DISSEMINATING SERVICE REGISTRY RECORDS

ANN APPS

MIMAS, Manchester Computing, The University of Manchester
M13 9PL, UK
e-mail: ann.apps@manchester.ac.uk

Abstract: The JISC Information Environment Service Registry (IESR) contains descriptions of collections of resources available to researchers, learners and teachers in the UK, along with technical service access details. This paper describes the data model and metadata description schema of IESR, and the services IESR provides to disseminate its records. There is a particular focus on the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) interface, including a possible use scenario. IESR's position within a wider information environment, its relationship to other initiatives, and future possible service registry directions are discussed.

Keywords: service registry; collection description; service metadata; OAI-PMH; middleware

INTRODUCTION

The JISC Information Environment Service Registry (IESR) [1] contains information about collections of resources that JISC (the Joint Information Systems Committee) makes available to researchers, learners and teachers within UK Higher and Further Education. In addition, IESR contains details of how to access technical services, both those that make the collections available, and others that play a significant role in the information environment.

The aim of IESR is to assist other applications, such as portals, virtual learning applications or research services, to discover materials that are relevant to their users' interests. It is a middleware, shared service, providing a single central registry within the JISC Information Environment [2], primarily intended for machine-to-machine access. But IESR should ultimately benefit end-users by facilitating more awareness of, and easier access to, relevant resources, for example through the single point of search that a portal provides.

Resources are described in IESR as collections, the intention being to provide discovery of collections of materials or datasets covering a topic matching a researcher's interests. The more specific discovery of an individual resource item would be part of the functionality of an application such as a portal with which the user interacts directly. The collections cover all disciplines, and currently include:

- Collections hosted by the JISC Data Centres (MIMAS and EDINA), including UK Census Datasets, the Archives Hub, DigiMap and Biosis
- The Resource Discovery Network (RDN) subject hubs
- Survey Datasets provided by the UK Data Archive

Future example inclusions may be University library catalogues and electronic journals.

IESR also contains descriptions of the services that provide access to a collection, known as 'informational' services. A further set of described services are broker services, called 'transactional' services, which do not provide access to explicit collections but play a significant role within the information environment. Example transactional services are OpenURL resolvers, or emerging Web Services within service oriented architectures such as those of the eLearning and eScience communities. Within IESR context the term 'service' refers to a technical, single access point to a resource collection. The meaning of the word 'service' tends to be overloaded in general use, being understood differently within a variety of contexts, for example a 'JISC service' loosely corresponds to an IESR collection.

Collection descriptions and access details are contributed to IESR by the resource providers, thus assuring accuracy, with a further quality check by the IESR Content Manager. The consequent publicity from registering and hence advertising a resource in IESR is expected to benefit a content provider by bringing additional use.

METHODOLOGY

IESR DATA DESCRIPTION

Resources are described within IESR by metadata according to an IESR-specific scheme, but based on open standards. The IESR data model comprises three types of entity:

- A Collection
- A technical, machine Service that is either: 'informational', providing a single access point to a collection; or 'transactional', acting as a stand-alone broker
- An Agent that owns a collection or administers a service

A collection may have many services, but must have at least one registered in IESR. A collection and a service may have one or many associated agents. An agent may be a collection owner or a service administrator, or both, and may be associated with multiple collections and services. A unique, global identifier, a URI (Internet Uniform Resource Identifier) is assigned to each entity on registration with IESR. These identifiers link between entities within published descriptions. But, being URIs, they are also available to identify IESR entities within the JISC Information Environment and beyond.

The IESR metadata scheme, including the semantics and occurrence requirements of the properties, is documented as a Dublin Core Application Profile [3]. For downstream machine processing the metadata is published within an XML encoding, defined by an XML schema. However the metadata model for each entity is a 'flat' set of repeatable properties each with a single value, rather than a hierarchical XML-style model, making it consistent with the Dublin Core Metadata Initiative's (DCMI) [4] abstract data model [5]. The metadata properties to describe collection and service entities are indicated briefly below. An earlier paper describes data capture by the IESR implementation [6] as composite descriptions within the IESR database, alongside a 'Meta Registry' that holds details of the relations between entities.

