ePrints and PURE: discussion paper

Presented to the Institutional Repository Steering Group

George Macgregor

December 2013

University of Strathclyde

Contents

Figures

Tables

1. Introduction

1.1 Motivation and context

The purpose of this paper is to re-examine the merits of the present repository configuration at the University of Strathclyde, specifically the parallel operation of both ePrints [1] and PURE [2]. The paper will also explore the implications of alternative repository scenarios. This paper fulfils an action assigned at the Institutional Repository Steering Group (IRSG) meeting held on 05 September 2013.

Several comparative analyses of ePrints and PURE have been undertaken by the IRSG over recent years. These analyses have tended to focus on the functionality of both systems with the aim of rationalising and determining which should become the University's principal research repository. It is not the intention of this paper to repeat these analyses because whilst they were deemed necessary at the time they were inadequate insofar as they focused on functionality at the expense of wider considerations. Previous analyses also failed to accept that Institutional Repositories (IRs) and Current Research Information Systems (CRISs), although demonstrating overlapping functionality and content, have evolved to fulfil different purposes. For these reasons this document can be considered a discussion paper to inform the decision making of the IRSG rather than a direct comparison of the technical features available in ePrints and PURE. This should enable informed decision making concerning the future of repositories at the institution.

It is impossible to cover every issue worthy of discussion in this paper, or every technical issue. This paper therefore restricts itself to those identified as the most significant under the repository typology used in section 2.1. A "conclusions" section has also been omitted to preserve the decision making remit of the ISRG. Where technical deficiencies of systems are identified, these are not intended to derogate and, in most cases, such deficiencies could be addressed with an appropriate programme of development work; their inclusion is more to inform a discussion about how alternative repository scenarios might function and the risks involved.

The paper is structured as follows: The remainder of this chapter explains the current institutional repository configuration and describes the issues surrounding any direct comparison of IR and CRIS implementations. Section 2 provides the majority of the discursive content, using an IR typology as the basis for discussion. Finally, section 3 sets out a series of feasible IR scenarios to be considered by the IRSG, with areas of risk, opportunity, etc. highlighted.

1.2 ePrints and PURE at the University of Strathclyde

As the readership of this paper will be aware, PURE has been the entry point for users wishing to deposit research content in ePrints (Strathprints) for several years. Validated content is pushed to ePrints using PURE's synchronisation framework and default connectors. PURE's connectors are a standardised, proprietary development framework enabling complex two-way exchange of data with IRs using WebDAV [3] and XSLT for object translation. Both simple submit operations and more advanced update and delete operations are handled by the connector. These IR connectors are part of PURE and have been developed to interoperate with leading IR platforms (i.e. ePrints, DSpace, FEDORA and Equella) but are maintained according to external needs. The result of PURE integration with an IR is that metadata and full-text objects are automatically deposited in the repository as they are submitted to PURE (following appropriate local validation processes). Research content deposited in PURE is also surfaced in the PUREPortal [2] (KnowledgeBase) which, owing to its integration within PURE, does not rely on any M2M interactions. Research content is therefore surfaced in two places: Strathprints and the KnowledgeBase [4].

Whilst the aforementioned repository configuration may appear peculiar to those within the University of Strathclyde, administering an IR in parallel with a CRIS is actually de rigueur in the UK [5] and beyond [6], [7]. Recent research conducted by the RepositoryNet+ Project [5] surveyed the UK repository landscape and found that of the small number of institutions deploying a CRIS* (31), the majority (27) ran an OAI-PMH compliant IR in parallel. The RepositoryNet+ survey also found a long tail of over 120 standalone IR implementations.

1.3 IRs and CRISs: definitions and overlap

The principal reason why most institutions have elected to run an IR and CRIS in parallel is because they are essentially different systems, originally devised for different purposes [8]. An IR is a system designed principally for gathering, disseminating and preserving the intellectual output of a research institution and, in so doing, fulfilling the goals of Open Access (OA), generating global visibility for institutional research and collecting institutional research content in a single digital location [8]–[12]. IRs also provide a useful mechanism for better storing, exposing and preserving other research content that might otherwise go unsurfaced, such as technical and project reports, working papers, patents, datasets, software, theses and other forms of "grey" literature

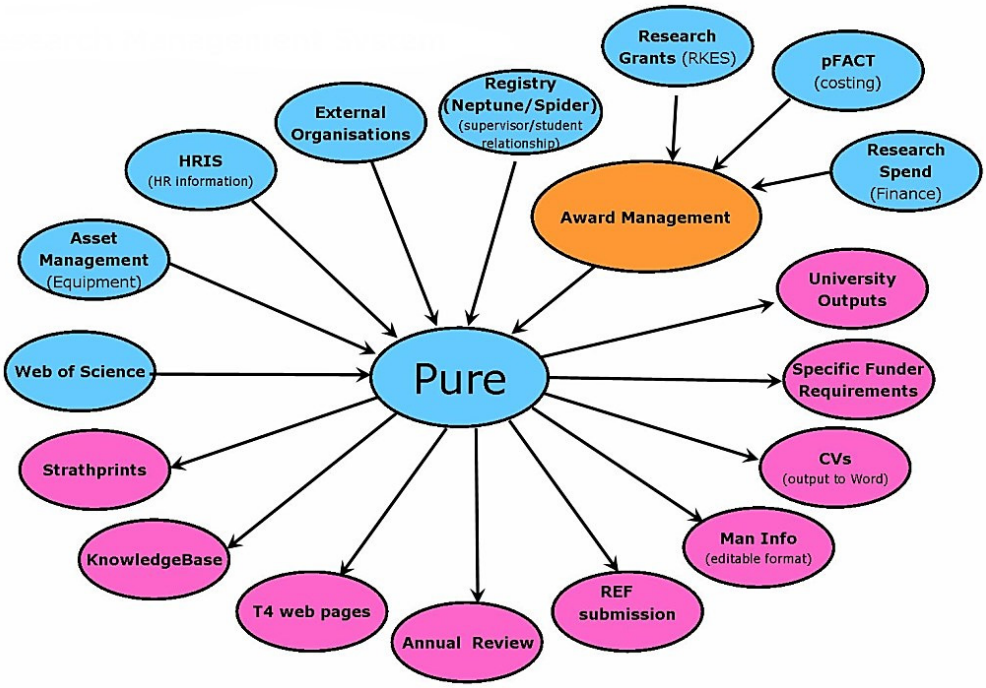

Figure 1: PURE at Strathclyde, as an example of a CRIS, pulling together disparate research relevant information and data.

[9], [10]. A CRIS, by contrast, tends to be more holistic in the nature of the information and data it is designed to curate; providing a comprehensive overview of contemporaneous institutional research activity by drawing together information from a number of disparate research-relevant sources, thus enabling improved administrative processes within research-intensive organisations [9], [11]. Such disparate information sources include HR, finance, funding award data, expertise DBs, research assets, and (notably for this paper) research repositories [13]. The management information derived from the CRIS supports business intelligence activity, policy making, research evaluation, but also enables the documentation of research activities and a formal log of research [14] (Fig. 1).The system definitions provided above are worth restating in order to highlight the difficulties in attempting direct

* A mixture of PURE, Symplectic, Avedas Converis and bespoke solutions.

comparisons between IRs and CRISs. The two are not cognate and treating them as such is a false analogy.

Overlap between IRs and CRIS systems nevertheless occurs because the raw content typically managed using IRs forms one of a wide number of important research-relevant sources targeted by a CRIS system. This is reflected in the CRIS models presented in the literature, most of which include a repository as a data source to be harnessed by the CRIS, e.g. [8], [9], [15]. It is also reflected in the Common European Research Information Format (CERIF) - used as a reference data model by PURE - which includes the major class `cfResPubl` (Result Publication) [16]. Products such as the PUREPortal module [2], as exemplified by the Strathclyde KnowledgeBase [4], accentuate this overlap by automatically exposing some CRIS content (typically research publication data and researcher profiles) via a surface-web frontend that simulates IR behaviour.

Further overlap has occurred in the opposite direction too. The importance of IR content in supporting the administrative functions of research institutions has been reflected in the evolution of IR platforms, which increasingly allow plugins and extensions to mimic some CRIS behaviour. For example, by installing the relevant plugins both ePrints and DSpace can serve CERIF compliant data [17], [18] and support institutions in the management of REF1 data (staff details) and REF2 data (research outputs) [18]. Functionality such as this has been crucial for the majority of UK HEIs without a CRIS and has been instrumental in supporting their preparations for REF2014. Enlighten [19], built on ePrints and demonstrating links with HR systems and grant funding data, is considered an exemplar of this approach [20], [21]. For most institutions, however, the available plugins offer only partial CRIS functionality, specifically in the area of research output management.

2. IR comparative discussion

2.1 An IR typology for discussion

Since direct comparisons cannot be made between an IR and an entire CRIS (Strathprints and PURE), the more appropriate point of comparison thus becomes the KnowledgeBase since it purports to simulate IR functionality [22] and essentially replicates the content delivered by Strathprints. Both also use the wider CRIS as the depositing mechanism. It is apposite to note that a small number of institutions (section 1.2) have also elected to use the PUREPortal as a de facto IR, most notably the University of Dundee [22], [23] and the University of Edinburgh [24]†. It is therefore sensible to consider what an institutional repository is supposed achieve and/or deliver and use this as the basis for discussion.

The literature agrees [5], [10], [26]–[31] that the principal functions of an IR should be to:

1. Provide open access to original research content (through self-archiving or other repository population methods) and thereby support underlying changes to scholarly communication;
2. Create and maximise the global visibility and potential impact of an institution's scholarly research;
3. Through improved research visibility, raise the international prestige of an institution;
4. Collect, curate and preserve institutional research content in a single digital location;
5. Store and digitally preserve "grey" digital assets, e.g. project reports, technical papers, etc.;
6. Support system interoperability standards to enable participation in global IR networks central to content aggregation, data mining, creation of new discovery tools, and generating new and unexpected knowledge from repository content.

The use of IRs as a means of sharing open educational resources (OERs) is also highlighted by some [27], [28] and is supported by generic IR platforms, e.g. EdShare (ePrints) [32], JORUM (DSpace) [33], etc. There are also many other examples of IRs and subject-based repositories managing such content [34]; however, at Strathclyde content falling into this category is managed within Digitool and is therefore outside the scope of this discussion paper.

Using the above typology it is possible to identify two broad areas which should steer the focus of this discussion paper:

1. Global visibility and IR interoperability, and;
2. Collecting, curating and digitally preserving.

2.2 Global visibility and IR interoperability

Research visibility and ergo discoverability is an issue of immense importance to the Strategic Plan (2011-2015) [35]. Openly accessible research is an obvious component in operationalizing the Principal's desire to "[…] optimise our research profile by supporting publication strategies to improve the quality, number and impact of research publications" and will likely feature in the forthcoming Strathclyde "citations policy" due to be disseminated by the Strategy and Policy Directorate [36].

The improved external visibility of research content is arguably the most significant function of institutional repositories [37] and remains a central argument in the debate favouring OA more generally [38]. A growing body of evidence [39] supports the assertion that Green OA deposits enjoy a higher impact by attracting citations from authors who would otherwise be unable to access it [40]–

† University of Edinburgh still maintain the Edinburgh Research Archive (ERA) [25], a DSpace instance which now only focuses on the collection, dissemination and preservation of theses, project reports, working papers and other grey materials.

[42]. In disciplines demonstrating strong support for Green OA scientometric analyses have estimated the consequent increases in citation impact to be in excess of 250% in some cases [43], [44]. Similar analyses also corroborate the importance of pre-print deposits in generating an immediate citation advantage [45] by accelerating citation impact, a consequence of research dissemination occurring prior to official publication [46], [47].

Within this context the technical mechanisms used by repositories to enable visibility become highly significant. Both Strathprints and the KnowledgeBase are visible to search engines to varying degrees. ePrints, like a number of IR platforms, also supports a wide variety of technical standards and protocols, all of which are designed to facilitate system interoperability. This interoperability supports a number of content discovery tools, contributes to content visibility, and enables participation in an international network of IRs providing unified access to research outputs that can then be (re)used by machines and/or researchers [31].

IR full-text downloads

One way of assessing IR visibility is of course to consider access / download logs. Figures for Strathprints can be accessed via IRStats [48] and IRUS-UK [49]. For the period 04-09-2013 (at IRUS-UK installation) to 30-11-2013 we yield the following figures:

- **IRStats**: 246,782
- **IRUS-UK**: 52,898 (across 4954 unique items)

Although IRStats can be useful to provide indicative figures on page visits, it remains clumsy because it does not distinguish between abstract views or full-text downloads, nor does it control for robots, usual user behaviour, suspicious access activity, etc. More meaningful are the full-text download figures that IRUS-UK delivers, as these better provide an indication of actual IR content discovery and usage [50], [51]. IRUS-UK counts full-text downloads only, as per COUNTER conventions [52] thus unusual usage and robots are also eliminated from the figures. In fact, IRUS-UK goes beyond COUNTER to include robots not identified in the COUNTER specification [49].

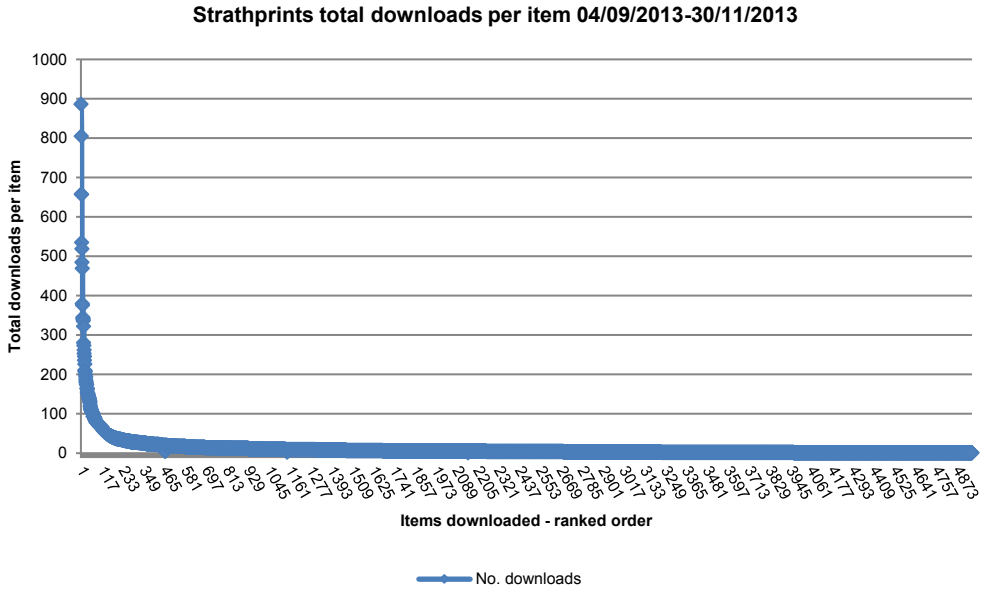

Figure 2: Strathprints total full-text downloads 09/2013-11/2013, using IRUS-UK.

