Grant Agreement ECP-2007-DILI-527003

ARROW

State of the art and guidelines for standards applicable – Edition 2

Deliverable number/name	*D4.4*
Dissemination level	*Public*
Delivery date	*31 July 2010*
Status	*Final*
Author(s)	*EDItEUR/BnF*

e*Content*plus

This project is funded under the *e*Content*plus* programme[1],
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

[1] OJ L 79, 24.3.2005, p. 1.

Introduction

This document aims at providing a "state of art" description of all the technical standards and metadata formats which might have a specific application within the ARROW project. A first edition released in July 2009 was a prospective state of art, with a concise description of each standard in the ARROW context. This document provided the ARROW consortium with an overview of the different standards and technical solutions used in each domain.

At the time of preparing this second edition, the prototype of the ARROW system is now implemented in Germany, the UK and Spain.[2] Metadata formats and technical standards have been selected in order to define the workflow and implement the system. This second edition aims at highlighting the strengths and weaknesses of each standard in the ARROW context and its current use in the system.

The list considers a wide range of metadata, identification, messaging, search-related standards which are currently used by the libraries, books in print agencies or collecting societies. It also covers technical standards such as protocols and web services. The standards that are covered are either specified by official standardisation bodies or *de facto* standards -- that is, standards which have been implemented in the sector but have not necessarily received formal approval by way of a standardization process.

The initial selection of standards to be covered in the first edition was based on the results of a series of questionnaires which were conducted in 2009 with all the stakeholders of the book value chain of the European countries participating to the ARROW project.

For the second edition, a small number of new entries have been created at the request of ARROW partners, to fill gaps in the coverage; and a number of entries have been revised to bring them up to date. All the entries have been slightly reordered to highlight particularly the use of the standard in ARROW.

We have also restructured and simplified our classification scheme for the standards (see page 5). We hope this will provide simpler navigation. We have also included a guide to the meaning of each section in the directory (see page 7).

This edition, like the first, was compiled by EDItEUR (www.editeur.org), the international trade standards organisation for the book and serial supply chain, acting as a sub-contractor to the ARROW Project, under the overall guidance of the Bibliothèque Nationale de France, the ARROW partner responsible for the interoperability work stream in ARROW. Individual entries have been prepared by appropriate subject specialists, but are intended to be broadly comprehensible to the general reader. We are optimistic that this document will be useful beyond the ARROW project itself, particularly in any future project where significantly different stakeholder groups need to come together to create cross-domain technical solutions.

[2] Its implementation in France is in progress and should be completed in autumn 2010.

Contents – alphabetical (by acronym)

Contents – thematic

	ODRL	47
	ONIX-LT	52
	ONIX-PL	53
	ONIX-RP	54
Published content	EPUB	20
	MPEG21 DID	40
	PDF	57
Search	CQL	14
	GILS	23
	MXG	43
	OpenSearch	55
	Topic Maps	67
	Z39.50	75
Technical protocol	FTP	22
	MIME	38
	OAI-PMH	45
	RDF	61
	REST	62
	Schematron	63
	SOAP	64
	SRU/SRW	65
	TLS/SSL	66
	WS	71
	XML	72
	XSL, XSLT, XSL-FO	73
	XSD, WXS	74

Guide to Entries

Name	The full name of the standard		
ARROW type	The classification of the standard following an ARROW typology	***Use in ARROW***	Whether and how the standard is used in the ARROW project
Acronym	The acronym used for the standard – usually the way the standard is most commonly referred to.	***Reference***	If a formal standard, its designation within ISO (or similar)
Governance	The organisation responsible for the standard		
URL	A link to web page with more information		
Status	Publication status	***Implementation***	How widely the standard is implemented.
Availability	Where and how the standard itself can be obtained (and what it costs if relevant) as well as additional information about the underlying infrastructure for assignment of identifiers where relevant.		
Description	A brief description of the standard, what it is used for, and where appropriate cross references to other standards covered in the document.		
Rights coverage	Since ARROW is specifically concerned with the management of rights, a note on the extent to which the standard has implications for rights management.		
Strengths	A brief statement of the strengths of the standard from an ARROW perspective.		
Weaknesses	A brief statement of the weaknesses of the standard from an ARROW perspective.		