Collection Description

IESR collection metadata is based on open standards as far as possible, being consistent with the NISO Metasearch Initiative Collection Description Specification [7] and the DCMI Collection Description Application Profile [8], IESR metadata definition having informed the NISO and DCMI working groups and vice versa. All these collection description schemes are derived from the earlier 'de facto' standard Research Support Libraries Programme (RSLP) Collection Description Schema [9], with some simplification by IESR to describe electronic resources only.

The collection metadata includes properties to enable resource discovery such as a collection's title, description and subject terms. Restrictions on property values are defined by vocabulary encoding schemes that are specified in the application profile. To provide coherent, consistent searching over all collections, IESR uses the Dewey Decimal System as a common, 'backbone' vocabulary, thus requesting that each collection description includes at least one Dewey term. A further, IESR-specific, metadata property, 'usesControlledList', captures the controlled vocabularies used by a collection to describe its items. The list of such schemes is an IESR-defined list, extensible on request, the growth of which reflects the multitude of different domain-specific vocabularies. This information may be useful to a terminology service that maps between vocabularies. It would inform a portal that, having discovered a collection of possible interest to a user, wants to provide an item level search over the collection using an appropriate vocabulary.

Figure 1 shows an example collection description in XML format, as published by IESR via its Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) [10] service (described in a later section). The actual IESR record has been abbreviated for inclusion in this paper, cutting short the abstract and omitting many of the subject terms. Also for brevity the administrative metadata has been omitted, but would be as shown in Figure 3.

```
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2006-02-20T15:16:43Z</responseDate>
  <request verb="GetRecord" identifier="oai:iesr.ac.uk:1088607373-15883"
    metadataPrefix="oai_iesr">http://iesr.ac.uk/service/iesroai</request>
  <GetRecord> <record> <header>
    <identifier>oai:iesr.ac.uk:1088607373-15883</identifier>
    <datestamp>2005-09-14</datestamp> </header>
    <metadata>
      <iesrd:iesrDescription xmlns:dc="http://purl.org/dc/elements/1.1/"
        xmlns:dcterms="http://purl.org/dc/terms/" xmlns:rslpcd="http://purl.org/rslp/terms#"
        xmlns:colldesctype="http://purl.org/cld/colldesc/type/"
        xmlns:iesr="http://iesr.ac.uk/terms/#" xmlns:iesrd="http://iesr.ac.uk/"
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:schemaLocation="http://iesr.ac.uk/ http://iesr.ac.uk/schemas/2005/09/06/iesr.xsd">
        <iesr:Collection>
          <dc:identifier xsi:type="dcterms:URI">
            http://purl.org/poi/iesr.ac.uk/1088607373-15883</dc:identifier>
          <dc:title>Arts and Humanities Data Service</dc:title>
          <dcterms:alternative>AHDS</dcterms:alternative>
          <dcterms:abstract>
            The Arts and Humanities Data Service (AHDS) is a digital archive that collects,
            preserves, promotes and provides access to the electronic resources which
            result from research and teaching throughout the arts and humanities in UK
            higher and further education.</dcterms:abstract>
          <dc:type xsi:type="dcterms:DCMIType">Collection</dc:type>
          <dc:subject xsi:type="dcterms:LCSH">Archaeology</dc:subject>
          <dc:subject xsi:type="dcterms:LCSH">Art</dc:subject>
          <iesr:usesControlledList xsi:type="iesr:CtrldVocabsList">
            AAT</iesr:usesControlledList>
          <dcterms:isReferencedBy xsi:type="dcterms:URI">
            http://www.ahds.ac.uk</dcterms:isReferencedBy>
          <iesr:hasService xsi:type="dcterms:URI">
            http://purl.org/poi/iesr.ac.uk/1088607166-14586</iesr:hasService>
          <rslpcd:owner xsi:type="dcterms:URI">
            http://purl.org/poi/iesr.ac.uk/1088606945-12868</rslpcd:owner>
        </iesr:Collection> </iesrd:iesrDescription>
    </metadata> </record> </GetRecord> </OAI-PMH>
```

FIGURE 1. AN EXAMPLE IESR COLLECTION DESCRIPTION SUPPLIED AS AN OAI-PMH GetRecord

Service Description

An IESR service description provides details of how to access or search a collection. IESR uses a bespoke scheme to provide a simple set of service metadata to support resource discovery, which effectively acts as a 'wrapper' for fuller access details encoded according to the appropriate schema for the service protocol.