52,898 full-text downloads of 4954 unique items in less than a three month period compares favourably with other institutions and infers an annual figure of circa 220,000. It should be remembered that although Strathprints has circa 32,000 “items” deposited, only 7941 (25%) have full-

text content and/or digital objects available. The figure of 52,898 is therefore generated from only 25% of IR content. These figures obviously do not include those users who discovered embargoed or unavailable Strathprints content (via metadata) and then submitted a request using the ePrints request button* [53] and/or followed a DOI to publisher version.

It is worth commenting that the full-text download figures from IRUS-UK installation to end November 2013 indicate a Pareto affect. In other words, almost 20% of the full-text content in Strathprints accounts for 80% (circa 39,000) of the total downloads. Although a (very) long tail of small download figures can be observed for the remaining content (Fig. 2), this still accounts for almost 15,000 downloads, and it should always be remembered that even a single download can be highly significant in the right context.

Unfortunately, COUNTER compliant full-text download statistics are unavailable for the KnowledgeBase thus preventing any meaningful comparative analysis. In the short time that Strathprints has been added to the institutional Google Analytic (GA) account† it can nevertheless be observed that the total amount of web traffic to both Strathprints and the KnowledgeBase is comparable. Both services reside within the top 5 most visited within the Strathclyde domain. Owing to the nature of the data, direct comparisons within GA are again problematic but, from data available on top path visits, it is evident that the KnowledgeBase is less of a conduit to research content than to staff profile pages, project summaries, departmental research group pages, etc. (Appendix A – Table 5). Such traffic is clearly important in helping to publicise Strathclyde as a research intensive institution, whilst also fostering the institution's global reputation as a leading technological university; but data appear to infer that direct access to research content (as distinct from other KnowledgeBase content) may not be as effective within the KnowledgeBase as it could be. An opportunity to make the KnowledgeBase "work harder" therefore presents itself and examples are highlighted elsewhere in this paper and outlined in section 3.1. By contrast the majority of traffic to Strathprints is to ePrints abstract pages, a large proportion of which we know – based on IRUS-UK data – are converted into full-text downloads (Appendix A – Table 6). Data also suggest that the KnowledgeBase may be less externally facing, as evidenced by the high proportion of the IP address visits from Glasgow and the Greater Glasgow area (Appendix A – Table 7). Strathprints is not without local page visits – and a proportion of local visits is to be expected in any IR – but the geographical spread of the visits Strathprints attracts and the significantly higher bounce rate is suggestive of higher levels of external search provider discovery (e.g. search engine, harvesting service, etc.) (Appendix A – Table 8).

OAI-PMH

Perhaps the most distinguishing feature of IRs are their adherence to the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) [54]. OAI-PMH provides a low-barrier mechanism for repository interoperability enabling repositories (i.e. data providers) to expose structured metadata adhering to a variety of metadata schema via the protocol. Discovery services (or "Service Providers", as defined by OAI-PMH) can initiate OAI-PMH service requests to harvest that structured data for local discovery purposes. Most IRs serve data adhering to DC and/or related application profiles. This is true of Strathprints, which also serves data according to the Metadata Object Description Schema (MODS) (using the Metadata Encoding and Transmission Standard (METS) as a wrapper) and the Digital Item Declaration Language (DIDL). The protocol also enables participation in the network of IRs central to fulfilling international initiatives, such as the Berlin Declaration on OA [12].

A number of high-profile discovery services based on harvesting exist, including CORE [55], BASE [56], JISC Institutional Repository Search [57] and OAIster [58]. Perhaps the most notable of these is the Knowledge Media Institute's (KMi) CORE service (Fig. 3). It is now the biggest full-text repository in the world, aggregating over 18 million OA full-text research papers from over 2,400 repositories and

* Estimated to be circa 10 per week. Similar functionality does not yet exist in the KnowledgeBase.

† Google Analytics was previously removed from Strathprints in 2011 owing to concerns surrounding the use of cookies under the EU ePrivacy Directive 2009/136/EC.

harnessing that data to provide a range of value-added services [59]. CORE's goal is to function as an OA platform to provide user discovery tools (including mobile apps), programmable access to aggregated metadata and content for those wishing to build new applications (e.g. text-mining, bibliometric tools, etc.) [60], and improving local repository visibility [59], [61]. HEFCE and RCUK have also expressed interest in the potential of using CORE as an OA compliance registry, i.e. a single service to interrogate for compliance with OA mandates. OAIster also remains highly significant as it provides a single discovery tool for the records of over 30 million digital resources harvested from OAI-PMH compliant collections worldwide which, in turn, also populates the over 2 billion items discoverable via WorldCat [62].

Strathprints is harvested by all the aforementioned services thus providing additional user discovery mechanisms and replicating metadata pertaining to Strathclyde research. There is also evidence that selective harvesting is conducted by certain subject-based Service Providers (using "setName") and smaller harvesters in countries such as France and India. Adherence to OAI-PMH also means that Strathprints features in the Registry of Open Access Repositories (ROAR) [63] and the Directory of Open Access Repositories (OpenDOAR) [64], thus enabling new Service Providers to target Strathprints as a data source thereby further exposing Strathclyde research content curated in Strathprints. It is also important to note that OAI-PMH is the principal mechanism by which the University Library feeds the SUPrimo [65] searches of Strathclyde research (Fig. 4).

Figure 3: Strathprints content harvested and surfaced by CORE.

The JISC's Open Mirror project is currently exploring the development of a UK discovery service delivering access to UK research content, harvested from UK repositories [66]. JISC is motivated by the need to provide a one-stop-shop for all openly accessible UK research output and a desire to improve the discoverability of UK research more generally. OAI-PMH is anticipated to be the protocol used to populate Open Mirror.

It is possible for PUREPortal content to be exposed via an OAI-PMH-*style* interface but this remains disabled in the KnowledgeBase. This is primarily because Strathprints represents an established OAI target for the institution and, as such, research content is already being routinely harvested and surfaced by search providers. It should nevertheless be noted that the protocol - as implemented by Atira/SciVal in the PUREPortal - remains OAI-PMH noncompliant and therefore unusable to most

harvesting services. Important protocol requests/responses are unsupported and DC metadata is exposed incorrectly [67]. These technical deficiencies have been noted for some time by Atira/SciVal and their rectification was raised as a priority for future development at the May 2013 PURE UK User Group, resulting in the generation of PURE JIRAs* #6061, #6063 and #6306. Development work will be directed according to a recently drafted OAI-PMH requirements specification, edited and largely drafted by James Toon (University of Edinburgh) [67], to make the PUREPortal OAI-PMH compliant. This work will include support for OAI-DC with support for the EU's OpenAIRE application profile [68]. It has been confirmed that support for other metadata schema and compliance with the RCUK application profile, RIOXX [69], are out of scope [67]. Atira/SciVal is expected to complete this work by mid-2014.

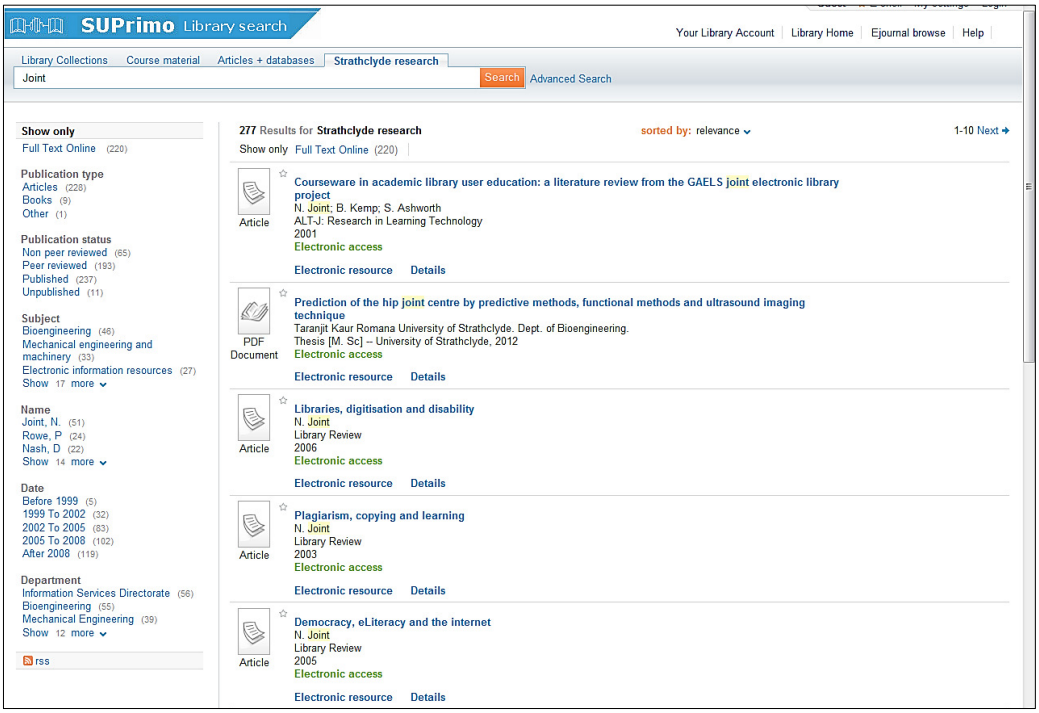

Figure 4: Strathprints research content harvested for SUPrimo search. Image shows content by Nick Joint.

Whilst Atira/SciVal has been criticised by the PURE UK User Group for its failure to "fix" problems with OAI-PMH quickly, it should be remembered that Atira/SciVal are managing development requirements across an entire CRIS, and not just the PUREPortal. Most institutions are operating an OAI-PMH compliant repository in parallel thereby disincentivising development activity from a software firm driven by commercial incentives. Easy integration with IR platforms is also cited as a key selling point for PURE more generally [2] thereby deprioritising OAI-PMH compliance of the PUREPortal. That it has taken some time to address this issue is therefore unsurprising.

OAI-PMH and compliance frameworks

Although tangential to research content visibility, it should finally be noted that OAI-PMH is the basis for a number of research compliance frameworks. IR application profiles for OpenAIRE [68] and RIOXX [69] are designed to provide compliance with the EU and RCUK funding policies on open access. For example, RIOXX - soon to be accompanied by V4OA [70] - focuses on metadata content rules and applying consistency to additional metadata elements used to record research funder and project/grant identifiers, thereby enabling the tracking of RCUK compliance across scholarly systems via harvesting. RIOXX compliance can be achieved in ePrints by installing the RIOXX plugin from the ePrints Bazaar [71]. Enlighten [19] provides a demonstrator of this. Guidelines on configuring IRs for OpenAIRE are also available [72].

* JIRA is the bug tracking and software development package used by Atira/SciVal to support the project management of PURE development activities.

At present neither Strathprints nor the KnowledgeBase adhere to OpenAIRE or RIOXX. Strathprints compliance with both could be implemented as per above but, based on the current parallel repository configuration, is precluded from doing so. Compliance with profiles such as these will nevertheless be increasingly necessary in future and, perhaps, mandatory. The KnowledgeBase will be OpenAIRE compliant once the aforementioned fixes have been applied to OAI-PMH and will make content discoverable via the OpenAIRE harvester [68], which provides aggregated access to openly accessible EU funded research. RIOXX will remain outside the scope of Atira/SciVal's work; the status of RIOXX, however, remains uncertain partly owing to the uncertainty surrounding the future of RCUK's OA policy. RIOXX is "strongly supported" by RCUK but RCUK has stopped short of mandating institutional compliance with the profile. This position may change with the finalisation of V4OA in 2014.

Although the process for monitoring OA compliance as part of REF2020 has yet to be decided, the HEFCE consultation [73] did infer that all research outputs would need to be deposited in an IR and that compliance checks would be made. It is difficult to imagine HEFCE conducting compliance checking within such a framework without using OAI-PMH and it is therefore conceivable that HEFCE will follow RCUK and the EU in this regard. The future of any repository configuration at the University of Strathclyde should therefore ensure it is capable of servicing the compliance requirements of funding bodies.

Other forms of open structured data

Many institutional repository platforms expose structured data about research content using a number of schema within pages. Some of these are used primarily to support web translator interoperability and/or citation exportation; but others are used to feed other discovery services and anticipate services that may wish to consume alternative forms of structured data for its indexes. Structured data is also used in search engine crawling, for example [37]. Metadata schema supported by ePrints includes Dublin Core, MODS, METS, DIDL, ePrints Application Profile and others. XML alternatives such as JSON are supported. Data are also exposed in RDF following the RDF/XML, N3 and N-triples serialisations, thus supporting semantically aware search agents.

Perhaps most crucially, in-page metadata is exposed by Strathprints according to the ePrints schema provided within HTML <meta> tags. This is significant because this schema is one of the few structured data formats formally supported by Google Scholar (GS) [74], [75]. Others schema falling within GS's inclusion guidelines are HighWire, BE Press and PRISM [74]. DC <meta> tags are also supported by GS and are also delivered by ePrints but are less favoured by GS as only "qualified" DC can provide the specificity required in the treatment of some bibliographic data elements. Sites that do not implement the aforementioned standards risk exclusion from GS [74]. It should also be noted that at time of writing GS recommends only three repository platforms for the delivery of content: ePrints, DSpace and Digital Commons [76].

These aforementioned factors may go some way to explaining the preference GS demonstrates for serving results from Strathprints over those of the KnowledgeBase (Fig. 5). Experiments exploring the customisation of DSpace to better adhere to GS inclusion guidelines have, for example, found dramatic differences in GS indexing ratios when DSpace - which does not by default support all of the above technical features – was developed to expose HTML metadata according to the HighWire schema [75]. PUREPortal pages do not support any of the above features and are therefore not optimised for GS inclusion. In light of this the University of Edinburgh has made bespoke changes to their PUREPortal pages to expose HighWire tags; although local expertise suggests that so far the results have been inconclusive [77]. This is perhaps attributable to the lack of indexing freshness demonstrated by GS when compared to the universal search offerings of Google and Bing. Whilst new IR content will normally be picked up by GS within a month, any subsequent updates to that content takes - by their own admission - on average 6-9 months to be indexed, with "several years" for larger websites [74]. This makes it very difficult to observe whether any adjustments to the

PUREPortal have made it more GS friendly. Edinburgh's change, however, immediately improves translator interoperability for reference management software (see Section 3.4) and can improve text summarisation and indexing in universal search engines [37]. Comparable benefits could therefore be achieved in the KnowledgeBase if implemented (Fig. 8).

