The JISC Information Environment Service Registry

Pete Johnston (UKOLN)

Services in the JISC Information Environment

The Joint Information Systems Committee (JISC) administers the provision of a wide range of digital content for use in higher and further education. That content is of diverse types (text, still images, maps, video, audio, datasets, software, learning resources) and is made available by different classes of content provider (individual educational institutions, sector-wide providers operated by JISC, commercial publishers, other public sector sources) and under varying different terms and conditions (open access, institutional subscription) [1].

The JISC Information Environment is a "set of networked services that allows people to discover, access, use and publish… resources… within the UK HE and FE community" [2]. The JISC IE is not itself a single system or service; rather, it is a framework of services that can be used in combination as components to deliver functionality of interest to a user. Furthermore, the environment is not a closed one: some of the service components used in the JISC IE to support educational users are also used to deliver services to other user communities.

Central to the concept of the Information Environment is the principle that service components are **used in combination** to deliver functions to a user. Since many of the resources of interest are physical resources (books, paper documents), users frequently combine the use of digital services (e.g. discovering, selecting and locating an item by querying a library catalogue) and physical services (requesting and accessing the located item by using the reference or lending service of a library, or submitting a request to a document delivery service).

In the past the capacity to combine digital service components has depended on **human** effort to 'join up' the components: for example, a researcher might obtain information from (the human-oriented interface of) one service and then use that information as input to a second service. In the course of that process the researcher locates and works with different interfaces - typically Web sites accessed through a Web browser, but perhaps also services accessed through desktop tools like bibliographic reference managers. The user may also need to make "adjustments" to the information obtained from one service before using it as input to a second service, in order to account for differences in the "semantics" used by the two interfaces. Often the task at hand requires the discovery of, and access to, multiple resources, through the use of multiple services – and the human labour involved may be considerable. The JISC IE technical architecture specifies a set of standards and protocols that service components should support so that they can exchange information effectively with other service components, that is, it provides standards for **machine**-oriented interfaces between service components, with the aim of reducing the effort required by the human user to use such services in combination [3, 4].

The IE Service Registry

To use a digital service effectively, a human user requires several pieces of information. They must first **discover** the existence of services of potential interest - perhaps by consulting a directory of services available. They then **select** those services most appropriate to the task at hand based on the information provided about the services, which may include factors such as the conditions of access to the service as well as the functions it offers. Different users may have different criteria for making that selection: some users may be willing to pay a fee for a service, or to accept a delay in obtaining the output, while others may require free access and immediate results; even for the same person, the factors affecting selection may vary depending on their context. Having made their selection, the user needs to **locate** the point at which a service can be accessed – for example the URL of a page providing an HTML input form – and they need to understand how to communicate with the selected service, that is, how to use the interfaces provided to **access** the service (provide arguments for search queries and interpret the results of such queries, for example).

These same requirements apply when the user of a service is a software application, and these are the functions that the JISC IE Service Registry is intended to support. That is, the IESR enables software applications to:

- Discover and select services of interest
- Determine how they can interact with a selected service (obtain information on the interfaces provided, and locate access points to specific implementations of those interfaces)

The IESR provides functions which, in theory at least are of potential interest to any other service component in the IE: in terms of the high-level functional classification of services used in the technical architecture, the IESR is a shared infrastructural service, or middleware service [5].
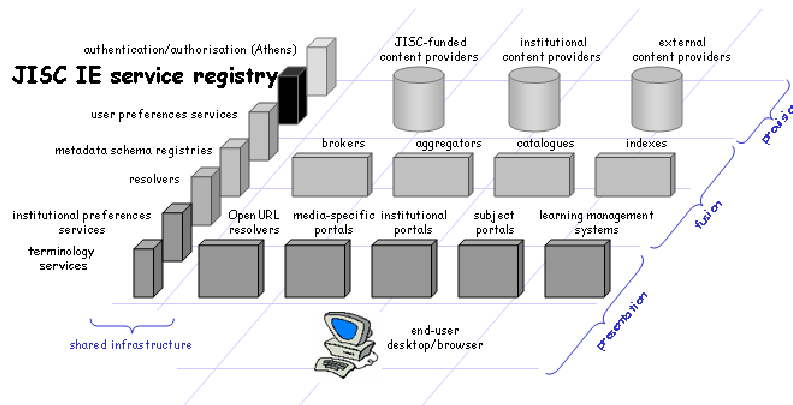