When designing the metadata scheme it was apparent that more information was required than simply RSLP's 'locator', which defines the access point to a collection. Links between the IESR entities, service protocol, and authentication information are needed, as well as title and description for a transactional service. In fact, in data modelling terms, an IESR service is actually a conflation, made for pragmatic reasons, of the 'location' of a collection and a 'service' provided at that location. Several standard service description schemes were considered, including Web Services Description Language (WSDL) [11] and ZeeRex [12], which is used to describe digital library service interfaces. But IESR, needing to cover a wider range of service types and to include some IESR-specific details, preferred a simpler scheme to record basic information about a service for resource discovery. Ideally a scheme needed to be consistent with the Dublin Core 'flat' data model, rather than a hierarchical XML model.

Every IESR service description has a single technical access method, for example Z39.50 (a standard machine protocol for information retrieval), SOAP (a W3C Web Services protocol), or OAI-PMH, and a location URL. A simple 'web page' is another possible service type, although this is not a machine-to-machine interface, many collections within the JISC Information Environment being accessible only in this way.

An 'interface' property, referencing a further definition in an appropriate format, is included for service types for which more details are needed to make a connection. For Z39.50 the 'interface' is a ZeeRex specification detailing the search attributes and result formats supported, whereas for SOAP it is a WSDL definition. The 'interface' property is not included for services whose protocols include dissemination of connection information. It is possible to determine details of an OAI-PMH service via an 'Identify' request, making duplication of such data in IESR redundant. Similarly the 'interface' property is not included where connection details are defined by a standard, such as an OpenURL resolver.

Figure 2 shows an example description of a service that provides OAI-PMH access to the collection of Figure 1. For brevity only the IESR metadata is shown, omitting the OAI-PMH wrapper in which it was disseminated and the namespace declarations.

```
<iesr:Service>
 <dc:identifier xsi:type="dcterms:URI">
  http://purl.org/poi/iesr.ac.uk/1088607166-14586</dc:identifier>
 <dc:title>AHDS OAI-PMH Service</dc:title>
 <dcterms:abstract>
  Arts and Humanities Data Service OAI-PMH Repository</dcterms:abstract>
 <rslpcd:locator xsi:type="dcterms:URI">
  http://ahds.ac.uk/srvc-oai/provider</rslpcd:locator>
 <dc:type xsi:type="iesr:AccMthdList">oai-pmh</dc:type>
 <dc:type xsi:type="dcterms:DCMIType">Service</dc:type>
 <dcterms:accessRights xsi:type="iesr:AuthList">none</dcterms:accessRights>
 <iesr:supportsStandard xsi:type="iesr:StdsList">oai-pmh-2_0</iesr:supportsStandard>
 <rslpcd:administrator xsi:type="dcterms:URI">
  http://purl.org/poi/iesr.ac.uk/1088606945-12868</rslpcd:administrator>
 <iesr:serves xsi:type="dcterms:URI">
  http://purl.org/poi/iesr.ac.uk/1088607373-15883</iesr:serves>
</iesr:Service>
```

FIGURE 2. AN EXAMPLE IESR SERVICE DESCRIPTION

Metadata Rights

Every IESR entity has an associated set of administrative metadata, which includes provenance data and a licence to reserve some rights concerning reuse of the entity description. IESR descriptions are freely available for non-commercial use under a Creative Commons licence [13], conditional on attribution of provenance and the same licence being maintained. Data suppliers agree to this licence when they contribute metadata descriptions of their resources. Requiring a common licence for all IESR records simplifies both the IESR implementation, and subsequent reuse requirements of IESR records, for a multitude of purposes including those that may not yet be apparent. This requirement does not appear unreasonable because IESR records are simple metadata descriptions, not complex resources, and because their initial purpose is to advertise the resources they describe. An example of IESR administrative metadata is shown in Figure 3.

IESR SERVICES

IESR provides access to its metadata by several services, based on standard interfaces, and further services are planned. Although of secondary purpose, it has a Web interface that allows searching and browsing of the content of the registry. Thus discovery of IESR descriptions is available to humans, maybe data suppliers checking their records, builders of static portals looking for relevant material to include, or anyone for general resource discovery. Also for human use is the IESR data entry and update service, a Web form editor, which enables the submission of entity descriptions and the relationships between them, according to the IESR metadata scheme.

The choice of the initial machine-to-machine services provided by IESR reflects the fact that its development is within a digital library related environment.