It should be noted that GS maintains a separate index to Google's universal search and also uses its own crawlers that are specifically optimised to index from appropriate content sources, such as repositories and publisher sites [75]. High visibility within GS as a consequence of following the inclusion guidelines does not therefore guarantee comprehensive indexing or high visibility within universal search. A similar indexing and crawling approach is adopted by Microsoft Academic Search [78] - a service which has more recently attracted favourable attention owing to its advantages over GS [79], [80] - however, again, high visibility with Academic Search does not guarantee high visibility in Bing.

Although the number of GS referrals to Strathprints is in the hundreds of thousands its significance should not be overstated in light of other referring websites. IRStats on Strathprints [48], for example, indicates that 10% of all referrals came from GS in the year to October 2013, with the rest coming from other search agents, repository users, and so forth. Given the above discussion the figure is likely to be less in the KnowledgeBase. But optimising any IR for a single search service - whilst important since even a single download could be highly significant in academic terms - should not become an obsession at the expense of optimisation for alternative discovery tools. In general, better data on web traffic is required to understand referrals and Google Analytics has recently been re-added to Strathprints.

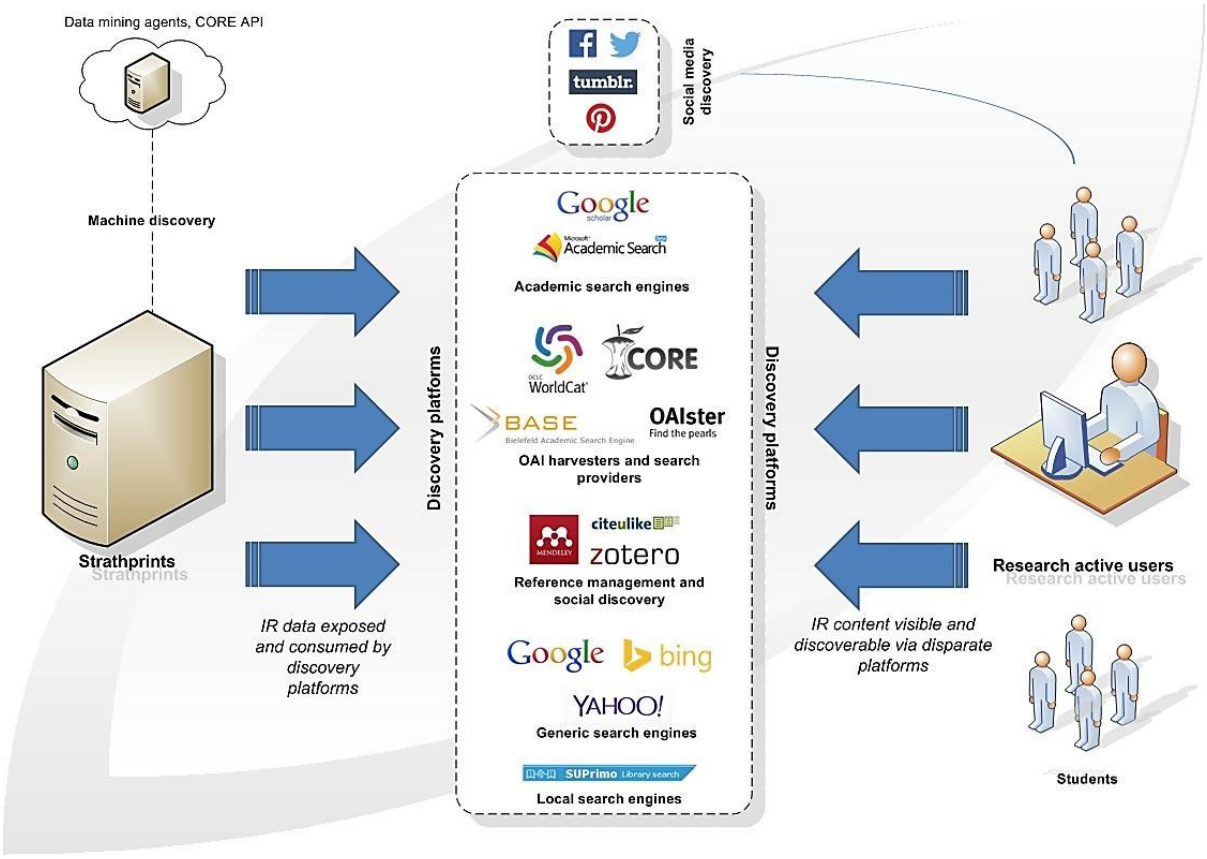

Figure 5: Diagram of Strathprints visibility to discovery platforms (including machine and social discovery), some enabled through ePrints adherence to technical protocols such as OAI-PMH and structured data exposure.

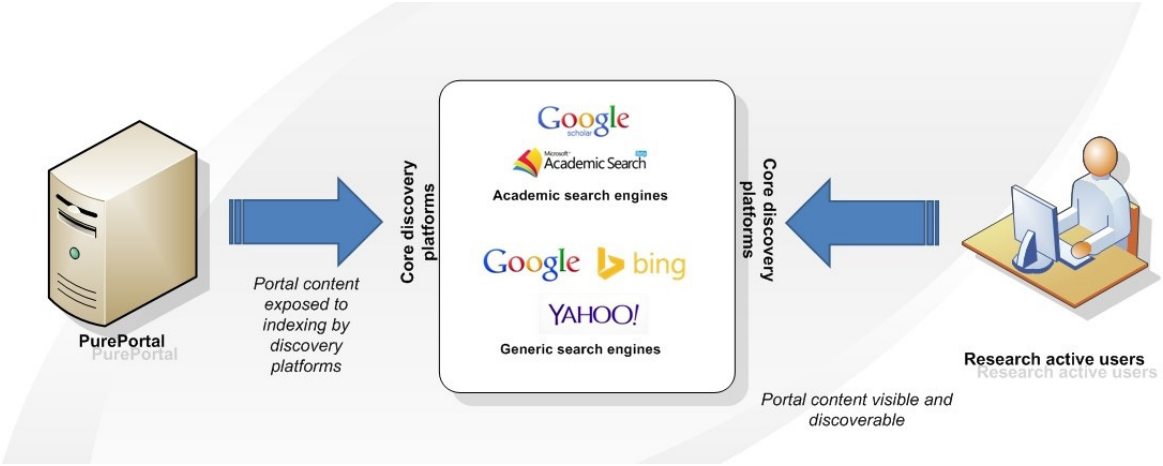

Figure 6: Diagram of current PUREPortal (KnowledgeBase) visibility to discovery platforms.

Sitemap support

The Sitemap protocol [81] was developed by Google, Microsoft and Yahoo! and provides a means of notifying search engines about the URLs on a website that are available for crawling. It forms an important part of search engine crawling strategies [82], [83]. A sitemap is essentially an XML file that lists crawlable URLs with additional metadata about each URL (e.g. when it was last updated, how often it usually changes, and how important it is relative to other URLs in the site) so that search engines can better crawl the site. Search engine discovery of new URLs is principally through new hyperlinks so sitemap data may not influence the initial discovery of URLs; but it supplements hyperlink discovery by enabling better and more intelligent crawling, particularly by helping crawlers identify recently updated content (the "freshness" of traditional indexing is a known limitation of standard crawlers) and/or better crawling content in future [84].

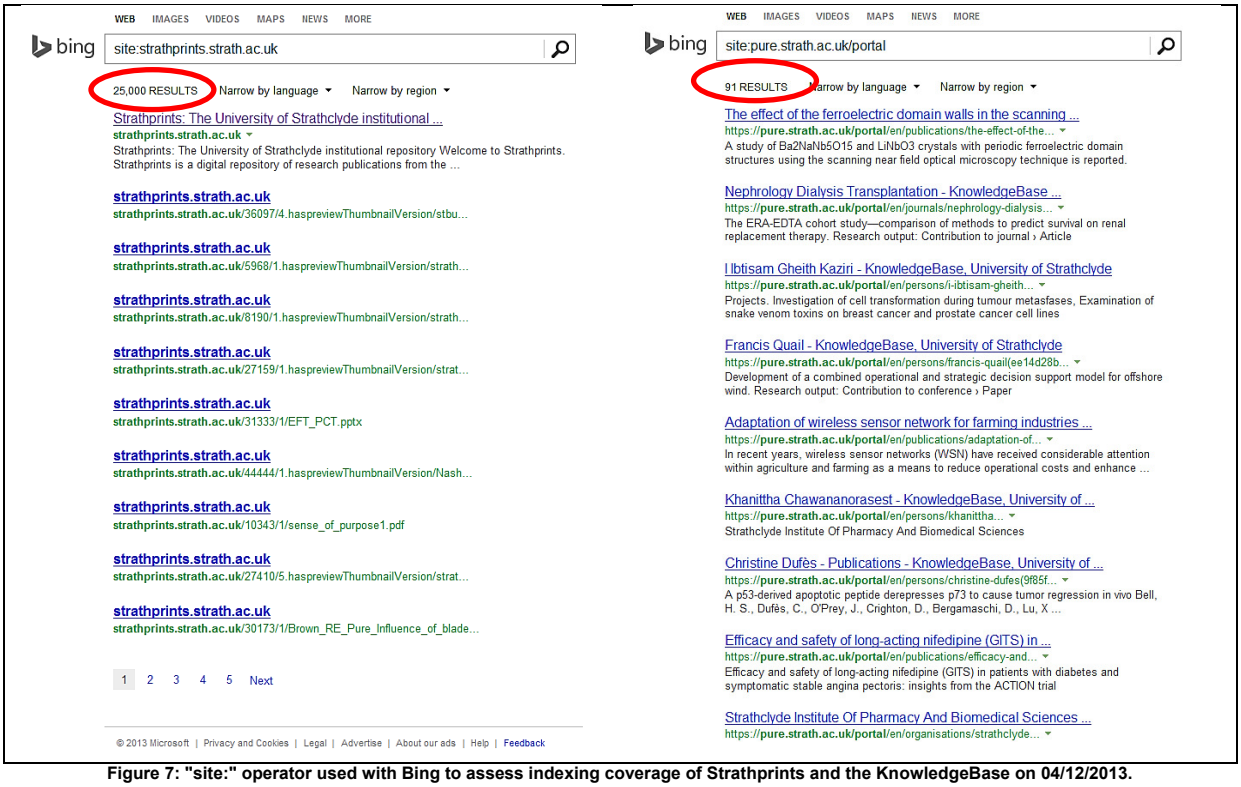

Figure 7: "site:" operator used with Bing to assess indexing coverage of Strathprints and the KnowledgeBase on 04/12/2013.

ePrints provides support for sitemaps; although the current default implementation augments the existing protocol by using extensions proposed by the DERI team [85] for the publishing "Semantic

Sitemaps" [86]. These are essentially RDF/XML dumps of Strathprints content published following standard Linked Data rules [87]. See for example: http://goo.gl/S9u2bs. The PUREPortal does not appear to support sitemaps [77]; but by using the `site:` operator command we can observe that Googlebot index penetration of the KnowledgeBase for "universal search" appears very healthy. This is less the case in Bing (Fig. 7) and Yahoo!, the latter of which uses the same indexes as Bing. This may be less of a concern; although it should be remembered that Bing is growing in popularity and now accounts for almost 20% of all web searches, thus total "Bing powered" searches now account for almost a third of all searches [88]. Ensuring high indexing penetration from these services is therefore important and sitemap support is a contributory tool for achieving this.

Web translator support

Section 3.2 noted the use of structured data to drive web translator interoperability and reference manager import. Web translators are automatically activated by users of reference management software (e.g. Mendeley, Zotero, CiteULike, etc.) when visiting data compliant discovery services or websites, thus allowing users to quickly save detailed bibliographic data about an item they are viewing or have downloaded and – depending on client software configuration – capture a digital version of the item (normally a PDF). Saved items can then be easily revisited within the software, annotated and cited in future work. Repository data compliancy essentially means exposing structured data according to schema such as DC and ePrints [89], [90], and enabling the clean parsing of DOIs for interrogating CrossRef [91].

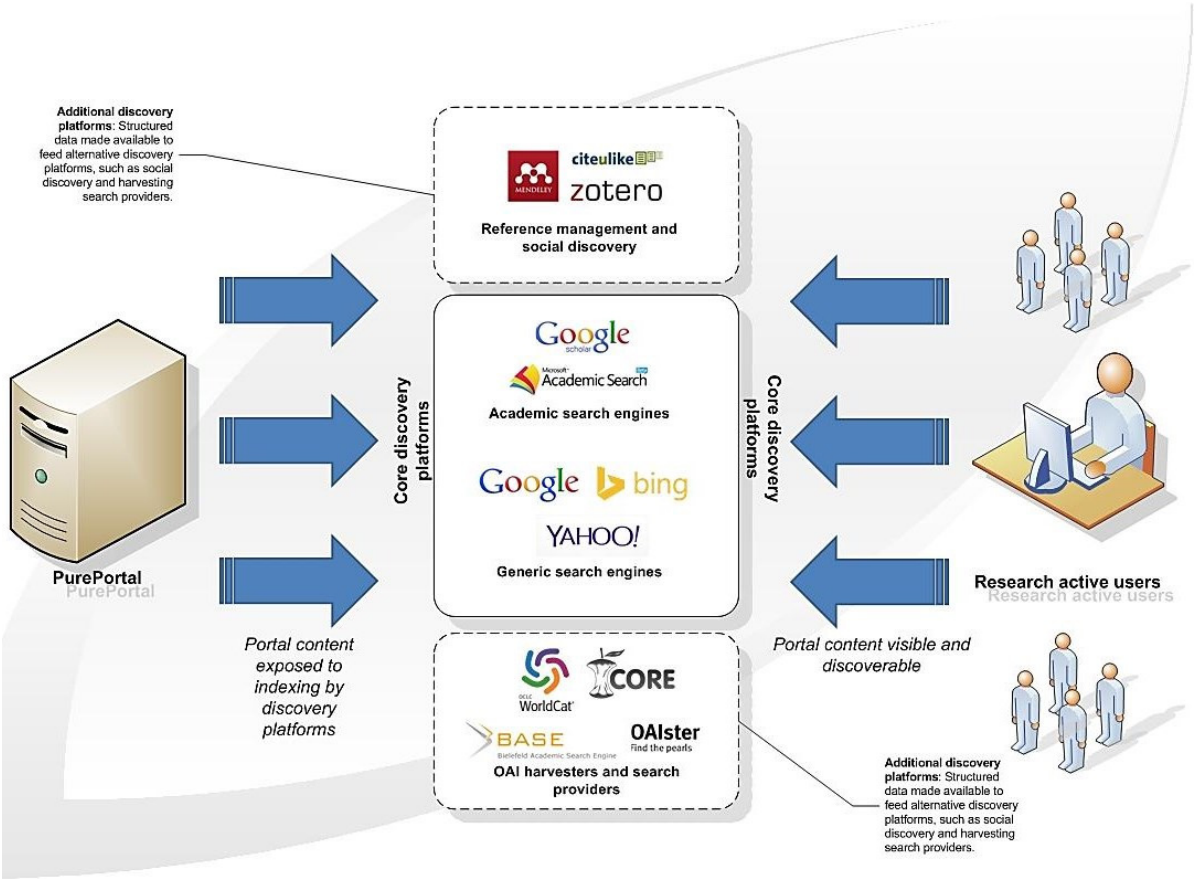

Figure 8: Diagram of PUREPortal (KnowledgeBase) visibility to discovery platforms after OAI-PMH compliance and structured data implemented.