The Standards Map

Name	Anglo-American Cataloguing Rules		
ARROW type	Metadata - library	***Use in ARROW***	Implicit in many MARC library catalogue records
Acronym	AACR2	***Reference***	
Governance	AACR Committee of Principals		
URL	http://www.aacr2.org/		
Status	AACR, Second Edition (1978), updated 2005	***Implementation***	Widespread use in cataloguing, especially in the US, UK and Canada.
Availability	The Anglo-American Cataloguing Rules, Second Edition, available for purchase in print form. Integrated into the Library of Congress' online Cataloguers Desktop tool.		
Description	The Anglo-American Cataloguing Rules (AACR) are jointly published by the professional library associations in the United States, Canada, and the United Kingdom. The Second Edition (AACR2), published in 1978, has subsequently been slightly revised several times, most recently in 2005. AACR provides the basic rules that have been used in cataloguing library materials for over forty years. The rules are "designed for use in the construction of catalogues and other lists in general libraries of all sizes. … The rules cover the description of, and the provision of access points for, all library materials commonly collected at the present time." AACR Part I deals with the provision of information describing the item being catalogued, and Part II deals with the determination and establishment of headings (access points) under which the descriptive information is to be presented to catalogue users, and with the making of references to those headings. In both parts the rules proceed from the general to the specific. A sweeping revision is underway, under the auspices of the Joint Steering Committee for the Development of RDA. RDA: Resource Description and Access was released in June 2010. See also: RDA, MARC21		
Rights coverage	N/A		
Strengths	Very widely used in the English-speaking world to define the content of catalogue records; provides uniformity of records designed to make it easier for library users to access precisely the information they seek.		
Weaknesses	May be coming towards the end of its widespread application with the development of RDA (although it will take a very long time for existing practice to be replaced)		

Name	Automated Content Access Protocol		
ARROW type	Metadata - rights/permissions	***Use in ARROW***	Not used (not currently relevant)
Acronym	ACAP	***Reference***	
Governance	Project, managed and financed by WAN, EPC and IPA		
URL	www.the-acap.org		
Status	v1.0 published Nov 2007, v2.0 in development	***Implementation***	2000+ websites (but see below)
Availability	Specification freely available from the website. No licence required for implementation.		
Description	The ACAP project was originally launched in response to publishers' concerns about "the search engine problem" - search engines being seen as monetising copyright content without the permission of the owners (and without a commensurate flow of value to the copyright owner). The ultimate scope of the project is to provide the necessary support for any onlune business model where automated (machine-to-machine) communication of permissions is required (particularly in business-to-business relationships). ACAP's initial implementation (to meet search engine requirements) is expressed as an extension to the Robots Exclusion Protocol; however, this is seen as a temporary solution until a more satisfactory and generally acceptable long term communication format is agreed. Perhaps inevitably, while a large number of publishers have now undertaken (very simple) implementations of ACAP on their websites, the search engines have proved to be resistant to any new approach to managing copyright on the network. As a result, ACAP's focus since the end of 2007 has been turned in the direction of public affairs (specifically making the case for copyright on the network and the need for technological tools to support it). This campaign continues to gather support in the light of growing challenges particularly for the news sector. ACAP commissioned EDItEUR to provide the semantics for ACAP v1.0, ensuring compatibility with the ONIX for Licensing Terms framework. v2.0, now in development, allows machine-readable permissions to be embedded in web-syndicated content.		
Rights coverage	ACAP's sole purpose is to associate machine-readable permissions with online resources. Currently elaborated specifically for communicating in the crawling/indexing environment, ACAP is extensible to any other application requiring the same type of capability. (see note above on v2.0).		
Strengths	A well understood brand position, both within publishing and increasingly among governments and regulators; has uniquely drawn together all sectors of publishing into a single project (with good contacts with other media). Flexible and extensible to any machine-to-machine permissions. Interoperable with equivalent ONIX semantics.		
Weaknesses	Currently developed only for limited crawl/index/display Use Cases, specific to search applications. Although proved to work in a technical pilot, not yet implemented in a live application. Still in project mode, with no long term governance structure yet in place.		