- An OAI-PMH [10] interface allows harvesting of records from IESR according to modification date criteria, and hence subsequent maintenance of the harvested data.
- A Z39.50 [14] service supplies XML records, with the option of full IESR XML descriptions, or simple Dublin Core versions. The IESR XML description published by the Z39.50 interface is composite, returning details of a collection and all its associated services and agents. This removes the need for subsequent retrievals by an application wanting to provide access to a discovered collection. This composite record reflects its internal storage within the IESR database.
- Again from the scholarly information community, an OpenURL 'Link-To' Resolver [15] service uses a standard way of passing information to enable retrieval of a single entity as IESR XML. The OpenURL 'Link-To' Resolver enables resolution of IESR identifiers.

THE IESR OAI-PMH SERVICE

IESR Items

By contrast to the Z39.50 service, the OAI-PMH service supplies single entity descriptions (collection, service or agent). The OAI-PMH specification requires the dissemination of a metadata record corresponding to a single item in the database, making the supply of a composite description inappropriate. Thus data for the OAI-PMH service is serviced via the Meta Registry, which holds basic details of each entity including its latest modification date and its relation to other entities. OAI-PMH item identifiers are derived by a simple transformation from IESR identifiers, made possible by the use of the PURL-based Object Identifier (POI) scheme [16] for identifier construction.

IESR being a registry, rather than a repository, is a collection of 'catalogue' records. Its constituent items are metadata records rather than the more complex items of information generally found in a repository. Thus the metadata records disseminated by IESR correspond directly to the metadata records within IESR rather than being metadata records about them.

IESR does not currently implement OAI-PMH Sets. However there are future possibilities to disseminate descriptions according to particular views onto IESR. An obvious splitting of IESR data into sets would be by entity type. One may envisage an application that has interest only in collection descriptions. Another possibility would be to organise sets by subject to assist sharing of descriptions with subject-specific applications or repositories. For instance a Physics portal would be interested only in resources related to Physics.

IESR Metadata Formats

OAI-PMH requires the dissemination of records using simple Dublin Core, for interoperability, and IESR does conform to this expectation. But the transformation of IESR entity descriptions to simple Dublin Core is a 'dumbing down' process that results in loss of specificity. Thus IESR provides an additional metadata format (metadataPrefix oai_iesr), in which records are disseminated as full IESR XML descriptions. Because this is a proprietary metadata format it will not be understood by harvesters in general, but it is assumed that harvesters outside the JISC Information Environment domain will simply ignore it. The benefits of providing full IESR descriptions seemed to outweigh the disadvantage of supplying a non-standard metadata format.

Full details of the mapping from the IESR metadata properties to OAI-PMH simple Dublin Core are available from the IESR Web site, as are examples of records supplied in this format. Inadequacies of the Dublin Core records are:

- Some IESR properties are omitted from the Dublin Core format because there is no appropriate Dublin Core property, for example 'usesControlledList' for a collection; or do not translate to simple Dublin Core because they describe contained items rather than the collection itself, for example 'itemType'.
- Simple Dublin Core does not permit the inclusion of vocabulary schemes, and the possibility of including these in a non-standard way was thought inappropriate. Thus all subject terms are supplied by their value only, making their usefulness to downstream applications doubtful.
- Inevitably many properties dumb down to 'relation' in simple Dublin Core. These include links to related entities as well as other relation properties within the description. Links within IESR may be distinguished by their identifier scheme, but their purpose cannot be deduced from the Dublin Core record.

Some alleviation of the inadequacy of the simple Dublin Core description is provided by the inclusion of a by-reference pointer as the value of an additional 'relation' to enable retrieval of the corresponding full IESR description. This by-reference link is an OpenURL ContextObject [15], a means of describing an identified resource in a standard way, retrieval on access being via the IESR OpenURL Link-To resolver service

By contrast IESR disseminates complete XML descriptions via the 'oai_iesr' metadata format, which include all the data registered with IESR for a particular entity. Related entities are identified by their IESR URI within the appropriate property, for example a collection's owner agent is identified by the value of the 'owner' property. Thus it would be possible to build a registry similar to IESR by harvesting the entity descriptions and reassembling them according to the implementation design of the replica registry. Updated descriptions are disseminated per entity.