Interoperating with such tools is important because they are now increasingly the principal means of research discovery for a growing number of users [92] and demonstrate growing popularity among academic researchers [93]. Mendeley [94] and Zotero [95], in particular, offer powerful social networking features supported by the cloud storage of users' research collections and references [93]. Users can choose to open their personal research collections or share them with particular groups.

As a discovery tool Mendeley functions much like Last.fm [96], [97], leading users to key resources, potential collaborators, and highlighting connections within their field of study. Apps created using the Mendeley API also leverage the crowd-based nature of discovery through recommendation and expertise, enabling the sharing of fully annotated research papers. These features are provided in addition to advanced retrieval tools which enable searching across collections curated by users (currently estimated to be in excess of 60 million [96]). The growing importance of user discovery through these types of tools is considered the principal motivation for Elsevier's recent acquisition of Mendeley [96].

Strathprints adheres to the specified data requirements of the leading reference management tools. Bespoke development of the PUREPortal by the University of Edinburgh, as noted previously, has enabled better interoperability with such research tools, and could be implemented in the KnowledgeBase with development [77] (Fig. 8).

Support for this should not be overstated. Mendeley, for example, has alternative ways of extracting metadata from content, e.g. extracting embedded file metadata, guessing probable metadata attributes by analysing document structure, etc.; however, these methods are less reliable and, if used, will often yield incomplete metadata [98].

2.3 Collecting, curating and digitally preserving

"Institutional content"

Central to a productive CRIS is the expectation that research active staff maintain and curate their research profiles and maintain CVs within the CRIS. It is satisfying to note that PURE has become an embedded service at Strathclyde and one with which most research active staff will engage frequently. This level of staff engagement is essential to ensure accurate business intelligence reporting on institutional research activity.

Recall from our IR typology that an important function of an IR is to enable the collection, curation and preservation of institutional content in a single digital location. Encompassed within the concept of "institutional content" is the idea that content is more than simply a collection of conventional scholarly communication outputs (e.g. journal articles and conference papers); it is essential to also capture, disseminate and digitally preserve "grey" digital assets. Research publication data pulled into PURE is pushed to Strathprints and mirrored in the KnowledgeBase. Both IRs therefore surface exactly the same content and both, in their current instantiation, technically fulfil the remit of this IR function. It is also worth noting that both IRs currently expose a small amount of grey content, encompassing metadata on a mixture of circa 1,200 commissioned reports, working or discussion papers, visual resources, software, datasets, and so forth, of which approximately 400 have a digital object attached. Improving the collection and exposure of grey content is therefore more a question of advocacy than a technical issue for Strathprints or the KnowledgeBase.

Where there is a conceptual distinction to be made is on the issue of what constitutes "institutional content". The CRIS requirement for comprehensive staff research profiles, combined with the ease with which publication data from Scopus, arXiv and PubMed can be imported, means that Strathprints and the KnowledgeBase increasingly surface content that is not institutionally "owned", i.e. content that may have been authored by staff members while they were employed at a different (perhaps competing?) institution.

From a holistic OA perspective this is perceived as a positive development, and some repository managers are content to expose non-institutional content because it transcends local considerations and is largely consistent with the global OA agenda. Others, however, are more reticent and have conceded that the collection policy of their IR has to be flexible if staff engagement with their IR is to be maintained [99]. In many ways, this tension crystallises one of the fundamental differences in

function between IRs and CRIS systems. The University of Edinburgh, for example, has admitted that this issue was one of several factors influencing their decision to retire their parallel IR/CRIS configuration [99]. It is difficult to estimate the proportion of IR content that is not Strathclyde owned within Strathprints and the KnowledgeBase; but, based on the proportion of output currently pushed for validation in a typical week, it is likely to be in excess of 20% of all IR content.

This raises questions about the institutional content collection policies of what are essentially different systems, devised for different purposes. The KnowledgeBase is underpinned by a CRIS and by definition has a shorter term view of the information lifecycle. Content collection is driven primarily by contemporaneous information on research activity to drive business intelligence on institutional research performance [9], [11]. Consideration of the longer term preservation responsibilities, and the benefits that may arise from this, are therefore not a priority within a CRIS environment. This is also reflected in the functionality of most CRISs (including PURE [2]), which do not provide preservation tools and instead rely on content preservation jobs to push content to external IRs (e.g. ePrints, DSpace, etc.).

By contrast an IR is less focussed on contemporaneous content collection only. As we have noted from our IR typology in section 2.1, an IR is important for exposing current research content and improving research visibility; but it also adopts a longer-term view of information lifecycles and remains an important tool for collecting, curating and digitally preserving institutional research content [26], [28], [29], [31], thus ensuring long-term access to institutional research. This may also include historic research outputs, the exposure of which may contribute to the concept of "institutional memory" (or "corporate memory") and function to enhance the reputation and prestige of the institution within the sector (e.g. promoting the rediscovery of seminal Strathclyde research). This latter requirement segues into the archival nature of IRs, and in many IR implementations there are looser collection policies such that miscellaneous items managed by an institution (either collected by or gifted to the institution) are deposited to support known or future research interests, thus enriching existing digital collections and promoting both the reputation and research potential of the institution. This activity can also support some administrative functions by providing a body of material that can be consulted to understand organisational history and management.

The CRIS emphasis on shorter term information lifecycles and its need to ingest data from a wide variety of sources (including publication data that might not be institutionally owned by Strathclyde) is a divergence from the traditional IR typology and may not be consistent with the need to fulfil the requirements of institutional memory, or to accurately preserve and expose the intellectual content of Strathclyde research over time.

Digital preservation

Digital preservation has been a topic of discussion within the IR community for as long as IRs have existed [100] and it remains an important component of any IR typology. Institutional repositories aim to preserve and make accessible digital content on a long-term basis. Digital preservation and long-term access are therefore inextricably linked and underpin the OA movement's goal of providing a growing corpus of research content that can be accessed by both humans and machines.

Delivering long-term access to digital objects in an IR requires considerable planning and resource, and it is apposite to note that at time of writing the institution has no detailed long-term digital preservation strategy; however, supporting the preservation of - and long-term access to - IR content also requires system support to plan, identify file format risks, provide open storage and cloud solutions, complex digital object export, history modules, rights metadata, and so forth. The maturity of an IR platform such as ePrints brings with it significant system support for digital preservation activities [101]. Successive ePrints digital preservation projects [102]–[105], some involving the University of Oxford, the British Library and the National Archives [105], [106], have produced a variety of preservation toolkits [107] and repository plugins [108], [109] all designed to support preservation activities and ergo the long-term access to institutional research content. Other tools

have also been integrated as part of ePrints from version 3, including METS and DIDL support for complex object export and conversion, a history module for preservation metadata maintenance, and preservation rights declaration management [110]. Some of this functionality was a direct output of the Preserv project [105], involving complex digital preservation problems encountered by partners such as the British Library [110]. The ePrints Bazaar, available to versions 3.3† and above, provides access to additional plugins, including the Arkivum A-Stor Storage Backend plugin [111], enabling cloud storage and archiving of content, file health monitoring, and migration. ePrints also demonstrates SWORD (Simple Web-service Offering Repository Deposit) compliance [112] thus enabling the automatic deposit of IR content in, say, other repositories (e.g. funder mandated IRs or subject repositories) but potentially to repositories with a dedicated preservation remit, such as the anticipated Open Mirror repository [66].

As we have noted, consideration of the longer term preservation responsibilities have not been addressed by CRIS vendors because their data and information management needs have understandably not necessitated it. Where digital preservation has arisen as a concern it has been considered outside the scope of a CRIS and has been resolved by pushing content to "long term storage", where IR preservation tools can be used. This is currently the approach adopted by Atira/SciVal for PURE.

It should be reported that PURE may in future demonstrate improved functionality in this area as it attempts to accommodate a desire from stakeholders to include support for the long-term management of research datasets. A programme of work led by the University of Strathclyde under the auspices of the Research Data Management Project (RDM) is currently working with Atira/SciVal and the PURE UK User Group to specify system functionality and data requirements for dataset management. It would be anticipated that any preservation planning and management tools implemented for RDM would be available for IR content within the PUREPortal. Atira/SciVal have also indicated that SWORD2 support may be included in a future release of PURE, although it is not yet a development priority and is omitted from the PURE Roadmap 2013-2015.

† Strathprints is currently operating on version 3.2.5.

3. Possible future IR scenarios

A number of alternative IR scenarios are feasible at the University of Strathclyde in light of the discussion points highlighted in sections 1.3 to 2.3. An additional objective of this paper (and an action from IRSG) is to list these for further discussion, highlighting potential areas of risk. A simple SWOT (Strengths, Weaknesses, Opportunities, Threats) framework appears most suitable for this task and is often deployed in system selection or implementation contexts, e.g. [113]–[115].

3.1 IR scenario #1: status quo

IR scenario #1 refers to the current IR configuration at Strathclyde, which is typified by the parallel operation of both Strathprints and KnowledgeBase (PURE) and the use of PURE's default connectors to push research content to Strathprints. Research content is therefore surfaced in two places: Strathprints and the KnowledgeBase. The principal benefit of this configuration concerns the way in which it can contribute to research visibility and the opportunities that are available to optimise content visibility through technical improvements to the KnowledgeBase, alongside Strathprints.

Table 1: IR scenario #1 SWOT matrix.

Internal environment / External environment		Strengths	Weaknesses
		• High visibility of Strathclyde research content via use of Strathprints technical protocols in harvesting services, social reference services, academic search engines (e.g. GS) and SUPrimo search, thus improving potential of research impact.	• Technical overhead of maintaining parallel systems serving same content; overhead for GM, systems team, infrastructure, RKES. • Difficult to understand true impact and use of research content spread across multiple IRs.
Opportunities	• Improvements to KnowledgeBase possible to maximise visibility of Strathclyde research content and promote research impact (e.g. OAI-PMH, HighWire data, sitemap support), thus combining with Strathprints to drive research visibility and impact. • Strathprints compliance with OAI-PMH will enable participation in – and extra visibility from - Open Mirror. • Scope for development of Strathprints (to ePrints V3.3+), with potential of improving UX, user engagement and data exposure. • Potential for improvements to data re-use across both IR platforms in a variety of institutional systems. • Opportunity to use Scopus institutional affiliation string for identification of historical Strathclyde research for backfilling KnowledgeBase / Strathprints.	• Parallel operation of Strathprints and KnowledgeBase provides multiple discovery channels for users. • Generally good surface web search agent indexing penetration of both KnowledgeBase and Strathprints. • Content metadata tools in PURE (e.g. Scopus, arXiv, etc.) enable simple staff deposit and data ingest • PURE cron job benefits (e.g. automatic assignation of ASJC classifications).	• Connector configuration prone to failure and requires maintenance, e.g. deposit errors, organisational changes pushed, etc.. • Potential end user confusion about the purpose of two systems with identical content. • The "Institutional memory" functions of IR poorly serviced in present configuration. • Overlap in aspects of Strathprints and KnowledgeBase service delivery to end users resulting in additional institutional cost.
Threats	• Lack of digital preservation support in PURE. • Long lead times for Atira/SciVal's implementation of PUREPortal technical improvements. • Atira/SciVal development of the PUREPortal deprioritised in favour of larger CRIS needs. • Lack of COUNTER compliant statistical tools for the KnowledgeBase, e.g. IRUS-UK.	• Maintenance of Strathprints ensures IR interoperability within global IR network (and text mining). • Wide staff engagement with PURE, thus contributing to Green OA deposit behaviour and population of Strathprints. • Re-use of Strathprints data demonstrated to populate T4 publication pages. • Productive collaborative relationship across ISD and RKES.	• Parallel IR operation complicates compliance with prevailing technical standards and protocols, as standards need to be reflected across two IRs (e.g. RIOXX/V4OA, OpenAIRE, etc.), one of which is driven by commercial decision. • Many improvements to Strathprints wedded to KnowledgeBase.

3.2 IR scenario #2: Strathprints retired

IR scenario #2 refers to a repository configuration in which Strathprints is retired entirely, thus leaving the KnowledgeBase as the only research repository at the University of Strathclyde. This scenario is best compared to the recently adopted position of the University of Dundee [23]. The principal attraction of this approach is the simplicity of maintaining only a single IR and the removal of any long-term IR storage systems. Significant risks accompany this approach, including a potential negative impact on research visibility and ergo the impact of Strathclyde research, the disenfranchisement of PURE/KnowledgeBase from a number of national and international initiatives, and the loss of a potentially valuable IR tool to manage and curate heterogeneous digital content (e.g. for the library, archives, long-term digital preservation, long-term OA, etc.). There also staffing implications for the management of KnowledgeBase research content. Maximising the impact of Strathclyde research is arguably the most important of these considerations and, as such, implementation of scenario #2 would in all likelihood require a lengthy preparatory period during which PURE / KnowledgeBase could be verified as supporting the same technical protocols as ePrints and was evaluated as delivering comparable visibility.