Name	Archival Resource Key		
ARROW type	Identification	***Use in ARROW***	A potential alternative manifestation identifier if no ISBN available
Acronym	ARK	***Reference***	
Governance	California Digital Library (University of California)		
URL	http://www.cdlib.org/inside/diglib/ark/arkspec.html		
Status	Open specification and IETF draft RFC (2008)	***Implementation***	Limited (27 organisations, not all active)
Availability	Scheme maintained at California Digital Library. No fee is involved. Any institution can become an assigner authority by contacting CDL and can then generate ARKs; CDL uses open-source software called 'noid' (nice opaque identifiers).		
Description	An ARK is a persistent identifier, currently expressed as a URL in a specific scheme. Referents may be digital, physical, or abstract. ARKs are intended to work with objects that last longer than the organizations that provide services for them. Neither the web server itself nor the current web protocols are expected to last longer than the identified objects. The key points are two-fold: a) Associated metadata and persistence commitment. ARKs resolve (using a standard query specification) to three things: (1) A digital object (e.g. a content object which forms part of a digital archive); (2) Metadata about that digital object; (3) A commitment statement by the provider. Metadata is not required to be in any particular scheme or precise; the commitment statement is a free text statement. b) ARK as a globally unique identifier can be represented in various ways: it is assigned by one body (Name Assigning Authority) but can be used by several (Name Mapping Authorities, which are “mutable and replaceable”): so the reference http://bnf.fr/ark:/13030/tf5p30086k might become the reference http://portico.org/ark:/13030/tf5p30086k (the identifier ark:/13030/tf5p30086k remains the same). Also defined in several versions (latest v15 (2008) of an IETF draft (http://tools.ietf.org/id/draft-kunze-ark-15.txt), but the CDL documentation is more complete and the RFC is not heavily quoted, so essentially ARK is an open but non-standardised tool. Assigners are mostly American libraries (NLM, Library of Congress and several leading university and digital libraries). Bibliothèque Nationale de France has used the ARK scheme since 2006 in public applications and back office systems; the only UK organisations represented are the DCC and the Centre for Ecology and Hydrology.		
Rights coverage	None in the specification, though "Metadata" could include any item of data.		
Strengths	Application in major archival institutions; persistence. ARK identifiers can be extended with qualifiers in order to manage granularity and versioning of digital objects		
Weaknesses	Lack of effective standardization in application		

Name	MIME-based Secure Peer-to-Peer Business Data Interchange using HTTP, Applicability Statement 2		
ARROW type	Commercial messaging	***Use in ARROW***	Not used (not relevant)
Acronym	AS2	***Reference***	IETF RFC 4130
Governance	Internet Society / Internet Engineering Task Force		
URL	http://tools.ietf.org/html/rfc4130		
Status	Proposed Standard	***Implementation***	N/A
Availability	All IETF RFCs ("Requests for Comment") are freely available.		
Description	IETF RFC 2026 defines an "Applicability Statement" as a specification of how existing Internet Standards can be applied to meet a specific business need. Two Applicability Statements have been published so far. The first (IETF RFC 3335, more commonly referred to as AS1) describes MIME-based secure peer-to-peer business data interchange using SMTP, the widely-used email protocol. AS2 addresses the same business need but using the web (HTTP) instead of email. Both AS1 and AS2 specify how to perform EDI transactions over the Internet instead of over proprietary Value-Added Networks (VANs). They show how existing Internet Standards can be employed to achieve similar security of transmission as is achieved using conventional EDI. The payload format is not specified as such, but can be the same as for conventional EDI (typically EDIFACT or ANSI X12) or any alternative (e.g. XML-based). AS2 is proving popular with businesses that already employ conventional EDI and wish to switch from the use of high-cost VAN infrastructures to the use of low-cost Internet and web-based infrastructures. AS2 is seen as a less costly and less disruptive replacement for conventional EDI than more radical alternatives such as ebXML and web services. AS2 therefore tends to appeal to larger trading entities, such as major manufacturers and retailers, with large customer networks, whereas web services appeals more to smaller businesses that were not traditional EDI users. AS2 is particularly popular in North America, while ebXML (an XML-based alternative to EDIFACT) is more popular in Europe and Asia.		
Rights coverage	AS2 does not specify the content of transaction payloads, and so does not specifically cover the communication of rights information.		
Strengths			
Weaknesses	Not relevant to ARROW		