The associated administrative metadata for an entity, which captures its provenance, is returned in the 'about' section of the OAI-PMH record. This is in contrast to the descriptions disseminated by other protocols that do not make this provision. In these cases the administrative metadata is included as a separate description within a containing description set, indicated by XML identifiers and references within the description set, a solution that

seems less elegant than the one provided by OAI-PMH. Figure 3 shows the administrative metadata that would be supplied as part of the OAI-PMH GetRecord of Figure 1.

```
<about>
  <oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
   xmlns:dc="http://purl.org/dc/elements/1.1/"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
    http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
   <dc:creator>http://www.ahds.ac.uk</dc:creator>
   <dc:publisher>http://iesr.ac.uk</dc:publisher>
   <dc:date>2005-09-14</dc:date>
   <dc:rights>http://creativecommons.org/licenses/by-nc-sa/2.0/uk/</dc:rights>
   <dc:rights>This IESR administrative metadata must always be retained
     with its associated entity description.</dc:rights>
  </oai_dc:dc> </about>
```

FIGURE 3. AN EXAMPLE ADMINISTRATIVE RECORD FOR A COLLECTION

Additionally expressions about rights for reuse of the supplied records are included in the 'about' section according to the appropriate OAI-PMH framework schemas. There are two rights fields. The first is a 'rights reference' to the Creative Commons licence. The second is a 'rights definition' containing the IESR provenance requirement. Figure 4 shows the rights statements that would be supplied as part of the OAI-PMH GetRecord of Figure 1.

```
<about>
  <rights xmlns="http://www.openarchives.org/OAI/2.0/rights/"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/rights/
    http://www.openarchives.org/OAI/2.0/rights.xsd">
   <rightsReference ref="http://creativecommons.org/licenses/by-nc-sa/2.0/uk/rdf" />
  </rights> </about>
<about>
  <rights xmlns="http://www.openarchives.org/OAI/2.0/rights/"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/rights/
    http://www.openarchives.org/OAI/2.0/rights.xsd">
   <rightsDefinition>
    <oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
     xmlns:dc="http://purl.org/dc/elements/1.1/"
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
      http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
     <dc:description>The IESR administrative metadata must always be retained
       with its associated entity description.</dc:description>
    </oai_dc:dc> </rightsDefinition> </rights> </about>
```

FIGURE 4. EXAMPLE RIGHTS FOR A COLLECTION

FUTURE SERVICES

Several further interfaces are planned including an RSS (a protocol for syndication) [17] service for new data alerts, a NISO Metasearch XML Gateway (MXG) [18] service, and a Web Services SOAP [19] interface. The particular design and details of the Web Services SOAP interface has not yet been decided, but, like other interfaces, it will supply IESR descriptions both as simple Dublin Core and as full IESR XML. There will be an interface that conforms to 'Search / Retrieve Web' (SRW) [20], which is a standard Web Services protocol, registered by NISO, within the digital library community. A Web Services SOAP interface would also enable the use of IESR within frameworks based on service oriented architectures, such as those of the eScience and eLearning communities.

RESULTS

USING IESR

IESR has had little real use as yet, its proposed deployment being still rather visionary. IESR envisages that applications such as metasearch portals will be some of its significant users. A portal "brings together content from a diverse range of resources, which it collates into an amalgamated form for presentation to an end-user" [21]. A metasearch portal provides a search across a range of resources that are appropriate for a user, for example those applicable within a particular subject domain. A possible use scenario of IESR via its OAI-PMH interface is described below. Some other possibilities of use are descried in an earlier paper [22]. IESR expects to define some more formal use cases in the near future, hoping that these will encourage practical use.

USING IESR VIA OAI-PMH

Possibly a portal would harvest IESR descriptions using OAI-PMH and cache them locally, maybe because it would prefer to operate from its own local service registry for performance reasons. Or perhaps a metasearch portal would harvest IESR records into its 'knowledge base' after conversion to an appropriate format. The portal would discover collections of interest to an end user, according to their topic of interest, by searching the copy of the IESR data within its local database, having translated the search topic into a broader, preferably Dewey, term appropriate for searching IESR data. At the same time the portal would determine service access details for the discovered collections. It would then provide to the end user a metasearch across these collections, using a service protocol appropriate for the portal and the user's environment. For example a portal that operates via Z39.50 would provide the metasearch using the Z39.50 details for access to the collections, and ignore any discovered collections that do not have a Z39.50 interface. Resources are described within IESR as collections rather than at item level, the intention being to provide discovery of a collection of resources that may cover a topic in which a user is interested. The subsequent inclusion of that collection in a metasearch is part of a portal's functionality.