Table 2: IR scenario #2 (Strathprints retired) SWOT matrix.

Internal environment / External environment		Strengths	Weaknesses
		• Generally good surface web search agent indexing penetration of the KnowledgeBase. • Simplicity of maintaining only a single IR and the removal of connector overheads / technical issues associated with parallel long-term storage IR, interoperability considerations, and ownership/control.	• Loss of Strathprints contributing to lower visibility of Strathclyde research content and a reduction in discovery channels. • Opportunity cost of users misdirected to obsolete / or inactive content on Strathprints.
Opportunities	• Improvements to KnowledgeBase possible to enhance visibility of Strathclyde research content and promote research impact, e.g. OAI-PMH, HighWire data, sitemap support, social media.	• Wide staff engagement with PURE, thus - at least theoretically - maintaining current levels of Green OA deposit. • Simplicity of promoting only a single service to institutional end users. • Content metadata tools in PURE (e.g. Scopus, arXiv, etc.) enable simple staff deposit and data ingest.	• Lack of IR interoperability leading to poor participation in global IR network and low usage of Strathclyde research in text mining and/or discovery via harvesting search providers. • Responsibility for the management of KnowledgeBase research content moved from ISD to RKES, with probable staffing and negative knowledge management implications.
Threats	• Potential damage to citation impact of Strathclyde research and failure to meet Strategic Plan (2011-2015) objective of supporting publication strategies to improve the quality, number and impact of research publications. • Once productive collaborative relationships across ISD and RKES in the area of IR issues and OA degrade with the retirement of Strathprints. • Atira/SciVal development of the PUREPortal deprioritised in favour of larger CRIS needs. • Long lead times for Atira/SciVal's implementation of PUREPortal technical improvements and slow responses to prevailing technical standards or new initiatives. • Lack of digital preservation support in PURE and inability to support longer term OA and/or OA mandates. • Loss of Strathprints contributes to reduced digital content and preservation strategy within library and wider ISD.	• PURE cron job benefits (e.g. automatic assignation of ASJC classifications). • Nominal cost savings associated with retiring Strathprints (e.g. on-going infrastructure, maintenance overheads, etc.).	• Institutional disenfranchisement from national and international IR initiatives owing to deficit in PUREPortal functionality, e.g. Open Mirror. • Current lack of COUNTER compliant statistical tools for the KnowledgeBase, e.g. IRUS-UK. • Loss of valuable IR tool (Strathprints) to manage and curate heterogeneous digital content, e.g. library, archives, long-term digital preservation, long-term OA, etc. • SUPrimo research search requiring redevelopment to populate via PUREPortal source. • T4 publication page data grab redevelopment required to use PURE.

3.3 IR scenario #3: connector retired

The PURE default IR connector is currently used to automatically deposit metadata and full-text objects in Strathprints after they have been validated in PURE. IR scenario #3 refers to a repository configuration in which the PURE default connector is retired, thus leaving Strathprints and the KnowledgeBase as standalone systems. This scenario would be technically similar to the recently adopted position of the University of Edinburgh, which maintains Edinburgh Research Explorer (PUREPortal) [24] and Edinburgh Research Archive (DSpace) [25] as separate IRs and without the use of the PURE connector.

The principal attraction of this approach is the removal of the connector itself, thus removing the maintenance of the connector and enabling both systems to follow alternative development paths and pursue local departmental priorities, perhaps deferring to the IR and CRIS definitions outlined in section 1.3. Whilst attractive, this scenario could carry significant risks, including a potential negative impact on research visibility and ergo the impact of Strathclyde research, technical and "soft" issues in populating Strathprints after disconnection from PURE, a lack of efficacy that may accompany any IR ecosystem in which often overlapping content is managed across disparate IRs, and the disenfranchisement of PURE from a number of national and international initiatives. Maximising the impact of Strathclyde research is arguably the most important of these considerations and, like scenario #2, implementation of scenario #3 would in all likelihood require a lengthy preparatory period during which PURE / KnowledgeBase could be verified as supporting the same technical protocols as ePrints and was evaluated as delivering comparable visibility.

Table 3: IR scenario #3 (connector retired) SWOT matrix.

Internal environment / External environment		Strengths	Weaknesses
		• Visibility of disparate range of Strathclyde digital content via use of Strathprints and KnowledgeBase. • Collection of "institutional content" better fulfilling "traditional" IR remit of institutional memory for Strathprints; KnowledgeBase better serving "current" CRIS research content. • Demarcation of Strathprints and PURE, with clear IR and CRIS remits and less user confusion; distinct services to promote to users.	• Negative impact on Strathclyde research visibility and ergo impact owing to lack of appropriate technical protocols in KnowledgeBase. • Potential "data in" / "content in" problems for Strathprints, e.g. lack of input mechanism and user deposit / content notification, data creation and ingest deficit, Strathprints no longer benefitting from wide staff engagement with PURE, etc.
Opportunities	• PURE refocus on contemporaneous content. • Improvements to KnowledgeBase possible to maximise visibility of research content and promote research impact, e.g. OAI-PMH, HighWire data, sitemap support, etc and participation in (inter)national initiatives, e.g. Open Mirror, CORE, etc. • Development of plugin for Strathprints using Scopus API (within existing SciVal subscription) thus obviating most "data in" issues for the archiving of research content. • Opportunity to engage in a detailed programme of digital preservation work within an IR exemplifying true "institutional content", i.e. Strathprints. • Potential opportunity to deploy digital preservation toolkits/plugins within Strathprints. • Scope for development of Strathprints as a standalone service (to ePrints V3.3+), with potential of improving UX, user engagement and exposure of misc. archival and grey content. • Potential for improvements to data re-use across both IR platforms in a variety of institutional systems.	• Repository priorities capable of being set locally to better serve departmental needs. • Good surface web search agent indexing penetration of both KnowledgeBase and Strathprints. • Content metadata tools in PURE (e.g. Scopus, arXiv, etc.) enable simple staff ingest and benefit. • PURE cron job benefits (e.g. automatic assignation of ASJC classifications) thus supporting metadata generation. • Strathprints continues to expose some Strathclyde research content (assuming "data/content in" issues ameliorated, thus providing additional user discovery mechanism).	• Potential lack of efficacy in IR ecosystem as content (often overlapping) is managed across disparate IRs by (potentially) different teams (e.g. overlapping programmes of work, divergence in standards / quality management, negative consequences of staff spread perhaps leading to additional staff, low knowledge exchange, etc.). • Difficult to understand true impact and use of research content spread across multiple IRs. • Data used to populate T4 publication pages would need to be sourced from PURE, requiring development work from Web Team. • Not all content exposed equally, as content in Strathprints more visible, perhaps incurring displeasure of academics.
Threats	• Once productive collaborative relationships across ISD and RKES in the area of IR issues and OA degrade with the retirement of the connector. • Limiting Strathprints content to Strathclyde only content could suppress the impact of Strathclyde researchers more generally, by only promoting those works generated while at the institution. This has potential to damage research impact objectives in the Strategic Plan (2011-2015). • Compliancy reporting (e.g. RCUK, HEFCE, EU, etc.) could be difficult from PURE unless better support is demonstrated by Atira/SciVal. • Long lead times for Atira/SciVal's implementation of PUREPortal technical improvements. • Atira/SciVal development of the PUREPortal deprioritised in favour of larger CRIS needs.		• Lack of IR interoperability (KnowledgeBase) leading to poor participation in global IR network and low usage of Strathclyde research in text mining and/or discovery via harvesting search providers. • Institutional disenfranchisement from national and international IR initiatives owing to deficit in KnowledgeBase (PURE) functionality, e.g. Open Mirror. • Current lack of COUNTER compliant statistical tools for the KnowledgeBase, e.g. IRUS-UK thus making "real world" measurement of contemporaneous research impact difficult.

3.4 IR scenario #4: "connector lite"

IR scenario #4, or the "connector lite" scenario, refers to an IR configuration that would seek to incorporate the best aspects of both scenarios #1 and #3 by demonstrating only a partial connection between Strathprints and PURE, thus enabling the selective pushing of research content from PURE to Strathprints. The principal benefit of this approach is the relative separation of Strathprints and the KnowledgeBase, enabling the development of "institutional content" within Strathprints while allowing a stronger CRIS focus on contemporaneous research content. It would also obviate some of the technical issues surrounding the population of Strathprints associated with scenario #3 but would require bespoke development of PURE and resources this entails.

Table 4: IR scenario #4 ("connector lite") SWOT matrix.

Internal environment / External environment		Strengths	Weaknesses
		• High visibility of Strathclyde research content via use of Strathprints technical protocols in harvesting services, social reference services, academic search engines (e.g. GS) and SUPrimo search, thus improving potential of research impact. • Collection of "institutional content" better fulfilling "traditional" IR remit of institutional memory for Strathprints.	• Technical overhead of maintaining two systems with "part-time" connection remains similar to scenario #1; overhead for GM, systems team, infrastructure, RKES. • Connector configuration prone to failure and requires maintenance.
Opportunities	• Opportunity to engage in a detailed programme of digital preservation work within an IR exemplifying true "institutional content". • Potential opportunity to deploy digital preservation toolkits/plugins within Strathprints. • Improvements to KnowledgeBase possible to maximise visibility of research content and promote research impact, e.g. OAI-PMH, HighWire data, sitemap support. • PURE refocus on contemporaneous content. • Scope for development of Strathprints as a standalone service (to ePrints V3.3+), with potential of improving UX, user engagement and exposure of misc. archival and grey content. • Potential for improvements to data re-use across both IR platforms in a variety of institutional systems.	• "Part time" parallel operation of Strathprints and KnowledgeBase provides multiple discovery paths to most Strathclyde research content. • Good surface web search agent indexing penetration of both KnowledgeBase and Strathprints. • Strathprints compliance with OAI-PMH will enable participation in – and extra visibility from - Open Mirror. • Content metadata tools in PURE (e.g. Scopus, arXiv, etc.) enable simple staff ingest and benefit population of Strathprints.	• Bespoke development of PURE required to support "connector lite" workflow and associated costs. • Difficult to understand true impact and use of research content spread across multiple IRs. • Data used to populate T4 publication pages would need to be sourced from PURE, requiring development work from Web Team. • Not all content exposed equally, as content in Strathprints more visible, perhaps incurring displeasure of academics.
Threats	• Any development to alter the validation workflow to support "connector lite" could impose additional maintenance requirements and/or technical risks. • Limiting Strathprints content to Strathclyde only content could suppress the impact of Strathclyde researchers more generally, by only promoting those works generated while at the institution. This has potential to damage research impact objectives in the Strategic Plan (2011-2015). • Compliancy reporting (e.g. RCUK, HEFCE, EU, etc.) could be difficult from PURE unless better support is demonstrated by Atira/SciVal. • Long lead times for Atira/SciVal's implementation of PUREPortal technical improvements. • Atira/SciVal development of the PUREPortal deprioritised in favour of larger CRIS needs. • Lack of COUNTER compliant statistical tools for the KnowledgeBase, e.g. IRUS-UK.	• Demarcation of Strathprints and PURE, with clear IR and CRIS remits and less user confusion; distinct services to promote to users. • PURE cron job benefits (e.g. automatic assignation of ASJC classifications) thus supporting metadata generation. • Wide staff engagement with PURE, thus contributing to CRIS business intelligence objectives but also Strathprints population. • Maintenance of productive collaborative relationship across ISD and RKES.	• Parallel IR operation complicates compliance with prevailing technical standards and protocols, as standards need to be reflected across two IRs (e.g. RIOXX/V4OA, OpenAIRE, etc.), one of which is driven by commercial decision. • Potential lack of efficacy as content is managed separately across disparate IRs by (potentially) different teams (e.g. divergence in standards / quality management, negative consequences of staff spread, etc.). • Many improvements to Strathprints remain wedded to PURE / KnowledgeBase, and vice versa.

4. References

[1] University of Southampton, “EPrints - Digital Repository Software,” 2013. [Online]. Available: http://www.eprints.org/. [Accessed: 07-Dec-2013].

[2] Elsevier, “Pure | SciVal,” 2013. [Online]. Available: http://info.scival.com/pure. [Accessed: 07-Dec-2013].

[3] WebDAV Community, “WebDAV Resources,” 2010. [Online]. Available: http://www.webdav.org/. [Accessed: 02-Dec-2013].

[4] University of Strathclyde, “KnowledgeBase,” 2013. [Online]. Available: http://pure.strath.ac.uk/portal/. [Accessed: 02-Dec-2013].

[5] A. Dorward and P. de Castro, “UK RepositoryNet+ Round Table Presentation on Lessons Learnt,” presented at the Repository Fringe 2013, University of Edinburgh, 2013.

[6] L. Wenaas, N. Karlstrøm, and T. Vatnan, “From a national CRIS along the road to Green Open Access – and back again : building infrastructure from CRIStin to institutional repositories in Norway,” in *E-Infrastructures for Research and Innovation: Linking Information Systems to Improve Scientific Knowledge Production: Proceedings of the 11th International Conference on Current Research Information Systems*, Prague, Czech Republic, 2012, pp. 289–294.

[7] J. Schirrwagen, P. Manghi, N. Manola, L. Bolikowski, N. Rettberg, and B. Schmidt, “Data Curation in the OpenAIRE Scholarly Communication Infrastructure,” *Information Standards Quarterly*, vol. 25, no. 3, p. 13, 2013.

[8] N. Sheppard, “Learning How to Play Nicely: Repositories and CRIS,” *Ariadne: Web Magazine for Information Professionals*, 2010. [Online]. Available: http://www.ariadne.ac.uk/issue64/wrn-repos-2010-05-rpt. [Accessed: 02-Dec-2013].

[9] L. Carr, “EPrints: a hybrid CRIS/repository,” presented at the Workshop on CRIS, CERIF and Institutional Repositories, 2010.

[10] R. Crow, “The Case for Institutional Repositories: A SPARC,” *Position paper*, 2006. [Online]. Available: http://scholarship.utm.edu/20/. [Accessed: 02-Dec-2013].

[11] K. Jeffery and A. Asserson, “Institutional Repositories and Current Research Information Systems,” *New Review of Information Networking*, vol. 14, no. 2, pp. 71–83, 2009.

[12] S. Harnad, “The Implementation of the Berlin Declaration on Open Access: Report on the Berlin 3 Meeting Held 28 February - 1 March 2005, Southampton, UK,” *D-Lib Magazine*, vol. 11, no. 03, Mar. 2005.