Name	Creative Commons		
ARROW type	Metadata – rights/permissions	***Use in ARROW***	Not used (not currently relevant)
Acronym	CC	***Reference***	
Governance	Creative Commons Corporation, a US non-profit company		
URL	http://creativecommons.org/		
Status	A range of 6 standard CC licences available since 2002	***Implementation***	Over 100 million documents were available under CC licences in 2008
Availability	CC licences are freely available		
Description	In its own words, "Creative Commons defines the spectrum of possibilities between full copyright and the public domain. From all rights reserved to no rights reserved. Our licenses help you keep your copyright while allowing certain uses of your work — a "some rights reserved" copyright. Creative Commons licenses are not an alternative to copyright. They work alongside copyright, so you can modify your copyright terms to best suit your needs. We've collaborated with intellectual property experts all around the world to ensure that our licenses work globally." Creative Commons offers any user a choice of six licences which are formed from a combination of 4 simple licence conditions: • *Attribution*: You let others copy, distribute, display, and perform your copyrighted work — and derivative works based upon it — but only if they give credit the way you request. • *Share alike*: You allow others to distribute derivative works only under a license identical to the license that governs your work • *Non-commercial*: You let others copy, distribute, display, and perform your work — and derivative works based upon it — but for non-commercial purposes only. • *No derivative works*: You let others copy, distribute, display, and perform only verbatim copies of your work, not derivative works based upon it. Licenses can be represented both by human readable icons, and by machine readable licence expressions. Localised versions of the CC licences have been developed in a number of different jurisdictions.		
Rights coverage	Firmly based in copyright, but limited in terms of available licence conditions and not normally considered suitable for use in commercial licensing.		
Strengths	The extensive implementation of CC licences is a strong indication of their value on the network to those who wish to publish content non-commercially. Standardisation brings considerable simplification for licensees and licensors alike.		
Weaknesses	CC licences have been criticised both for being too restrictive and for being too permissive – but these weaknesses are entirely contextual.		

Name	CIDOC Conceptual Reference Model (CRM)		
ARROW type	Conceptual model	***Use in ARROW***	Not used (not applicable)
Acronym	CIDOC-CRM	***Reference***	ISO 21127
Governance	ISO		
URL	http://cidoc-crm.org		
Status	ISO standard (2006)	***Implementation***	
Availability	As published ISO standard and in an update version available on the Web.		

Description	The CIDOC Conceptual Reference Model (CRM) provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation. The CIDOC CRM is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to. It is intended to be a common language for domain experts and implementers to formulate requirements for information systems and to serve as a guide for good practice of conceptual modelling. In this way, it can provide the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives. Sharing much of the same logical analysis as FRBR, since 2003 there has been an international committee working on expressing the IFLA FRBR model with the concepts, tools, mechanisms, and notation conventions provided by the CIDOC CRM, and aligning (possibly even merging) the two object-oriented models with the aim of contributing to the solution of the problem of semantic interoperability between the documentation structures used for library and museum information. In practice this has led to the development of FRBRoo, a formal ontology for FRBR expressed as an extension of the CIDOC CRM. This ontology was released in version 1.0.1 in January 2010.
Rights coverage	'Right' is declared as a class in the model and can be related to both 'Actor' and 'Legal Object'; however, the introduction states explicitly that 'of necessity, some concepts covered by the CRM are less thoroughly elaborated than others: E39 Actor and E30 Right, for example. [...] These "underdeveloped" concepts can be considered as hooks for compatible extensions.'
Strengths	A strong reference model
Weaknesses	Not directly relevant from an ARROW standpoint

Name	Contextual Query Language		
ARROW type	Search	***Use in ARROW***	Not used (federated search not implemented in ARROW)
Acronym	CQL	***Reference***	
Governance	The Library of Congress		
URL	http://www.loc.gov/standards/sru/sru1-1archive/cql.html		
Status	v1.2 published Nov	***Implementation***	
Availability	Specification freely available from the website. No licence required for implementation. Different implementations are available for free downloading for different programming languages such as java, perl, ruby.		
Description	The Contextual Query Language is the underlying query syntax used by SRU/W protocol.. CQL is a formal language for representing queries to information retrieval systems such as web indexes, bibliographic catalogs, etc. The design objective is that queries be human readable and writable, and that language be intuitive while maintaining the expressiveness of more complex languages. So it proposes as a powerful and expressive language such as expert ones (e.g. SQL, XQuery, etc.) and simple and intuitive to express concepts such as not powerful ones (e.g. CCL and Google). CQL is so-named "Context Query Language" because it is founded on the concept of searching by semantics or context, rather than by syntax. The same search may be performed in a different way on very different underlying data structures on different servers; the important thing is that both servers understand the intent behind the query. In order for multiple communities to define their own semantics, CQL uses Context Sets to ensure cross-domain interoperability. Context sets permit users to create, for example, their own indexes without fear of choosing the same name as someone else and thereby having an ambiguous query. CQL is based on the definition of a set of abstract access points, such as title, author, subject and refinements of those such as personal author, uniform title, geographical subject. While large data bases generally have some form of indexing structure associated with them, and the abstract access points of the CQL are often called abstract “indexes”, CQL does not actually mandate the existence of “physical” indexes at the target but the ability to retrieve as if there were. CQL does not make presumptions about the database design but it is biased toward searching metadata that is identified (i.e., records as data rather than as documents) to enable “smart” searching. A server can claim different level of conformance to CQL. A higher level corresponds to greater expressiveness.		
Rights coverage	N/A		
Strengths	Expressiveness and powerful. It provides the ability of contextualizing search indexes (e.g. namespace)		
Weaknesses	Not relevant to ARROW		