If a portal harvests records frequently its use of IESR is effectively dynamic. Thus the portal builder does not need to know about all available resources, and more will appear when they are registered in IESR. This means that potentially users will discover resources applicable to their domain of interest of which they were unaware.

DISTRIBUTED SERVICE REGISTRIES

Provision of IESR's OAI-PMH interface opens up the possibility of sharing data between service registries, both within the JISC Information Environment and beyond, assuming they use the same, or a derivable, metadata schema. Clearly a single global registry is not viable, being an unscalable solution in terms of maintenance and quality assurance. Data ownership concerns would be better addressed if resource providers were able to contribute their own nodes to a distributed registry. IESR is collaborating with the US OCKHAM Digital Library Services Registry [23], which is using the IESR metadata description profile to capture details of resources developed with US National Science Foundation (NSF) funding. Each OCKHAM node maintains its own descriptions, which are disseminated to other nodes via OAI-PMH harvesting, using a DNS-like model, which ensures that each node has an up-to-date copy of the entire registry. Searching over the aggregated registry is local at the end user's institution.

DISCUSSION

FEDERATED SERVICE REGISTRIES
A wider vision than that of a distributed service registry described above may be a federated grid of service registries, each node being responsible for descriptions of collections and services under its own remit, but contributing to a shared, searchable, virtual service registry. The problem than becomes one of searching across this grid's many nodes.

- One solution may via a metasearch using an appropriate protocol such as MXG. But this maybe too high a barrier for use by a portal application, with an additional concern about performance degradation over a large federated registry.
- Another solution may be to use UDDI [24]. But UDDI is perceived as difficult to use, so also may be too high a barrier. (UDDI is described below with further concerns.)
- If each registry within the grid provides an OAI-PMH interface, harvesting descriptions of interest into an aggregated database, possibility selected according to some criterion such as subject, would provide a means of implementing a metasearch over the entire content.

SHARING RESOURCE DESCRIPTIONS
The current scope of IESR is the registration of resources provided by JISC within the Information Environment to researchers, learners and teachers in the UK. It is possible that this may be widened to include more UK resources, without going as far as the distributed or federated registry model. Currently descriptions within IESR have been contributed directly. Some resources that would be applicable to the current user domain may have already been catalogued elsewhere. In particular descriptions of freely available resources, and materials such as library catalogues and institutional repositories may be already available. In such cases it would seem sensible to import existing descriptions rather than recreate them. However they may need some manual augmentation if imported into IESR. For example a registry of institutional or similar repositories may have details of only OAI-PMH and Web interfaces, whereas a repository may have further interfaces such as SRU which could be recorded in IESR.

Similarly other services may wish to harvest IESR data to populate registries for other purposes. Such an application may be a more restricted registry, such as a local institution calalogue. Or it may a directory of repositories that is interested in OAI-PMH interfaces only. A scenario may be seen developing here where a collection description is created once only by a resource provider, then shared by multiple applications. Each instance of this description may be augmented in different ways. For example IESR may include service access information, whereas an organisation's catalogue may include details pertinent to the focus of its particular users.

OAI REGISTRIES
Several registries exist of repositories that have OAI-PMH interfaces. Some of these, such as OAIster [25], register any resource that is OAI-PMH compliant. Thus IESR itself is registered with OAIster. Other directories such as ROAR [26], OpenDOAR [27] in the UK, and DAREnet [28] in the Netherlands record a more specific set of repositories of open access information. These would include institutional repositories of research outputs, and subject-specific repositories of research literature and associated experimental data. These registries include some subject categorisation of a repository's content, in OAIster's case being manually added for appropriate repositories. However, generally the descriptions of repositories within these registries is less detailed than an equivalent IESR record would be. Additionally they do not record service interfaces to the collection of materials within the repository other than the OAI-PMH and Web page access points.

REUSE RIGHTS

Sharing resource descriptions between applications raises issues about catering for different licences concerning re-use of records. The IESR requirement of a single Creative Commons licence simplifies this problem in a scenario where IESR registers only records that have been contributed directly. But if descriptions were to be harvested from elsewhere, the IESR data profile would need extension to record several different licences. Obviously a much simpler solution would be for all registries to agree on a common licence, but it is doubtful if such a utopian situation is achievable. The Creative Commons licence used by IESR, which allows free non-commercial use conditional on attribution of provenance being maintained, appears to provide the best solution. This maintains some control over re-use whilst permitting 'derivatives', for example, adding local information to records. There may be some commercial interest in IESR records, but this is not forbidden by the licence, rather requiring some negotiation for the use of descriptions created by public funding.