[13] K. G. Jeffery and A. Asserson, “Supporting the Research Process with a CRIS,” 2006.

[14] euroCRIS, “CRIS concept and CRIS benefits,” 2010. [Online]. Available: http://www.eurocris.org/Index.php?page=concepts_benefits&t=1. [Accessed: 11-Nov-2013].

[15] A. Clements, “Using CERIF-XML to integrate heterogeneous research information from several institutions into a single portal: JISC final report,” 2010. [Online]. Available: http://goo.gl/FzC4di.

[16] “CERIF specification,” 26-Oct-2013. [Online]. Available: http://www.eurocris.org/Index.php?page=CERIF-1.5&t=1. [Accessed: 26-Jun-2013].

[17] JISC, “Readiness for REF,” 2013. [Online]. Available: http://www.jisc.ac.uk/whatwedo/programmes/inf11/sue2/r4r. [Accessed: 02-Dec-2013].

[18] University of Southampton, “EPrints - Are you ready for REF 2014?,” 2013. [Online]. Available: http://www.eprints.org/ref2014/. [Accessed: 11-Dec-2013].

[19] University of Glasgow, “Enlighten,” 2013. [Online]. Available: http://eprints.gla.ac.uk/. [Accessed: 02-Dec-2013].

[20] W. Nixon, “Enrich: improving integration between an institutional repository and a CRIS at the University of Glasgow,” in *Connecting Science with Society: The Role of Research Information in a Knowledge Based Society*, M. Stempfhuber and N. Thidemann, Eds. Aalborg, Denmark: Toptryk Grafisk ApS, 2010, pp. 55–64.

[21] W. Nixon, S. Ashworth, and V. McCutcheon, “Enlighten: research and APC funding workflows at the University of Glasgow,” *Insights: the UKSG journal*, vol. 26, no. 2, pp. 159–167, Jul. 2013.

[22] B. Alroe and H. Steen, “CRIS + Repositories, all in one?,” presented at the 9th euroCRIS Strategic Seminar, Brussels, 2011.

[23] University of Dundee, “Discovery : the University’s institutional repository,” 2012. [Online]. Available: http://www.dundee.ac.uk/library/research/discovery/. [Accessed: 02-Dec-2013].

[24] University of Edinburgh, “Edinburgh Research Explorer,” 2013. [Online]. Available: http://www.research.ed.ac.uk/portal/. [Accessed: 02-Dec-2013].

[25] University of Edinburgh, “Edinburgh Research Archive,” 2013. [Online]. Available: https://www.era.lib.ed.ac.uk/. [Accessed: 02-Dec-2013].

[26] A. Swan and L. Carr, “Institutions, Their Repositories and the Web,” *Serials Review*, vol. 34, no. 1, pp. 31–35, Mar. 2008.

[27] A. Swan, “Open Access institutional repositories: a briefing paper,” *Enabling Open Scholarship (EOS)*, 2009. [Online]. Available: http://www.openscholarship.org/upload/docs/application/pdf/2009-09/open_access_institutional_repositories.pdf. [Accessed: 02-Dec-2013].

[28] R. Kennison, S. Shreeves, and S. Harnad, “Point & Counterpoint: The Purpose of Institutional Repositories: Green OA or Beyond?,” *Journal of Librarianship and Scholarly Communication*, vol. 1, no. 4, Sep. 2013.

[29] S. Rumsey, “The purpose of institutional repositories in UK higher education : a repository manager’s view,” *International Journal of Information Management*, vol. 26, no. 3, pp. 181–186, Jun. 2006.

[30] C. S. Burns, A. Lana, and J. M. Budd, “Institutional Repositories: Exploration of Costs and Value,” *D-Lib Magazine*, vol. 19, no. 1/2, Jan. 2013.

[31] Confederation of Open Access Repositories, “The case for interoperability for open access repositories,” 2011. [Online]. Available: http://www.coar-repositories.org/files/A-Case-for-Interoperability-Final-Version.pdf. [Accessed: 02-Dec-2013].

[32] University of Southampton, “EdShare,” 2013. [Online]. Available: http://www.edshare.soton.ac.uk/. [Accessed: 02-Dec-2013].

[33] MIMAS and JISC, “JORUM,” 2013. [Online]. Available: http://open.jorum.ac.uk/xmlui/. [Accessed: 02-Dec-2013].

[34] University of Southampton, “ROAR: ‘Learning and Teaching Objects’ repository type matches,” 2013. [Online]. Available: http://goo.gl/0721WV. [Accessed: 02-Dec-2013].

[35] University of Strathclyde, “Our Strategic Plan - University of Strathclyde,” 2011. [Online]. Available: http://www.strath.ac.uk/strategicplan/. [Accessed: 31-Jul-2012].

[36] University of Strathclyde, “Strategy and Policy Directorate,” 2013. [Online]. Available: http://www.strath.ac.uk/ps/strategyandpolicy/. [Accessed: 10-Apr-2013].
[37] A. M. Tmava and D. G. Alemneh, “Enhancing Content Visibility in Institutional Repositories: Overview of Factors that Affect Digital Resources Discoverability,” *iConference, 2013, Fort Worth, Texas, United States*, Feb. 2013.
[38] S. Harnad, “Maximizing Research Impact Through Institutional and National Open-Access Self-Archiving Mandates,” presented at the CRIS2006. Current Research Information Systems: Open Access Institutional Repositories, 2006.
[39] S. Hitchcock, “The effect of open access and downloads (’hits’) on citation impact: a bibliography of studies,” 2013. [Online]. Available: http://opcit.eprints.org/oacitation-biblio.html. [Accessed: 02-Dec-2013].
[40] M. Norris, C. Oppenheim, and F. Rowland, “The citation advantage of open-access articles,” *Journal of the American Society for Information Science and Technology*, vol. 59, no. 12, pp. 1963–1972, 2008.
[41] Y. Gargouri, C. Hajjem, V. Larivière, Y. Gingras, L. Carr, T. Brody, and S. Harnad, “Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research,” *PLoS ONE*, vol. 5, no. 10, p. e13636, Oct. 2010.
[42] G. Eysenbach, “Citation Advantage of Open Access Articles,” *PLoS Biol*, vol. 4, no. 5, p. e157, May 2006.
[43] C. Hajjem, S. Harnad, and Y. Gingras, “Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact,” arXiv e-print cs/0606079, Jun. 2006.
[44] S. Lawrence, “Free online availability substantially increases a paper’s impact,” *Nature*, vol. 411, no. 6837, p. 521, 2001.
[45] A. Gentil-Beccot, S. Mele, and T. Brooks, “Citing and Reading Behaviours in High-Energy Physics. How a Community Stopped Worrying about Journals and Learned to Love Repositories,” arXiv e-print 0906.5418, Jun. 2009.
[46] H. F. Moed, “The effect of ‘open access’ on citation impact: An analysis of ArXiv’s condensed matter section,” *Journal of the American Society for Information Science and Technology*, vol. 58, no. 13, pp. 2047–2054, 2007.
[47] V. Aman, “The potential of preprints to accelerate scholarly communication - A bibliometric analysis based on selected journals,” arXiv e-print 1306.4856, Jun. 2013.
[48] University of Strathclyde, “Strathprints: IRStats,” 2013. [Online]. Available: http://strathprints.strath.ac.uk/cgi/irstats.cgi. [Accessed: 02-Dec-2013].
[49] MIMAS, Cranfield University, Evidence Base, and JISC, “IRUS-UK,” 2013. [Online]. Available: http://www.irus.mimas.ac.uk/. [Accessed: 05-Dec-2013].
[50] P. Needham and G. Stone, “IRUS-UK: making scholarly statistics count in UK repositories,” *Insights: the UKSG journal*, vol. 25, no. 3, pp. 262–266, Nov. 2012.
[51] G. Stone, “Huddersfield Open Access Publishing,” *Information Services and Use*, vol. 31, no. 3, pp. 215–223, Jan. 2011.
[52] ProjectCOUNTER, “COUNTER - Counting Online Usage of Networked Electronic Resources | Home,” 2013. [Online]. Available: http://www.projectcounter.org/. [Accessed: 05-Dec-2013].
[53] S. Hitchcock, “EPrints ‘Request eprint’ button,” 2006. [Online]. Available: http://www.eprints.org/news/features/request_button.php. [Accessed: 05-Dec-2013].
[54] C. Lagoze, H. V. de Sompel, M. Nelson, and S. Warner, “Open Archives Initiative Protocol for Metadata Harvesting,” 2008. [Online]. Available: http://www.openarchives.org/pmh/. [Accessed: 02-Dec-2013].
[55] Knowledge Media Institute, “CORE,” 2013. [Online]. Available: http://core.kmi.open.ac.uk/search. [Accessed: 02-Dec-2013].
[56] Bielefeld University, “BASE (Bielefeld Academic Search Engine): Basic Search,” 2013. [Online]. Available: http://www.base-search.net/. [Accessed: 02-Dec-2013].
[57] MIMAS and JISC RepositoryNet, “Institutional Repository Search,” 2013. [Online]. Available: http://irs.mimas.ac.uk/demonstrator/. [Accessed: 02-Dec-2013].
[58] OCLC, “OAIster,” 2013. [Online]. Available: http://www.oclc.org/oaister.en.html. [Accessed: 02-Dec-2013].
[59] P. Knoth and Z. Zdrahal, “CORE: Aggregation use cases for open access,” in *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, 2013, pp. 441–442.
[60] Knowledge Media Institute, “CORE API,” 2013. [Online]. Available: http://core.kmi.open.ac.uk/intro/api. [Accessed: 04-Dec-2013].
[61] P. Knoth, “From Open Access metadata to Open Access content: two principles for increased visibility of Open Access content,” presented at the Open Repositories 2013 (OR 2013), Charlottetown, Prince Edward Island, Canada, 2013.
[62] OCLC, “WorldCat.org: The World’s Largest Library Catalog,” 2013. [Online]. Available: http://www.worldcat.org/. [Accessed: 02-Dec-2013].
[63] “Registry of Open Access Repositories,” 2013. [Online]. Available: http://roar.eprints.org/. [Accessed: 29-Oct-2013].
[64] JISC and University of Nottingham, “OpenDOAR: Directory of Open Access Repositories,” 2011. [Online]. Available: http://www.opendoar.org/. [Accessed: 02-Dec-2013].
[65] University of Strathclyde, “SUPrimo,” 2013. [Online]. Available: http://www.strath.ac.uk/library/. [Accessed: 02-Dec-2013].
[66] N. Jacobs, “What could you do with an Open Mirror?,” *Jisc*, 2013. [Online]. Available: http://www.jisc.ac.uk/blog/what-could-you-do-with-an-open-mirror-22-oct-2013. [Accessed: 02-Dec-2013].
[67] J. Toon, “Requirements for implementation of OAI-PMH in PURE. Drafted on behalf of the PURE UK User Group - OA and Repositories Sub Group,” 2013. [Online]. Available: http://goo.gl/GMJ0e9. [Accessed: 02-Dec-2013].
[68] OpenAIRE Consortium, “OpenAIRE,” 2013. [Online]. Available: http://www.openaire.eu/en/component/openaire/browse/default/390. [Accessed: 07-Dec-2013].
[69] RCUK, JISC, and UKOLN, “RIOXX: Home of the RIOXX application profile and guidelines,” 2013. [Online]. Available: http://rioxx.net/. [Accessed: 02-Dec-2013].
[70] JISC and UKOLN, “V4OA: Vocabularies for Open Access,” 2013. [Online]. Available: http://v4oa.net/. [Accessed: 02-Dec-2013].
[71] T. Brody, “RIOXX Application Profile Export - Bazaar Store,” 2013. [Online]. Available: http://bazaar.eprints.org/316/. [Accessed: 02-Dec-2013].
[72] OpenAIRE Consortium, “OpenAIRE Guidelines 2.0: Guidelines for content providers of the OpenAIRE information space,” 2013. [Online]. Available: http://goo.gl/gW7hTy. [Accessed: 02-Dec-2013].
[73] HEFCE, “Consultation on open access in the post-2014 Research Excellence Framework,” 2013. [Online]. Available: http://www.hefce.ac.uk/pubs/year/2013/201316/. [Accessed: 02-Dec-2013].
[74] Google, “Google Scholar help: inclusion guidelines for webmasters - indexing,” 2013. [Online]. Available: http://www.google.com/intl/en/scholar/inclusion.html#indexing. [Accessed: 02-Dec-2013].
[75] K. Arlitsch and P. S. O’Brien, “Invisible institutional repositories: Addressing the low indexing ratios of IRs in Google Scholar,” *Library Hi Tech*, vol. 30, no. 1, pp. 60–81, 2012.