Figure 1 :The JISC Information Environment Technical Architecture

The functionality provided by the IESR is intended primarily for use by software tools, and the interfaces provided by the IESR are machine-oriented, rather than human-oriented. The IESR does not seek to provide a human-readable Web-based directory or the JISC IE – the IESR is not a presentational service. Some of the information made available by the IESR may be presented to human users, but that presentation function is performed by other service components. For example, a presentation service might display information about collections obtained from the IESR so that a human user of that presentation service can select other services with which to interact (as targets for a broadcast query, for example).

However, it should be noted that the ability of an application to interact with a "newly-discovered" service depends on the application having "prior knowledge" of the semantics of the interface to that service. Where services are based on standard protocols, this may be possible since the requesting application can be programmed in advance to support that protocol: for example, if an application obtains the information that a service component supports the OAI Protocol for Metadata Harvesting, and it obtains the base URL for the OAI repository, then it can interact with that service using the "verbs" defined by the OAI-PMH. In other cases, however, the semantics of the interface are service-specific, and the capacity of the requesting service to access the described service is conditional on the intervention of a human administrator to interpret those semantics and configure the requesting application accordingly.

Collection-level description and the JISC IESR

Content in the JISC IE is provided in the form of collections. A subset of services in the JISC IE has as their primary function the provision of access to items within a collection, or to metadata records about items within a collection (a catalogue); in the terms of the technical architecture, these services are **informational** services [6].

The decision to make use of an informational service is based on attribute(s) of the collections to which the service provides access, as well as attributes of the service itself. For example, a human user may elect to search a library catalogue on the basis that it contains records describing items relevant to a particular subject area. Similarly, a portal dedicated to a specific subject area or discipline selects a range of catalogues to make available for searching on the basis that they are likely to contain descriptions of relevant items. The initial version of the JISC IE technical architecture describes a distinct collection description service and service description service [3]. However, it was always recognised that the usage of these two potential components was closely related. The functionality of some services means that any assessment of their relevance requires information about the collection to which they provide access; and conversely, it would be of limited value for an application to discover the existence of a collection unless it could also obtain information on the services that

provide access to that collection. Simplifying slightly, the IESR can be regarded an application that itself provides services based on a database of collection descriptions and service descriptions.

The IESR also provides access to descriptions of services which are not related to access to collections (e.g. [examples here]); these are referred to as **transactional** services. The remainder of this paper concentrate primarily on informational services and the role of the IESR in facilitating access to collections, but it should be remembered that these services represent only a subset of the services in the JISC IE.

The IESR pilot project

During 2002-2003, the JISC Shared Services Programme funded a first phase of a project, based at MIMAS, University of Manchester, to explore the use requirements for the IESR and to develop a pilot service registry, with collection and service metadata provided by a small number of JISC service providers [7]. A second phase of the project has recently been funded to consolidate this initial work.

The IESR Data Model

The pilot IESR database is based on a data model derived from the RSLP Collection Description model [8], or perhaps more accurately from the subset of the model that underpins the RSLP Collection Description Schema [9]. The primary entity types in the IESR data model are:

- the Collection, ("an aggregation of items")
- the Owner of the Collection ("the Agent who has legal possession of a Collection")
- the Service ("the provision of, or system of supplying, one or more functions of interest to a user")
  - *subtype*: informational Service ("that provide access to, or metadata about, items or Collections at a digital location")
  - *subtype*: transactional Service (that provide other functionality)
- the Administrator of the Service ("the Agent responsible for the digital environment in which the Collection is held")
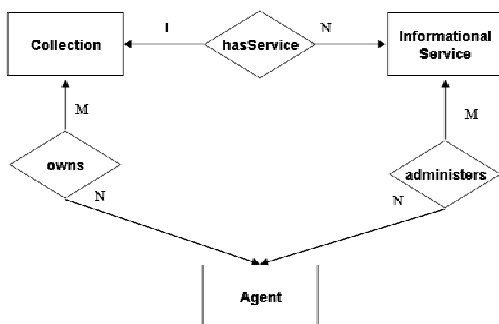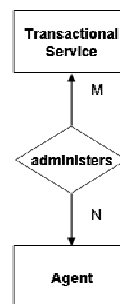