Name	CrossRef		
ARROW type	Identification	***Use in ARROW***	Not used (heavily implemented for serials but limited in books)
Acronym	CrossRef	***Reference***	n/a
Governance	Publishers International Linking Assocation, Inc. (an application of DOI)		
URL	http://www.crossref.org		
Status	DOI ISO standardisation now at FDIS stage. De facto standard since 2000.	***Implementation***	2,800 publishers, >20,000 journals; Assigned to a large number (~35M) of articles including back files, and carried in many A&I services. Individual access to existing information is free (both from bibliographic data or from DOI); access to full text may require appropriate permissions. Larger scale commercial services using the database are available for libraries and others (e.g. indexing services).
Availability	Registration of a DOI requires membership of CrossRef.		
Description	CrossRef (a DOI implementation, and DOI registration agency) is a cross-publisher citation linking system, which assigns DOIs to scholarly articles (and increasingly, other related materials such as books, conference proceedings etc) and so processes citations (pre-or post-publication) to populate a reference list with persistent citation links. Publishers may use any format for their identifier (SICI, PII, ISSN-based, private etc) which then forms the suffix of a DOI. CrossRef prescribes a metadata scheme to facilitate look up services, and maintains a number of services for affiliate libraries. CrossRef also works closely with library link resolvers, both commercial (eg Ex-Libris) and non-commercial, to offer linkage to an appropriate (allowed) copy through e.g. OpenURL services. Additional services of interest to the CrossRef community continue to be developed (e.g. plagiarism detection, "CrossCheck"); or considered (e.g. author/institution identification).		
Rights coverage	Users click on a reference citation (a DOI) on one publisher's platform and link directly to the cited content on another publisher's platform, subject to the target publisher's access control practices. Also works with OpenURL. CrossRef is run by publishers.		
Strengths	Very widely implemented to identify current and backlist content in the scientific and technical publishing communities.		
Weaknesses	At this point, limited application to books – primarily used for journal articles.		

Name	Dublin Core Metadata Element Set		
ARROW type	Metadata - generic	***Use in ARROW***	Not used (not appropriate for ARROW because of requirement for automated processing)
Acronym	DCMES	***Reference***	ISO 15836
Governance	ISO/TC46/SC4; Dublin Core Metadata Initiative (DCMI) Executive, Oversight Committee and Usage Board		
URL	http://dublincore.org/		
Status	International Standard revised 2009	**Implementation**	Widespread loose adherence
Availability	Specifications freely downloadable from the DCMI web site; no licence required. Various tools freely available online, with a subset of the Elements reasonably widely deployed in describing online and offline resources.		
Description	The Dublin Core Metadata Element Set (DCMES) is the principal deliverable of the long-running Dublin Core Metadata Initiative (DCMI), which evolved from a series of workshops begun in 1995. Intended to provide straightforward and domain-neutral descriptions of both digital and physical resources, the fifteen elements of the DCMES are supplemented by substantial additional guidance, terminology and modelling work. An active and open international community continues to support various public email lists, working groups and an annual conference. Originally intended as a relatively simple 'pidgin' capable of supporting both the creation of new resource descriptions and providing some degree of interoperability between richer metadata standards within specific domains, DCMES has subsequently become more complex as diverse communities of interest have sought to extend it for their own ends. Recently revised work in the area of Application Profiles sees the Initiative grappling once more with the tension between domain richness and global interoperability, "by providing a framework for designing a Dublin Core Application Profile [that] defines metadata records which meet specific application needs while providing semantic interoperability with other applications on the basis of globally defined vocabularies and models." This structure of metadata is required when exchanging metadata in OAI-PMH. See also: OAI-PHM		
Rights coverage	'Rights' is one of the 15 elements of the Dublin Core, although somewhat loosely defined in most generic applications; "Typically, rights information includes a statement about various property rights associated with the resource, including intellectual property rights."		
Strengths	Widely implemented and simple to use.		
Weaknesses	Very "loose adherence" in application implies that the value of metadata in a DC record is inevitably limited. Most appropriate for human interpretation and where no other metadata record is available.		