[76] Google, “Google Scholar help: inclusion guidelines for Webmasters - overview,” 2013. [Online]. Available: http://www.google.com/intl/en/scholar/inclusion.html#overview. [Accessed: 02-Dec-2013].
[77] J. Toon and G. Macgregor, “PurePortal: Research Explorer HighWire tags (Email correspondence),” 25-Nov-2013.
[78] Microsoft, “Microsoft Academic Search,” 2013. [Online]. Available: http://academic.research.microsoft.com/. [Accessed: 02-Dec-2013].
[79] P. Jacsó, “The pros and cons of Microsoft Academic Search from a bibliometric perspective,” *Online Information Review*, vol. 35, no. 6, pp. 983–997, Nov. 2011.
[80] J. L. Ortega and I. F. Aguillo, “Microsoft Academic Search and Google Scholar citations: a comparative analysis of author profiles,” *Journal of the American Society for Information Science and Technology*, In press.
[81] Microsoft, “sitemaps.org,” 2008. [Online]. Available: http://www.sitemaps.org/. [Accessed: 02-Dec-2013].
[82] Google, “Google Webmaster Guidelines,” 2013. [Online]. Available: https://support.google.com/webmasters/answer/35769?hl=en. [Accessed: 02-Dec-2013].
[83] Microsoft, “Bing Webmaster Guidelines,” 2013. [Online]. Available: http://www.bing.com/webmaster/help/webmaster-guidelines-30fba23a. [Accessed: 02-Dec-2013].
[84] U. Schonfeld and N. Shivakumar, “Sitemaps: above and beyond the crawl of duty,” in *Proceedings of the 18th international conference on World wide web*, New York, NY, USA, 2009, pp. 991–1000.
[85] R. Cyganiak, H. Stenzhorn, R. Delbru, S. Decker, and G. Tummarello, “Semantic Sitemaps: Efficient and Flexible Access to Datasets on the Semantic Web,” in *The Semantic Web: Research and Applications*, S. Bechhofer, M. Hauswirth, J. Hoffmann, and M. Koubarakis, Eds. Springer Berlin Heidelberg, 2008, pp. 690–704.
[86] R. Cyganiak, R. Delbru, and G. Tummarello, “Semantic Web Crawling: A Sitemap Extension,” 2007. [Online]. Available: http://sw.deri.org/2007/07/sitemapextension/. [Accessed: 02-Dec-2013].
[87] W3C, “Linked Data,” *W3C*, 2013. [Online]. Available: http://www.w3.org/standards/semanticweb/data. [Accessed: 07-Dec-2013].
[88] D. Goodwin, “Google Fails to Gain Search Market Share, Bing Steals From Yahoo,” *Search Engine Watch*, 14-Nov-2013. [Online]. Available: http://searchenginewatch.com/article/2307115/Google-Fails-to-Gain-Search-Market-Share-Bing-Steals-From-Yahoo. [Accessed: 02-Dec-2013].
[89] “dev:exposing_metadata [Zotero Documentation].” [Online]. Available: http://www.zotero.org/support/dev/exposing_metadata. [Accessed: 02-Dec-2013].
[90] Mendeley, “Mendeley Web Importer - Information for Publishers,” 2013. [Online]. Available: http://www.mendeley.com/import/information-for-publishers/. [Accessed: 02-Dec-2013].
[91] Publishers International Linking Association, “crossref.org,” 2013. [Online]. Available: http://www.crossref.org/. [Accessed: 02-Dec-2013].
[92] D. MacMillan, “Mendeley: Teaching Scholarly Communication and Collaboration through Social Networking,” in *2012 IATUL Proceedings*, Nanyang Technological University, Singapore, 2012.
[93] X. Li, M. Thelwall, and D. Giustini, “Validating online reference managers for scholarly impact measurement,” *Scientometrics*, vol. 91, no. 2, pp. 461–471, May 2012.
[94] Mendeley, “Free reference manager and PDF organizer | Mendeley,” 2013. [Online]. Available: http://www.mendeley.com/. [Accessed: 02-Dec-2013].
[95] Roy Rosenzweig Center for History and New Media, “Zotero,” 2013. [Online]. Available: http://www.zotero.org/. [Accessed: 02-Dec-2013].
[96] D. Dobbs, “When the Rebel Alliance Sells Out,” *The New Yorker Blogs*, 12-Apr-2013. .
[97] V. Henning and J. Reichelt, “Mendeley - A Last.fm For Research?,” in *IEEE Fourth International Conference on eScience, 2008. eScience ’08*, 2008, pp. 327–328.
[98] Mendeley, “How does the automatic document details ...,” 2013. [Online]. Available: http://support.mendeley.com/customer/portal/articles/227883-how-does-the-automatic-document-details-extraction-work-. [Accessed: 02-Dec-2013].
[99] G. Macgregor, C. Keene, J. Toon, and G. Hall, “Collection policy query: exposing historical research outputs (Email correspondence to the UKCoRR list),” 12-Aug-2013.
[100] R. K. Johnson, “Institutional repositories: Partnering with faculty to enhance scholarly communication,” *D-Lib Magazine*, vol. 8, no. 11, 2002.
[101] S. Hitchcock, “Institutional digital repositories: What role do they have in curation?,” 29-Jun-2011. [Online]. Available: http://eprints.soton.ac.uk/273243/. [Accessed: 16-Dec-2013].
[102] University of Southampton, “EPrints Digital Preservation,” 2009. [Online]. Available: http://preservation.eprints.org/. [Accessed: 16-Dec-2013].
[103] University of Southampton, “KeepIt: Kultur, eCrystals, EdShare (and NECTAR) - Preserve It!,” 2010. [Online]. Available: http://preservation.eprints.org/keepit/. [Accessed: 16-Dec-2013].
[104] M. Pickton, D. Morris, S. Meece, S. Coles, and S. Hitchcock, “Preserving repository content: practical tools for repository managers,” *Journal of Digital Information*, vol. 12, no. 2, Apr. 2011.
[105] University of Southampton, “Preserv: Repository Preservation and Interoperability,” 2009. [Online]. Available: http://preserv.eprints.org/. [Accessed: 17-Dec-2013].
[106] S. Hitchcock, D. Tarrant, and L. Carr, “Towards repository preservation services: Final report from the JISC Preserv 2 project,” *Final Report to JISC*, 21-Aug-2012. [Online]. Available: http://repository.jisc.ac.uk/381/. [Accessed: 16-Dec-2013].
[107] D. Tarrant, “Preservation Toolkit,” *ePrints Bazaar Store*, 01-Aug-2011. [Online]. Available: http://bazaar.eprints.org/142/. [Accessed: 16-Dec-2013].
[108] D. Tarrant, “Droid,” 2012. [Online]. Available: http://bazaar.eprints.org/143/. [Accessed: 16-Dec-2013].
[109] D. Tarrant and T. Brody, “Preserv2/EPrints Preservation Plugins 1,” 2010. [Online]. Available: http://files.eprints.org/422/. [Accessed: 16-Dec-2013].
[110] University of Southampton, “Preservation Support,” 2010. [Online]. Available: http://wiki.eprints.org/w/Preservation_Support. [Accessed: 16-Dec-2013].
[111] R. Lowe and M. Pembury, “Arkivum A-Stor Storage Backend,” 28-Jun-2013. [Online]. Available: http://bazaar.eprints.org/313/. [Accessed: 16-Dec-2013].
[112] JISC and UKOLN, “SWORD,” 2013. [Online]. Available: http://swordapp.org/. [Accessed: 17-Dec-2013].
[113] U. Isikdag, S. Zlatanova, and TU Delft: OTB Research Institute, “A SWOT analysis on the implementation of Building Information Models within the geospatial environment.” Taylor & Francis Group, 31-Dec-2009.

[114] J. Tan, Ed., *Developments in Healthcare Information Systems and Technologies: Models and Methods*. IGI Global, 2010.
[115] D.-H. Kim, B.-J. Ahn, J.-H. Kim, and J.-J. Kim, “The Strategic Approach Using SWOT Analysis to Develop an Intelligent Program Management Information System (iPMIS) for Urban Renewal Projects,” in *Fourth International Conference on Computer Sciences and Convergence Information Technology, 2009. ICCIT ’09*, 2009, pp. 320–324.

5. Appendix A

Table 5: KnowledgeBase top 50 pageviews for the period 09/12/2013-16/12/2013, as counted by the institutional Google Analytics account.

KnowledgeBase path	Pageviews	Unique Pageviews	Avg. Time on Page	Bounce Rate	% Exit
/persons/search.html?advanced=true	162	110	9.97	33.33%	8.64%
/organisations/index.html	112	91	19.69	0.00%	9.82%
/publications/search.html	101	82	15.99	0.00%	9.90%
/	98	73	56.89	7.69%	8.16%
/projects/search.html	91	73	33.23	0.00%	10.99%
/impacts/search.html?advanced=true	37	29	40.14	0.00%	21.62%
/persons/jules-simo(0728db21-1dbf-4b99-8436-7780628ec38c).html	36	16	109.38	28.57%	33.33%
/activities/search.html?advanced=true	34	28	13.71	0.00%	17.65%
/journals/search.html	32	30	13.85	0.00%	18.75%
/organisations/strathclyde-institute-of-pharmacy-and-biomedical-sciences(a2d0d1ee-5c6b-47d4-9a69-9559a744e3be).html	26	17	13.53	33.33%	42.31%
/equipment/search.html	23	18	4.90	0.00%	13.04%
/organisations/electronic-and-electrical-engineering(b97a184c-1352-4101-ae9b-2c48bcb874fc).html	22	17	16.88	0.00%	22.73%
/persons/francis-quail(ee14d28b-95f7-4c7f-989c-6fbb8dca16e8).html	22	16	64.80	50.00%	31.82%
/persons/lie-xu(e3156411-5f45-47bf-aeb6-755e4faf2da1)/publications.html	22	8	17.68	0.00%	0.00%
/persons/jennifer-hastie(dbc5dc95-2b4b-4f95-81f1-2bd207ee1885).html	21	10	88.24	66.67%	19.05%
/persons/keith-mathieson(2a47ac65-b002-4f61-b295-b1965cacdd29).html	18	15	52.80	45.45%	44.44%
/persons/andrew-colin-robertson(2bdb5b28-2929-4082-b050-1ca63be64cc9)/publications.html	16	7	65.75	33.33%	25.00%
/persons/william-leithead(d08359ca-2ff5-4b93-862b-7686d244964c).html	16	9	42.27	57.14%	31.25%
/organisations/mechanical-and-aerospace-engineering(3c556187-f468-4880-87b9-1b838544878b).html	15	12	62.00	0.00%	20.00%
/persons/graeme-burt(4faf261c-5aa3-44e1-b7bc-2f4ae9d1c2be).html	15	10	159.85	33.33%	13.33%
/organisations/electronic-and-electrical-engineering(b97a184c-1352-4101-ae9b-2c48bcb874fc)/persons.html?filter=current	14	3	20.50	0.00%	0.00%
/organisations/index.html?faculty=Engineering	14	13	9.86	0.00%	0.00%

KnowledgeBase path	Pageviews	Unique Pageviews	Avg. Time on Page	Bounce Rate	% Exit
/persons/duncan-graham(fffefd19-16b4-468b-ad7d-3faf32af11ff).html	14	11	92.00	11.11%	7.14%
/persons/bernard-harris(a6576388-0649-491f-9df0-bfe5bd40efd1).html	13	6	195.00	0.00%	7.69%
/persons/duncan-graham(fffefd19-16b4-468b-ad7d-3faf32af11ff)/activities.html	13	5	17.20	0.00%	23.08%
/publications/globalization-and-global-politics(8117079a-fd33-49b7-a50d-66259698d300)/export.html	13	9	87.75	88.89%	69.23%
/organisations/index.html?faculty=Science	12	11	7.58	0.00%	0.00%
/organisations/pure-and-applied-chemistry(b17c127f-2782-4bc7-a810-32be7a6ea91b).html	12	7	78.78	0.00%	25.00%
/persons/campbell-booth(8928524b-e866-4980-9706-74c6dd6e8097).html	12	11	30.00	50.00%	33.33%
/persons/graham-ault(0d537d36-ff77-4f55-af76-3ab20c06e4bf).html	12	9	14.57	60.00%	41.67%
/persons/joseph-clarke(af7b1779-f0f5-4b8f-af27-0c6c973dbb82).html	12	8	38.38	42.86%	33.33%
/persons/lie-xu(e3156411-5f45-47bf-aeb6-755e4faf2da1).html	12	8	31.10	14.29%	16.67%
/persons/simon-adams(88633fc4-7822-4e79-aa7e-478d1b6cdb3a).html	12	4	41.90	0.00%	16.67%
/publications/strategies-to-transfer-research-findings-to-professionals-in-maternity-services(2994bda8-1ec5-4d49-9027-92dcfa465e4d).html	12	5	263.36	0.00%	8.33%
/activities/international-journal-of-production-planning-and-control(e64cae18-696b-42e3-ad2a-a79d5824469d).html	11	9	151.25	66.67%	63.64%
/organisations/electronic-and-electrical-engineering(b97a184c-1352-4101-ae9b-2c48bcb874fc)/persons.html	11	8	64.45	0.00%	0.00%
/organisations/electronic-and-electrical-engineering(b97a184c-1352-4101-ae9b-2c48bcb874fc)/persons.html?page=2&filter=current	11	2	20.10	0.00%	9.09%
/persons/francis-quail(ee14d28b-95f7-4c7f-989c-6fbb8dca16e8)/projects.html	11	6	13.73	0.00%	0.00%
/persons/patrick-norman(9d9e7aa9-f5bb-4895-b96a-8557c0832fd4)/publications.html	11	4	29.25	0.00%	27.27%
/persons/paul-gerard-tuohy(61e6fea1-3589-4158-8d32-b8fb6c8c47a5).html	11	8	56.00	57.14%	63.64%
/projects/eu-fp7--seahorse(4bdfe218-f5e7-4ea9-8c22-99201ab3b1db).html	11	3	33.00	0.00%	9.09%
/organisations/biomedical-engineering(6e353bdf-26a7-4209-b042-f2a0a80491f2).html	10	6	18.22	0.00%	10.00%
/persons/bernard-harris(a6576388-0649-491f-9df0-bfe5bd40efd1)/publications.html	10	7	11.63	0.00%	20.00%
/persons/duncan-graham(fffefd19-16b4-468b-ad7d-3faf32af11ff)/projects.html	10	2	28.80	0.00%	0.00%
/persons/francis-quail(ee14d28b-95f7-4c7f-989c-6fbb8dca16e8)/supervision.html	10	9	51.17	100.00%	40.00%
/persons/iraklis-lazakis(d3612df8-5db8-49dd-9399-88f52bff61f4).html	10	6	31.33	40.00%	40.00%
/persons/jules-simo(0728db21-1dbf-4b99-8436-7780628ec38c)/publications.html	10	8	17.00	0.00%	20.00%
/persons/lina-stankovic(00700796-ea0e-450b-9185-eb3f2b2aa267)/projects.html	10	2	22.30	0.00%	0.00%

KnowledgeBase path	Pageviews	Unique Pageviews	Avg. Time on Page	Bounce Rate	% Exit
/persons/mahdi-khorasanchi(6d2944ee-14c7-45c8-80e0-d302fc3a27d2).html	10	3	7.78	0.00%	10.00%
/persons/william-mcgeown(df653b73-b372-43f9-a292-dbf00c638a96).html	10	3	49.75	0.00%	20.00%
	9099	6980	45.59	55.09%	29.68%

Table 6: Strathprints top 50 pageviews for the period 09/12/2013-16/12/2013, as counted by the institutional Google Analytics account.