Figure 2 : Informational Services   **Figure 3 : Transactional Services**

The IESR metadata schema

The IESR metadata schema [10] is based on the RSLP CD Schema, and like the RSLP CD Schema it supports the description of a number of related entities and the relationships between those entities. The schema provides properties to represent:

- attributes of a Collection
- attributes of the Owner of a Collection
- attributes of a Service
- attributes of the Administrator of a Service
- the relationships between these entities
- relationships between multiple Collections (is-Part-Of)
- relationships between these entities and some external entities (logos, publications)

(The IESR schema also provides a small set of administrative metadata properties which support the creation of metadata about descriptions of these entities.)

For the description of collections, the IESR schema reuses a subset of the RSLP CD Schema collection properties, many of which are themselves drawn from the Dublin Core metadata vocabularies, and extends that set of properties with a small number of additional properties that are useful for the discovery and selection of collection in the JISC context (notably dcterms:educationLevel and iesr:usesControlledList. See tables below.)

The RSLP CD Schema provides only a minimal set of properties for the description of network services ("digital locations" in the RSLP CD Schema), and the IESR metadata schema provides more descriptive capability in this area. One of the challenges is that service description tends to be heavily protocol-specific because different protocols support different functionality. The features of a Z39.50 target, for example, are different from those of an OAI-PMH repository. The IESR approach to this challenge has been to provide a set of service attributes that are broadly common to services regardless of the protocol on which they are based, and to include a pointer to a protocol-specific specification (such as a WSDL document [11] or a Zeerex record [12]). The IESR schema also prescribes the use of controlled vocabularies ("encoding schemes") for several properties, both for collection and service description.

Note: The following tables provide only a summary overview of the metadata schema used in the IESR pilot project. For a full description of the schema and guidelines for its use, see the "Metadata" section of the IESR project Web site [10]. The metadata schema is due to be reviewed during the second phase of the project in 2004.

Collection description

Descriptive properties of the Collection

Label	Property Name	Definition	Obligation	Occurrences Permitted	Encoding Schemes
Identifier	**dc:identifier**	A formal global identifier for the collection.	Mandatory		
Title	**dc:title**	The name of the collection.	Mandatory		
Alternative title	**dcterms:alternative**	An alternative name, or sub-title of the collection.	Optional	Repeatable	
Description	**dc:description**	A free text summary description of the collection.	Optional		
Type	**dc:type**	The type of the collection.	Mandatory	Repeatable	dcterms:DCMIType rslpcd:CLDT
Physical Characteristics	**dc:format**	The physical or digital characteristics of the collection.	Optional	Repeatable	
Language	**dc:language**	The language of the items in the collection.	Optional	Repeatable	dcterms:RFC3066
Rights	**dc:rights**	Copyright statement about the collection.	Optional		
Use Rights	**iesr:useRights**	A statement about allowed usage of data from the collection, for example terms and conditions.	Optional		
Access Rights	**dcterms:accessRights**	Licence requirements to access the collection.	Optional		
Contents Date Range	**rslpcd:contentsDateRange**	The range of dates of the individual items within the collection.	Optional	Repeatable	dcterms:W3CDTF
Uses Controlled List	**iesr:usesControlledList**	A classification scheme or thesaurus used by the collection	Mandatory	Repeatable	iesr:CtrldVocabsList
Education Level	**dcterms:educationLevel**	Education level of group for whom the collection is intended or useful	Optional	Repeatable	meg:UKEL

Subject/coverage of items in Collection

Subject	**dc:subject**	A single concept (keyword or subject descriptor) that is the subject of the collection, or of the items in the collection.	Mandatory	Repeatable	dcterms:DDC iesr:HASSET iesr:JACS dcterms:LCSH dcterms:MESH iesr:UNESCO
Spatial Coverage	**dcterms:spatial**	The spatial coverage of the items in the collection.	Optional	Repeatable	iesr:HASSET dcterms:ISO3166 dcterms:TGN iesr:UNESCO
Temporal Coverage	**dcterms:temporal**	The temporal coverage of the items in the collection.	Optional	Repeatable	dcterms:W3CDTF