Name	Digital Object Identifier		
ARROW type	Identification	***Use in ARROW***	Not used (see CrossRef and ISBN-A)
Acronym	DOI	***Reference***	ISO FDIS 26324
Governance	International DOI Foundation, Inc (USA)		
URL	http://www.doi.org		
Status	Final Draft ISO (2010); active implementation since 2000	***Implementation***	International; ~50 m assigned
Availability	Through a federation of appointed DOI Registration Agencies (RAs), under policies and common infrastructure provided by the International DOI Foundation. Used in a range of publishing applications since 2000 (mainly bibliographic sector and professional level, eg CrossRef); by early 2009 c 40 million assigned. Requires assignment by RAs and a fee to support a service using the DOI System. Fees are set independently by each individual Registration Agency.		
Description	Provides a specification and implementation of assignment, syntax, metadata and resolution to provide persistent, actionable, semantically interoperable identification of any entity (physical, digital or abstract) on digital networks. Includes a social component (RA federation) to ensure persistence and consistency. A DOI name identifies an object as a first-class entity, not simply the place where the object is located, and can be associated with defined services on a network. Initial implementations of redirection to a single URL are now being supplemented by functionalities of multiple linkage and structured metadata models. Uses two underlying technologies (Handle System and the indecs content model) and inherits the features and capabilities of each. Applicable to any entity; main applications to date are to bibliographic or data resources. In principle independent of any technology, but currently widely used with Web http. Specifications also exist for incorporation of other identifier schemes into the DOI system, e.g. ISBN-A.		
Rights coverage	The system is independent of specific business model or legal framework, but is designed with applications to current content businesses in mind, and requires that assigners conform to relevant copyright legislation. IDF encourages rights applications.		
Strengths	Very broadly applicable identifier; considerable potential in multiple resolution capability provided by Handle technology.		
Weaknesses	Not widely implemented outside CrossRef application, although new applications (including for example DataCite (http://www.datacite.org/) now developing rapidly. Lack of widespread native support for Handle in internet applications. DOI may not pass the final stage of ISO standardisation, but this is not very significant because the DOI is already very widely implemented. The cost of establishing DOI Registration Agencies (and subsequently the cost of implementing DOI) has been seen as a barrier in some markets.		

Name	Electronic Data Interchange For Administration, Commerce and Transport		
ARROW type	Commercial messaging	***Use in ARROW***	Not used (not relevant)
Acronym	EDIFACT	***Reference***	ISO 9735
Governance	United Nations Economic Commission for Europe (UNECE) on behalf of the United Nations Center for Trade Facilitation and Electronic Business (UN/CEFACT) ISO/TC154		
URL	http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=35032		
Status	International Standard	***Implementation***	Very widely implemented
Availability	All Parts of ISO 9735 are available for purchase from ISO, Geneva, and from many national standards bodies that have transposed them as national standards.		
Description	EDIFACT is a ten-part International Standard, published first in 1988, updated in 1998 and further revised for a Second Edition published in 2002. The standard defines: (a) syntax rules for the construction of EDI messages, which can be exchanged in either a batch or interactive mode; (b) a protocol (I-EDI) for interactive message exchange; (c) a set of standard message formats. Several organisations have developed profiles of subsets of the standard message set for different applications, including the EANCOM set published by GS1, which has in its turn been profiled for use in specific trading sectors. EDItEUR first developed its EDIFACT message set for use in the publishing sector from the EANCOM set in 1996. EDIFACT defines a compact syntax in which the body of a message is composed of data segments, each of which is composed of one or more component data elements. Strict syntax rules make it possible to compress a message so that it is typically roughly one-tenth the size of a comparable message in XML syntax. There are two syntax levels: level A uses entirely plain text characters, while level B uses three non-printing control characters as separators. EDIFACT has been widely adopted in Europe, but less widely in Asia, where electronic commerce became established more recently and the availability of XML alternatives to EDIFACT has had more impact. EDIFACT has had little impact in North America, where ANSI X12 has dominated the market for EDI standards.		
Rights coverage	The standard EDIFACT message set does not include specific support for communication of rights-related information. There is the capability to include "associated data" (Part 8), but there is no known use of this capability for embedding rights-related information in EDFACT messages.		
Strengths	Very widespread use in commercial applications, including in libraries		
Weaknesses	With the development of the Internet, the use of Value Added Networks (once the sole medium for EDI messages) has more-or-less disappeared, with EDI migrating to the Internet. At the same time, the requirement for compactness has become less of an issue, and the greater expressiveness of XML formats for unambiguous communication of transactional messages is likely to supplant traditional EDI formats. However, this will be a slow process, in view of the fact that conventional EDI is widely implemented and fulfils its requirements very effectively.		