Strathprints path	Pageviews	Unique Pageviews	Avg. Time on Page	Bounce Rate	% Exit
/cgi/	926	716	164.33	52.17%	22.79%
/view/	569	430	273.82	67.32%	34.45%
/	183	139	146.00	36.84%	37.70%
/downloads/	79	69	296.63	0.00%	49.37%
/42361/	70	51	188.85	70.59%	71.43%
/9593/	61	44	24.53	75.00%	68.85%
/4914/	36	1	57.91	0.00%	2.78%
/7361/	36	31	2828.80	83.87%	86.11%
/13436/	35	25	2853.10	70.83%	71.43%
/openaccess.html	31	28	69.00	100.00%	41.94%
/2603/	29	21	110.78	71.43%	68.97%
/43871/	27	20	342.57	78.95%	74.07%
/13304/	26	17	658.36	47.06%	57.69%
/26994/	24	23	279.50	91.30%	91.67%
/29448/	24	15	734.44	73.33%	62.50%
/35374/	23	16	103.14	66.67%	69.57%
/5213/	22	17	2036.33	82.35%	72.73%
/38070/	20	15	353.83	66.67%	70.00%
/15582/	19	15	303.43	66.67%	63.16%
/26496/	19	11	8548.38	45.45%	57.89%
/7405/	18	15	368.00	80.00%	83.33%
/26548/	17	13	4161.00	76.92%	76.47%
/7872/	17	15	22.50	86.67%	88.24%
/1806/	16	10	47.00	70.00%	56.25%
/7996/	16	15	155.00	86.67%	87.50%

/8833/	16	13	518.67	84.62%	81.25%
/9673/	16	10	128.50	50.00%	50.00%
/15474/	15	13	51.50	92.31%	86.67%
/37619/	14	13	501.00	92.31%	92.86%
/27825/	13	8	276.80	42.86%	61.54%
/8908/	13	7	195.43	57.14%	46.15%
/15446/	12	12	0.00	100.00%	100.00%
/16115/	12	9	654.67	66.67%	75.00%
/2308/	12	10	14.00	80.00%	75.00%
/33378/	12	12	0.00	100.00%	100.00%
/40194/	12	8	1455.00	75.00%	66.67%
/4021/	12	11	28.00	81.82%	83.33%
/6502/	12	12	0.00	100.00%	100.00%
/9582/	12	8	64.00	62.50%	66.67%
/9666/	12	5	23.00	75.00%	41.67%
/16367/	11	11	0.00	100.00%	100.00%
/28548/	11	7	22.25	57.14%	63.64%
/4113/	11	5	1215.57	40.00%	36.36%
/42661/	11	8	83.67	62.50%	72.73%
/45173/	11	5	1258.83	20.00%	45.45%
/4977/	11	8	37.86	37.50%	36.36%
/7446/	11	9	39.50	88.89%	81.82%
/7706/	11	8	35.50	62.50%	63.64%
/835/	11	8	12.33	75.00%	72.73%
/18040/	10	8	2237.50	75.00%	80.00%
	8821	6868	446.42	73.13%	60.55%

Table 7: KnowledgeBase top 50 pageviews based on the city from which access originated for the period 09/12/2013-16/12/2013, as counted by the institutional Google Analytics account.

KnowledgeBase path	City	Pageviews	Unique Pageviews	Avg. Time on Page	Bounce Rate	% Exit
/persons/search.html?advanced=true	Glasgow	62	46	6.45	0.00%	9.68%
/projects/search.html	Glasgow	35	30	54.07	0.00%	17.14%
/	Glasgow	33	25	13.41	33.33%	12.12%
/publications/search.html	Glasgow	33	27	13.68	0.00%	6.06%
/organisations/index.html	Glasgow	32	27	33.63	0.00%	0.00%
/persons/jules-simo(0728db21-1dbf-4b99-8436-7780628ec38c).html	Paderborn	27	13	157.69	27.27%	40.74%
/persons/search.html?advanced=true	Hamilton	27	5	4.93	0.00%	0.00%
/activities/search.html?advanced=true	Glasgow	21	17	14.50	0.00%	14.29%
/persons/lie-xu(e3156411-5f45-47bf-aeb6-755e4faf2da1)/publications.html	Glasgow	17	5	17.35	0.00%	0.00%
/impacts/search.html?advanced=true	Glasgow	15	11	69.90	0.00%	33.33%
/persons/duncan-graham(fffefd19-16b4-468b-ad7d-3faf32af11ff).html	Glasgow	13	10	93.00	12.50%	7.69%
/persons/duncan-graham(fffefd19-16b4-468b-ad7d-3faf32af11ff)/activities.html	Glasgow	13	5	17.20	0.00%	23.08%
/persons/bernard-harris(a6576388-0649-491f-9df0-bfe5bd40efd1).html	Eastleigh	12	5	207.82	0.00%	8.33%
/organisations/electronic-and-electrical-engineering(b97a184c-1352-4101-ae9b-2c48bcb874fc)/persons.html?filter=current	Chisinau	11	1	18.18	0.00%	0.00%
/persons/jennifer-hastie(dbc5dc95-2b4b-4f95-81f1-2bd207ee1885).html	Dundee	11	2	125.60	0.00%	9.09%
/publications/strategies-to-transfer-research-findings-to-professionals-in-maternity-services(2994bda8-1ec5-4d49-9027-92dcfa465e4d).html	Larissa	11	4	265.70	0.00%	9.09%
/	Eastleigh	10	5	325.10	0.00%	0.00%
/equipment/search.html	Glasgow	10	8	3.29	0.00%	30.00%
/persons/duncan-graham(fffefd19-16b4-468b-ad7d-3faf32af11ff)/projects.html	Glasgow	10	2	28.80	0.00%	0.00%
/persons/mahdi-khorasanchi(6d2944ee-14c7-45c8-80e0-d302fc3a27d2).html	Glasgow	10	3	7.78	0.00%	10.00%
/persons/patrick-norman(9d9e7aa9-f5bb-4895-b96a-8557c0832fd4)/publications.html	Glasgow	10	3	30.71	0.00%	30.00%
/persons/william-leithead(d08359ca-2ff5-4b93-862b-7686d244964c).html	Glasgow	10	4	54.29	66.67%	30.00%
/persons/william-mcgeown(df653b73-b372-43f9-a292-dbf00c638a96)/publications.html	Glasgow	10	2	56.20	0.00%	0.00%
/journals/search.html	Glasgow	9	8	11.25	0.00%	11.11%
/organisations/electronic-and-electrical-engineering(b97a184c-1352-4101-ae9b-	Chisinau	9	1	21.00	0.00%	0.00%

KnowledgeBase path	City	Pageviews	Unique Pageviews	Avg. Time on Page	Bounce Rate	% Exit
2c48bcb874fc)/persons.html?page=2&filter=current						
/persons/andrew-colin-robertson(2bdb5b28-2929-4082-b050-1ca63be64cc9)/publications.html	East Kilbride	9	5	59.80	40.00%	44.44%
/persons/david-garcia(f24280b5-cd62-4386-8492-35720d9c0511).html	Glasgow	9	3	43.86	0.00%	22.22%
/persons/lie-xu(e3156411-5f45-47bf-aeb6-755e4faf2da1).html	Glasgow	9	5	35.25	0.00%	11.11%
/persons/lina-stankovic(00700796-ea0e-450b-9185-eb3f2b2aa267)/projects.html	Hamilton	9	1	22.89	0.00%	0.00%
/persons/robert-atkinson(a419c042-b11d-4edb-97cc-134d3ceade51)/supervision.html	Hamilton	9	1	3.44	0.00%	0.00%
/persons/william-mcgeown(df653b73-b372-43f9-a292-dbf00c638a96).html	Glasgow	9	2	54.14	0.00%	22.22%
/organisations/electronic-and-electrical-engineering(b97a184c-1352-4101-ae9b-2c48bcb874fc)/persons.html?filter=current&page=1	Chisinau	8	3	49.83	100.00%	25.00%
/organisations/electronic-and-electrical-engineering(b97a184c-1352-4101-ae9b-2c48bcb874fc)/persons.html?page=3&filter=current	Chisinau	8	3	24.60	100.00%	37.50%
/persons/francis-quail(ee14d28b-95f7-4c7f-989c-6fbb8dca16e8)/projects.html	Glasgow	8	3	12.88	0.00%	0.00%
/persons/gavin-halbert(2c983cbd-d5b5-43ec-b14a-6751fb2126bf)/supervision.html	Glasgow	8	1	12.14	0.00%	12.50%
/persons/jules-simo(0728db21-1dbf-4b99-8436-7780628ec38c)/publications.html	Paderborn	8	6	10.00	0.00%	25.00%
/persons/keith-mathieson(2a47ac65-b002-4f61-b295-b1965cacdd29).html	Glasgow	8	7	14.00	25.00%	37.50%
/persons/mahdi-khorasanchi(6d2944ee-14c7-45c8-80e0-d302fc3a27d2)/publications.html	Glasgow	8	4	128.00	100.00%	12.50%
/persons/search.html?advanced=true	Aalborg	8	7	6.63	0.00%	0.00%
/persons/william-mcgeown(df653b73-b372-43f9-a292-dbf00c638a96)/projects.html	Glasgow	8	1	125.38	0.00%	0.00%
/projects/ke-development-fund(416ad6b8-fd43-44aa-8a1e-03c0caeff7b5).html	London	8	1	8.88	0.00%	0.00%
/activities/directed-assembly-of-functional-nanomaterials-epsrc-grand-challenge-conference(d47d8bc6-452d-4181-b94a-bd0982315241).html	Akron	7	1	43.57	0.00%	0.00%
/persons/amir-siddiq(3c233b0e-c521-4a37-8ba0-4481f6f423db).html	Glasgow	7	3	25.80	0.00%	28.57%
/persons/andrew-colin-robertson(2bdb5b28-2929-4082-b050-1ca63be64cc9).html	Glasgow	7	2	100.00	0.00%	0.00%
/persons/andrew-colin-robertson(2bdb5b28-2929-4082-b050-1ca63be64cc9)/publications.html	Glasgow	7	2	70.00	0.00%	0.00%
/persons/arthur-mcivor(f91cecfe-44d0-4222-b7b5-d559cd912f44)/publications.html?page=2	Glasgow	7	1	15.00	0.00%	14.29%
/persons/duncan-graham(fffefd19-16b4-468b-ad7d-3faf32af11ff)/publications.html	Glasgow	7	4	98.40	0.00%	28.57%
/persons/francis-quail(ee14d28b-95f7-4c7f-989c-6fbb8dca16e8).html	Widnes	7	3	66.50	0.00%	14.29%
/persons/friedrich-philipp-seib(8da06a18-0175-4a1e-a89b-9ed4432696b3).html	Glasgow	7	4	8.00	33.33%	28.57%

KnowledgeBase path	City	Pageviews	Unique Pageviews	Avg. Time on Page	Bounce Rate	% Exit
/persons/jules-simo(0728db21-1dbf-4b99-8436-7780628ec38c).html	Glasgow	7	1	12.57	0.00%	0.00%
		9099	6980	45.59	55.09%	29.68%

Table 8: Strathprints top 50 pageviews based on the city from which access originated for the period 09/12/2013-16/12/2013, as counted by the institutional Google Analytics account.

Strathprints path	City	Pageviews	Unique Pageviews	Avg. Time on Page	Bounce Rate	% Exit
/view/	Glasgow	189	119	150.08	53.85%	15.87%
/cgi/	Glasgow	179	121	417.16	28.57%	10.06%
/	Glasgow	81	58	78.72	31.37%	33.33%
/cgi/	London	43	34	41.75	0.00%	25.58%
/4914/	Milan	36	1	57.91	0.00%	2.78%
/cgi/	Hamilton	32	18	21.72	0.00%	9.38%
/view/	Preston	31	21	14.71	0.00%	9.68%
/cgi/	Lake Forest	24	10	95.09	0.00%	4.17%
/cgi/	Xinxiang	23	16	29.71	0.00%	8.70%
/cgi/	Guangzhou	20	11	24.22	100.00%	10.00%
/13436/	Leicester	17	12	14.00	72.73%	70.59%
/cgi/	(not set)	16	15	49.08	0.00%	18.75%
/view/	(not set)	14	14	16.00	77.78%	64.29%
/37619/	Santiago	13	12	501.00	91.67%	92.31%
/9593/	(not set)	13	10	9.00	80.00%	76.92%
/cgi/	Algiers	13	9	29.33	0.00%	7.69%
/cgi/	Edinburgh	13	11	43.71	0.00%	46.15%
/cgi/	Thessaloniki	13	11	40.82	0.00%	15.38%
/view/	Thessaloniki	13	6	410.20	33.33%	23.08%
/cgi/	Coventry	11	8	87.00	0.00%	27.27%
/cgi/	Newcastle upon Tyne	11	8	28.70	0.00%	9.09%
/view/	London	11	8	7378.50	50.00%	45.45%
/view/	Motherwell	11	8	14.50	0.00%	9.09%
/view/	Kharkiv	11	6	32.60	0.00%	9.09%
/9666/	Glasgow	10	3	23.00	66.67%	30.00%

Strathprints path	City	Pageviews	Unique Pageviews	Avg. Time on Page	Bounce Rate	% Exit
/cgi/	Changchun	10	8	166.44	0.00%	10.00%
/cgi/	Massy	10	2	465.33	0.00%	10.00%
/cgi/	Hong Kong	10	10	12.00	0.00%	40.00%
/view/	Islamabad	10	10	299.50	0.00%	20.00%
/view/	Samara	10	2	58.11	0.00%	10.00%
/7282/	Shibuya	9	8	48.00	87.50%	88.89%
/cgi/	Dendermonde	9	6	177.44	0.00%	0.00%
/cgi/	Berne	9	5	29.50	0.00%	11.11%
/cgi/	Birmingham	9	5	136.67	0.00%	33.33%
/view/	Guildford	9	6	12.00	0.00%	11.11%
/	Hamilton	8	6	56.14	0.00%	12.50%
/20088/	Seoul	8	1	285.75	0.00%	0.00%
/30882/	Glasgow	8	4	94.50	25.00%	50.00%
/31032/	Petergof	8	3	9699.40	33.33%	37.50%
/46292/	Glasgow	8	1	133.14	0.00%	12.50%
/7599/	Iowa City	8	6	274.00	50.00%	62.50%
/cgi/	Aberdeen	8	7	4.67	100.00%	25.00%
/cgi/	Bangalore	8	6	25.25	100.00%	50.00%
/cgi/	Tirunelveli	8	2	77.71	0.00%	12.50%
/view/	Liverpool	8	7	6.50	100.00%	25.00%
/view/	Hamilton	8	7	168.00	66.67%	50.00%
/13436/	Indore	7	5	70.00	60.00%	71.43%
/2603/	Chennai	7	3	16.00	33.33%	42.86%
/26994/	(not set)	7	7	0.00	100.00%	100.00%
/27272/	Dublin	7	2	10883.80	0.00%	28.57%
		8821	6868	446.42	73.13%	60.55%