UDDI

The viability of a UDDI (Universal Description Discovery and Integration) [24] standard service based on the IESR data is currently under investigation within the project. UDDI is a metadata standard and protocol for the discovery and publicising of services on offer, which appears to match IESR's remit. UDDI is mainly used in the eBusiness domain, although generally within a company's intranet, not having achieved its initial vision of a single global registry of everything. There is some use within eScience where several projects have developed UDDI registries of Web Services [29], but again generally within projects and applications. There is little apparent use of UDDI within the digital library and scholarly information domains. Although an item on some 'wish lists', it is not clear whether it is not used because there are no implementations available, or vice versa. UDDI was designed to register Web Services that have WSDL interface, and most implementations register SOAP services, although it is possible to describe other service protocols with some contrivance. Thus UDDI does not match IESR's more agnostic view of the registration of a multitude of service protocols.

To build a UDDI version of IESR will require a mapping from IESR metadata. UDDI has a concept of a business entity which appears to correspond to an administrator agent in IESR. UDDI's business service seems to correspond to a conflation of IESR's collection and service, possibly reflecting the more general usage of the term 'service'. Particular service protocols could be defined using a UDDI template. It is possible to attach some extra metadata properties to a business service. But the possibility of capturing the full, rich collection metadata of IESR within a UDDI business service seems unlikely. Thus it appears that an IESR UDDI service, if implemented, would provide a partial view of IESR data. There may be value in providing such an interface if it is needed by applications in the information environment. There is also value in providing an IESR UDDI interface to evaluate practically whether such a service would be used.

A parallel development within the eScience domain is the Grimoires [30] service registry software available within the common, shared corpus of UK eScience software. Grimoires does provide a UDDI interface, this being a requirement of some applications within the eScience arena. However Grimoires provides additional, richer metadata descriptions, with associated alternative interfaces, indicating a similar conclusion to IESR's that a UDDI registry is able to provide only a partial view of the registered descriptions.

The UDDI protocol is designed to support federated service registries, keeping nodes up-to-date. Thus it would seem the appropriate choice to implement such an infrastructure. However the partial, 'lossy' view of the IESR collection descriptions implicit in a UDDI implementation, appears to preclude its viability as a platform for a federated service registry based on the IESR data model.

CONCLUSIONS

IESR has developed a registry of collections and their associated services within the JISC Information Environment. There is still work to be done, adding further resource descriptions and developing further services to access the IESR metadata. In the immediate future IESR aims to move from its current visionary status, by the development of use case examples and practical demonstrations. A significant corpus of material within IESR, some guarantee of its persistence, and applications to whose performance it is critical, should all help to encourage use.

IESR is agnostic about particular service access protocols. It has a remit to register any service type that is widely used within the Information Environment, unlike some other environments or registry models that are based on one protocol such as Web Services SOAP or OAI-PMH. Similarly it provides multiple service interfaces into its content. However development of realistic use cases, designs of distributed registries and methods of sharing descriptions, and discussions about federated service registries, all point to the significant viability of OAI-PMH as the transport protocol of choice. In addition the OAI-PMH protocol provides appropriate containers for the dissemination of metadata about descriptions, such as provenance and rights details, information that requires a proprietary solution within other protocols.

IESR has built a practical registry based on collection, service and agent descriptions. In particular IESR collection description is based on emerging standards in this area, the NISO Metasearch Initiative Collection Description Specification and the DCMI Collection Description Application Profile. Members of the IESR project have made significant contribution to the development of these standards, as well as developing a concrete registry that provides an exemplar of their use. Additionally IESR disseminates collection descriptions using OAI-PMH harvesting. This combination of a standard metadata description and a widely used dissemination format demonstrates the viability of sharing descriptions between registries, lowering the global cost of their production and ensuring their quality.