Name	EDItEUR XML Document Formats		
ARROW type	Commercial messaging	***Use in ARROW***	Not used (not relevant)
Acronym	EDItX	***Reference***	N/A
Governance	EDItEUR		
URL	http://www.editeur.org/		
Status	Published trade standard	***Implementation***	Limited (AU, DE, GB, SE)
Availability	The EDItX specifications and schemas are freely available from the EDItEUR website.		
Description	EDItX is a family of XML transaction message formats designed by EDItEUR with two primary objectives: (a) to provide an alternative to traditional EDI for sections of the trade and library book supply chains that have not implemented traditional EDI and might prefer to implement XML-based messaging; (b) to satisfy business requirements in those supply chains that have not been met and are unlikely ever to be met by established EDI solutions. The first EDItX formats were published in 2004, but only included a small subset of the message formats needed for transactions in regular use. A more complete set for trade book supply was published in 2007, with additional formats for library book supply being published in 2008. EDItX formats follow design rules derived in part from the traditional EDI formats in use in the book trade, and in part from XML/EDI formats developed for use in other business sectors. Unlike traditional EDI, EDItX formats are designed specifically to meet the needs of the book trade, as is evident in the naming conventions applied both to tag names and to code values. Naming conventions in EDItX are generally verbose, to aid human readability. The main implementations of EDItX are in the German and Swedish book trade. In the UK several of the EDItX formats have been used as a basis for the development of web service standards for the book trade, published by Book Industry Communication. EDItX formats are not related to XML-based formats used in other business sectors. This has not been an issue for a trade that historically has been largely self-contained, but may become an issue as more book products are supplied into the general retail sector.		
Rights coverage	None of the existing EDItX formats support the inclusion of rich bibliographic or market supply data, which is generally where rights-related information is to be found. Such information is typically communicated using the ONIX Book Product Information message format, another EDItEUR format but not part of the EDItX family.		
Strengths	A growing suite of transactional messages designed to support specific book trade requirements; new messages continue to be developed and existing messages modified to support specific requirements as these are identified (particularly in support of digital publishing).		
Weaknesses	Still only patchy implementation, as conventional EDI messages (see EDIFACT) remain largely “fit for purpose”. EDItX is specific to the book trade and not related to other XML EDI standards.		

Name	EPUB		
ARROW type	Published content	***Use in ARROW***	Not used (not relevant)
Acronym	epub	***Reference***	
Governance	International Digital Publishing Forum (IDPF)		
URL	http://www.idpf.org/specs.htm		
Status	Published November 2007; revision currently in progress	***Implementation***	Gaining ground rapidly
Availability	Specifications freely available from the IDPF website; no licence required for implementation		
Description	".epub" is the file extension of an XML format for reflowable digital books and publications. ".epub" is composed of three open standards, the Open Publication Structure (OPS), Open Packaging Format (OPF) and Open Container Format (OCF), produced by the IDPF. ".epub" allows publishers to produce and send a single digital publication file through distribution and offers consumers interoperability between software/hardware for unencrypted reflowable digital books and other publications. The Open eBook Publication Structure or "OEB", originally produced in 1999, is the precursor to OPS. A growing number of ebook platforms support the EPUB format; however, in practice, many platforms continue to use proprietary formats, meaning that publishers continue to have to support many and different XML schemes for different platforms.		
Rights coverage	The EPUB package has a metadata structure which has the capacity to carry rights information		
Strengths	A standard format has been long sought for the publication of ebooks, and EPUB provides the essential kernel on which a more comprehensive standard can emerge over time.		
Weaknesses	Apart from the fact that not all platforms support EPUB, the standard as published supports only relatively straightforward publications (layout etc). As a result the standard is being developed very rapidly and without a great deal of formality, as the ebook market itself develops. This is probably inevitable, and is clearly in the interests of the major users of the standard; however, from the point of view of an outsider, it might appear a point of weakness.		

