators and suppliers of software systems such as HR systems that encode and store IDs</td>
<td>3.48 13%</td>
<td>3.54 14%</td>
<td>3.35 11%</td>
<td>3.90 0%</td>
<td>3.31 13%</td>
<td>3.42 15%</td>
<td>3.42 10%</td>
</tr>
<tr>
<td>10. Politicians and planners</td>
<td>2.88 15%</td>
<td>2.92 17%</td>
<td>2.80 12%</td>
<td>3.06 15%</td>
<td>2.76 16%</td>
<td>2.93 13%</td>
<td>2.81 15%</td>
</tr>
<tr>
<td>11. Social science researchers who are interested in, for example, the mobility of researchers</td>
<td>3.09 11%</td>
<td>3.18 10%</td>
<td>3.15 9%</td>
<td>2.83 10%</td>
<td>3.31 13%</td>
<td>3.47 6%</td>
<td>2.83 11%</td>
</tr>
</tbody>
</table>
Appendix 5 - Interviewees
(excluding those wishing to remain anonymous).

** took part in the validation workshop (see Section 3.4).

<table>
<thead>
<tr>
<th>Name</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liz Allen</td>
<td>Wellcome Trust</td>
</tr>
<tr>
<td><strong>Niamh Brennan</strong></td>
<td>Assistant Librarian, Trinity College Dublin</td>
</tr>
<tr>
<td>Ian Carter</td>
<td>Director of Research and Enterprise, University of Sussex</td>
</tr>
<tr>
<td>Neil Chue Hong</td>
<td>Director, Software Sustainability Institute</td>
</tr>
<tr>
<td>Andrew Green</td>
<td>Chief Executive and Librarian, National Library of Wales</td>
</tr>
<tr>
<td>Kimberley Hackett</td>
<td>HEFCE</td>
</tr>
<tr>
<td>Amanda Hill</td>
<td>Hillbraith, Names and Mimas</td>
</tr>
<tr>
<td>Keith Jeffery</td>
<td>Director of IT and International Strategy, STFC</td>
</tr>
<tr>
<td>Stuart Lewis</td>
<td>Head of Digital Library, University of Edinburgh,</td>
</tr>
<tr>
<td>Andrew MacEwan</td>
<td>Authority Control Co-ordinator, British Library</td>
</tr>
<tr>
<td>Brian Matthews</td>
<td>Group leader, Scientific Information Group, Scientific Computing Dept, STFC</td>
</tr>
<tr>
<td>Valerie McCutcheon</td>
<td>Operations Manager, University of Glasgow</td>
</tr>
<tr>
<td>Daniel Needham</td>
<td>Mimas</td>
</tr>
<tr>
<td>Cameron Neylon</td>
<td>Director of Advocacy, PLoS (Public Library of Science)</td>
</tr>
<tr>
<td><strong>Mark Patterson</strong></td>
<td>Managing Executive Editor of eLife</td>
</tr>
<tr>
<td><strong>Stephen Pinfield</strong></td>
<td>Chief Information Officer, University of Nottingham</td>
</tr>
<tr>
<td>Richard Puttock</td>
<td>Head of Data and Management Information, HEFCE</td>
</tr>
<tr>
<td>Sally Rumsey</td>
<td>Principal Investigator, Oxford University Research Archive</td>
</tr>
<tr>
<td>Caroline Sutton</td>
<td>Publisher, Co-Action Publishing</td>
</tr>
<tr>
<td><strong>Michael Taylor</strong></td>
<td>Elsevier, UK</td>
</tr>
<tr>
<td>Paul Watson</td>
<td>Professor of Computer Science and Director of the Digital Institute, University of Newcastle</td>
</tr>
<tr>
<td>Jeremy Yates</td>
<td>Lecturer in Physics and Astronomy, UCL</td>
</tr>
<tr>
<td>David Zeitlyn</td>
<td>Professor of Social Anthropology, University of Oxford</td>
</tr>
</tbody>
</table>