Relationships between Collection and Service, Collection and Agents

Location	**iesr:hasService**	(The global identifier of) a service providing access to the collection	Mandatory	Repeatable	
Owner	**rslpcd:owner**	(The global identifier of) the agent that legally owns or has possession of the collection	Mandatory	Repeatable	

Super-collection	**dcterms:isPartOf**	(The global identifier of) another collection that contains this collection	Optional	Repeatable	
Associated collection	**rslpcd:has Association**	(The global identifier of) another collection that has association by provenance with this collection	Optional	Repeatable	
Associated publication	**rslpcd:hasPublication**	(The global identifier of) a publication that provides more information about the collection, such as a general information webpage; a help guide for using the collection; a report based on the use, study, or analysis of the collection	Optional	Repeatable	
Logo	**iesr:logo**	(The URL of) a logo for the collection	Optional		

Service description

Descriptive properties of the Service

Identifier	**dc:identifier**	A formal global identifier of the service.	Mandatory		
Name	**dc:title**	Name of the service.	Mandatory		
Description	**dc:description**	A free text summary description of the service	Optional		
Locator	**rslpcd:locator**	The access point for the service	Mandatory		
Interface	**iesr:interface**	A resource that provides interface information about the service, such as available inputs and outputs, in a machine readable format	Optional		
Type	**dc:type**	Technical type of interface providing access to service	Mandatory		iesr:AccMthdList
Output	**iesr:output**	Output format available from the service	Optional	Repeatable	dcterms:IMT
Access Rights	**dcterms:accessRights**	Access control for the service	Mandatory	Repeatable	iesr:AuthList
Supports Standard	**iesr:supportsStandard**	Indication of standards and profiles supported by the service	Optional	Repeatable	iesr:StdsList

Relationships between Service and Agents

Administrator	**rslpcd:administrator**	(The global identifier of) the agent who has responsibility for the electronic environment in which the collection is held	Mandatory	Repeatable	

External Relationships

See also	**rslpcd:seeAlso**	(The global identifier of) a document that provides more information about using the service	Optional	Repeatable	

Agent description

Identifier	**dc:identifier**	A formal global identifier for the agent.	Mandatory		
Name	**vcard:FN**	Full name of contact person for agent	Optional	Repeatable	
Organisation	**vcard:ORG**	The organisation that is the agent	Mandatory		
Role	**vcard:ROLE**	The role of the agent	Optional	Repeatable	
Telephone	**vcard:TEL**	The telephone number of the agent	Optional		
Email	**vcard:EMAIL**	Email contact address for agent	Mandatory		
URL	**vcard:URL**	URL for agent	Optional		
Logo	**iesr:logo**	(The URL of) a logo for the agent	Optional		

The JISC IE, the IESR and Web Services

The IESR pilot registry is implemented as an XML database, which is queried using the Cheshire information retrieval system [13]. A simple Web site provides access for the human user; and a Z39.50 target is available for software applications.