ACKNOWLEDGEMENTS

NOTES AND REFERENCES

1 IESR: JISC Information Environment Service Registry. http://iesr.ac.uk/
2 Investing in the Future: Developing an Online Information Environment. JISC. 2003. http://www.jisc.ac.uk/index.cfm?name=ie_home
3 CWA 14855, Dublin Core Application Profile Guidelines. CEN/ISSS. 2003. ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/MMI-DC/cwa14855-00-2003-Nov.pdf
4 DCMI: The Dublin Core Metadata Initiative. http://www.dublincore.org/
5 Powell, A., Nilsson, M., Naeve, A. and Johnston, P. DCMI Abstract Model. 2005. http://www.dublincore.org/documents/abstract-model/
6 Apps, Ann. A Registry of Collections and their Services: from Metadata to Implementation. *DC-2004: Proceedings of the International Conference on Dublin Core and Metadata Applications, October 11-14 2004, Shanghai, China*. Shanghai Scientific & Technical Literature Publishing House. 2004. pp. 67-73.

7 ANSI/NISO Z39.91-200x Collection Description Specification. 2005.
http://www.niso.org/standards/resources/Z39-91-DSFTU.pdf
8 DCMI Collection Description Application Profile. 2005. via
http://www.dublincore.org/groups/collections/
9 Powell, A. RSLP Collection Description Schema. 2000.
http://www.ukoln.ac.uk/metadata/rslp/schema/
10 Lagoze, C., Van de Sompel, H., Nelson, M. and Warner, S. The Open Archives Initiative
Protocol for Metadata Harvesting. 2004.
http://www.openarchives.org/OAI/openarchivesprotocol.html
11 Christensen, E., Curbera, F., Meredith, G. and Weerawarana, S. Web Services Description
Language (WSDL) 1.1. 2001. http://www.w3.org/TR/wsdl
12 ZeeRex: The Explainable "Explain" Service. 2004. http://explain.z3950.org/
13 Creative Commons. http://creativecommons.org/
14 ANSI/NISO Z39.50-1995 Information Retrieval. 1995.
http://www.niso.org/standards/resources/z3950.pdf
15 Apps, Ann and MacIntyre Ross. Emerging Uses for the OpenURL Framework.
*ELPUB2005: Proceedings of the 9th ICCC International Conference on Electronic
Publishing, 8-10 June 2005, Leuven, Belgium.* Leuven: Peeters Publishing. 2005. pp 283-289.
16 Powell, A., Young, J. and Hickey, T. The PURL-based Object Identifier. 2004.
http://www.ukoln.ac.uk/distributed-systems/poi/
17 Hammond, T., Hannay, T. and Lund, B. The Role of RSS in Science Publishing. *D-Lib
Magazine.* 2004. Vol 10 No 12.
18 NISO Metasearch XML Gateway Implementers Guide. 2005.
http://www.niso.org/standards/resources/MI-MXG_v0_3.pdf
19 W3C Web Services SOAP. 2000. http://www.w3.org/2000/xp/Group/
20 Search/Retrieve Web Service. 2004. http://www.loc.gov/standards/sru/srw/
21 Portals Frequently Asked Questions. JISC. 2003.
http://www.jisc.ac.uk/index.cfm?name=ie_portalsfaq
22 Apps, Ann. A Middleware Registry for the Discovery of Collections and Services.
*NCeSS2005: Proceedings of the First International Conference on e-Social Science, 22-24
June 2005, Manchester, UK.* 2005.
23 Frumkin, J. The problem of mainstreaming digital libraries. *OCLC Systems and Services:
International Digital Library Perspectives.* 2004. Vol 20 No 3. pp. 106-109.
24 UDDI Specification. 2006. http://www.uddi.org/specification.html
25 OAIster. 2006. http://oaister.umdl.umich.edu/o/oaister/
26 ROAR: Registry of Open Access Repositories. 2006. http://archives.eprints.org/
27 OpenDOAR: Directory of Open Access Repositories. 2006. http://www.opendoar.org/
28 DAREnet: Digital Academic Repositories. 2006. http://www.darenet.nl/
29 Allen, R., et al. Building the eScience Grid in the UK: Grid Information Services.
*AHM2003: Proceedings of the UK e-Science All Hands Meeting, 2-4 September 2003,
Nottingham, UK.* 2003.
30 Wong, S.C., Tan, V., Fang, W., Miles, S. and Moreau, L. Grimoires: Grid Registry with
Metadata Oriented Interface: Robustness, Efficiency, Security – Work-in-Progress.
CCGrid05: Cluster, Computing and Grid, 9-12 May 2005, Cardiff, UK. 2005.