Figure 4 provides a high level view of a possible set of interactions with the IESR by (a) a provider of a service who wishes to disclose the existence of their service and provide details of how to access it; and (b) a second service component (here a presentation service but potentially a component in any of the functional categories in the IE) that discovers content provision services and then accesses those services. The use of the arrows on the diagrams is not intended to suggest that these different interactions use the same information exchange protocols: interactions between the service requestor and the IESR (steps 2, 3) may be based on different protocols from those that the service requestor uses to communicate with the discovered services (step 4), for example.
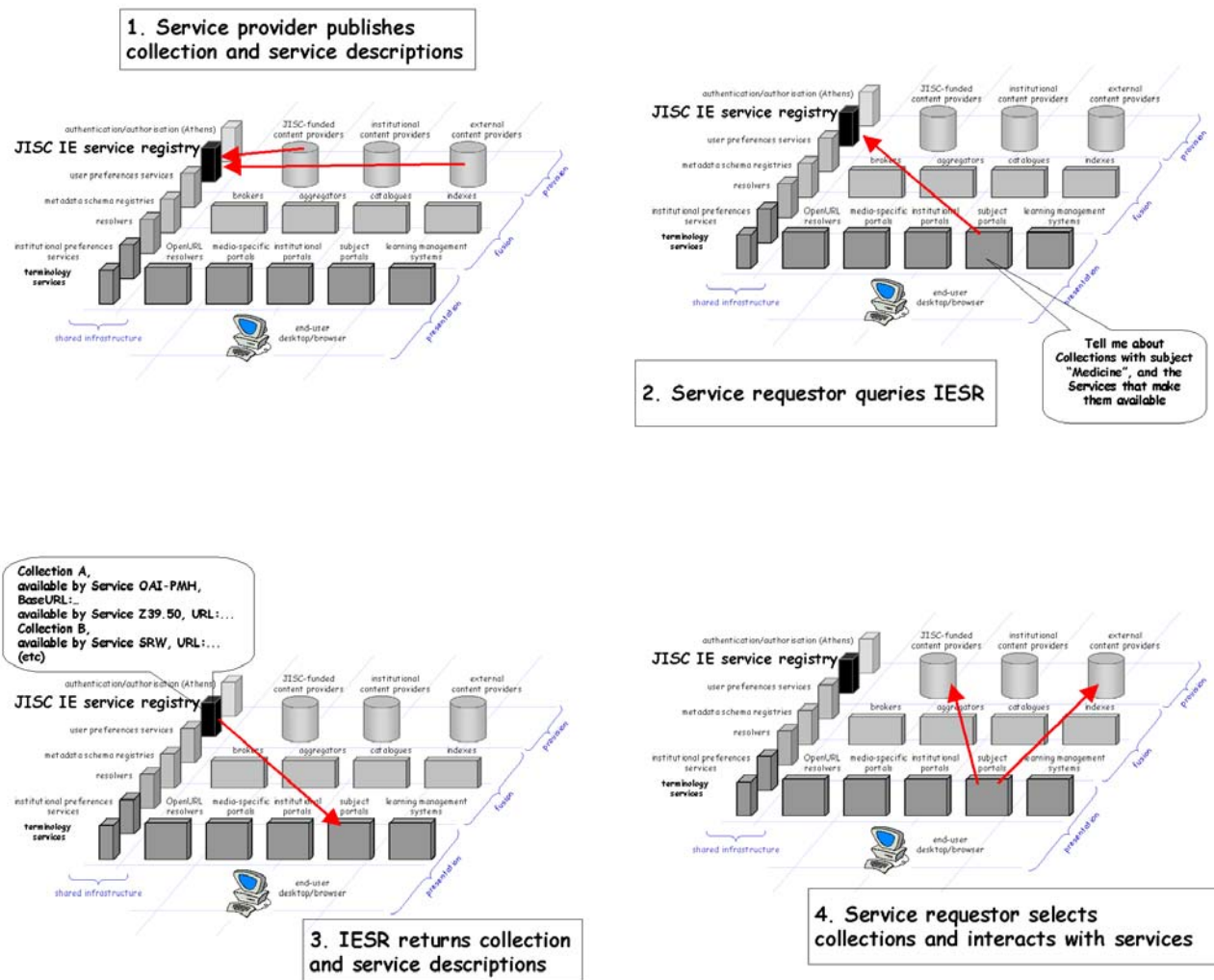
Figure 4 : Interacting with the IESR

During the second phase of the project, it is planned to investigate the provision of SRW/SRU and also OAI-PMH interfaces to the IESR to provide more flexibility in the access mechanisms available. It is important to note that use of the IESR does not depend on the use of Z39.50, and the "real time" search of the IESR outlined in figure 4 is not the only mechanism for using the collection and service metadata that the IESR discloses. An application might harvest records from the IESR at scheduled intervals and manage those harvested records as a "local" database, using whatever storage formats and access mechanisms are most useful for the local application.

The JISC IE technical architecture is broadly similar to and compatible with the conceptual architecture underpinning Web Services, with "service providers" publishing descriptions of their services to the IESR so that "service requestors" can discover them and interact with them [14, 15]. The Universal Description, Discovery and Integration (UDDI) protocol is an OASIS Standard for Web Service description and discovery [16]. UDDI provides both

- a data model for describing services, their providers, and the access mechanisms for the services; and
- a Web Service API for searching and contributing to a UDDI registry

The term "UDDI" is also sometimes used to refer to the UDDI registry implementation provided at uddi.org, also known as the UDDI Business Registry (UBR). Some preliminary studies have been conducted on the use of UDDI within the JISC IE [17], and the possibility of providing a UDDI interface to the IESR is to be explored in phase II of the IESR project.

Powell and Lyon [14] emphasise the importance of aligning the JISC IE technical architecture with the architectures based on Web Services that are being deployed more widely: the "JISC IE" is not a "closed" environment: Services developed primarily for use "within" the JISC IE may also be used by "external" applications, and some of the services used within the IE may have been developed for use by other communities too. In the longer term, it may well be the case that the functionality of the "infrastructural services" is provided not just by a single set of JISC-administered shared applications, but by a number of such services delivered by different providers. Applications within the JISC IE may use service

registries other than the JISC IESR to discover services; JISC service providers may disclose their services through other service registries; and service registries themselves may exchange data. In such a context, the provision of a standard interface to the IESR becomes significant: the developers of applications using multiple service registries do not expect to use a different interface specification to interact with each service registry, and there may be a growing requirement for the exchange of data between service registries themselves.

Summary

The IESR is a key component in JISC's efforts to enhance access to a set of distributed collections of heterogeneous content. The development work to date has built on previous work on collection-level description, particularly the RSLP Collection Description Model and Schema, and has provided an important input to work on collection description within the Dublin Core Metadata Initiative. It has also highlighted the close relationship between collection description and service description. The take-up of the Web Services suite of standards and specifications within the JISC IE and more generally highlights the importance of ensuring that the interfaces provided by IESR are compatible with those provided by other services providing parallel functionality to other communities.

References

[1] *The Joint Information Systems Committee (JISC)*. HTML: <`http://www.jisc.ac.uk/`>

[2] Powell, Andy. *JISC Information Environment General FAQ*. April 2003.
HTML: <`http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/faq/general/`>

[3] Powell, Andy and Liz Lyon. *The DNER Technical Architecture: scoping the information environment.* May 2001.
HTML: <`http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/dner-arch.html`>

[4] Powell, Andy. *JISC Information Environment Technical Standards Framework*. Version 1.1. May 2004.
HTML: <`http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/standards/`>

[5] Powell, Andy. Mapping the Information Environment, *Ariadne* 36 (July 2003)
HTML: <`http://www.ariadne.ac.uk/issue36/powell/`>

[6] Powell, Andy. *JISC Information Environment Glossary*. May 2003.
HTML: <`http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/glossary/`>

[7] *JISC Information Environment Service Registry Pilot Projec*t
HTML: <`http://www.mimas.ac.uk/iesr/`>

[8] Heaney, Michael. *An Analytic Model of Collections and their Catalogues.* UKOLN/OCLC. Third edition, revised. January 2000. HTML: <`http://www.ukoln.ac.uk/metadata/rslp/model/`>

[9] *RSLP Collection Description Schema*. HTML: <`http://www.ukoln.ac.uk/metadata/rslp/schema/`>

[10] *IESR Metadata*. HTML: <`http://www.mimas.ac.uk/iesr/metadata/`>

[11] *Web Services Description Language (WSDL) 1.1*. W3C Note 15 March 2001.
HTML: <`http://www.w3.org/TR/wsdl`>

[12] *Zeerex:The Explainable "Explain" Service*. HTML: <`http://explain.z3950.org/`>

[13] Cheshire II Information Retrieval System. HTML: <`http://cheshire.lib.berkeley.edu/`>

[14] Powell, Andy and Liz Lyon. The JISC Information Environment and Web Services, *Ariadne* 31 (July 2003)
HTML: <`http://www.ariadne.ac.uk/issue31/information-environments/`>

[15] Gardner, Tracy. An Introduction to Web Services, *Ariadne* 29 (October 2001)
HTML: <`http://www.ariadne.ac.uk/issue29/gardner/`>

[16] *Universal Description, Discovery and Integration (UDDI)*
HTML: <`http://www.uddi.org/`>

[17] Matthew Dovey. *JISC Information Environment Service Level Descriptions* (2001)
HTML: <`http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/service-description/study/`>