Name	Functional Requirements for Bibliographic Records		
ARROW type	Conceptual model	***Use in ARROW***	Cannot be implemented directly; however FRBR has had a considerable influence on thinking both in and beyond the library community.
Acronym	FRBR	***Reference***	
Governance	IFLA's FRBR Review Group		
URL	http://www.ifla.org/en/frbr-rg		
Status	1998 report of an IFLA Study Group. No formal status, although much cited.	***Implementation***	This report has informed discussion and debate for the past decade, most notably influencing the RDA standardisation effort.
Availability	Functional Requirements for Bibliographic Records is freely available for download from the IFLA web site; http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records		
Description	Developed by a working group of the International Federation of Library Associations and Institutions (IFLA) in the late Nineties, the Functional Requirements for Bibliographic Records (FRBR) defines a conceptual model to describe interactions with bibliographic systems and data from the perspective of a user. FRBR is independent of formal cataloguing standards such as the Anglo-American Cataloguing Rules (AACR), and has influenced recent efforts such as RDA that attempt to re-imagine these rules. At the heart of FRBR lie a number of key 'entities,' and the relationships between these are key to FRBR's view of the world. Creations, for example, are classed as being Works, Expressions, Manifestations or Items; a set of classifications spanning everything from the creator's original concept to a specific example of the final work such as one particular copy of a print run of an edition of a book. "FRBR may serve as a reference point for testing the validity and robustness of extant [bibliographic] data models and data structures. It can therefore be used to improve extant formats as well as to provide guidance for the process of developing new formats. It also can be extremely valuable in helping design OPACs." The FRBRoo (FRBR-object oriented) initiative is a joint effort the CIDOC CRM and FRBR international working groups to establish "a formal ontology intended to capture and represent the underlying semantics of bibliographic information and to facilitate the integration, mediation, and interchange of bibliographic and museum information." The most recent version (1.0.1) was published in January 2010. See also: CIDOC Conceptual Reference Model (CRM) and RDA		
Rights coverage	n/a		
Strengths	A very influential conceptual model, which is informing a great deal of library thinking about the future of cataloguing. Work has been done to make compatible with CIDOC; also has many features in common with the <indecs> model, with which it is contemporaneous.		
Weaknesses	A conceptual model – requires interpretation and reification.		

Name	File Transfer Protocol		
ARROW type	Technical protocol	***Use in ARROW***	Not used (web services preferred)
Acronym	FTP	***Reference***	IETF RFC 959
Governance	Internet Society / Internet Engineering Task Force		
URL	http://tools.ietf.org/html/rfc959		
Status	Standard	***Implementation***	Very widely used
Availability	All IETF RFCs ("Requests for Comment") are freely available.		
Description	The origins of FTP are in RFC 114, published in 1971. The first stable version of FTP was published in 1980 (RFC 765), and this was replaced by the current standard in 1985. FTP is a protocol for the exchange of files between a user and a server connected via the Internet. Unusually, the protocol involves two types of connection being made at the same time: a control connection and a data connection. The control connection is for the exchange of request and response text messages between user and server, while the data connection is for the exchange of file data. No data can be transferred without a control connection being established and maintained throughout the session. An FTP server may require user authentication or may allow connection by "anonymous" users. Once connected, and depending upon what the server will allow the user to do, a user may request the server to perform a range of simple directory and file management tasks in addition to file transfer. FTP servers vary in their capabilities, but as a minimum will enable files to be transferred between user and server. FTP does not use any form of encryption for either the control or data channels, which makes the protocol inherently insecure. Several attempts have been made to develop secure forms of the protocol (e.g. FTP over SSH, SFTP, Secure Copy), but none has become well established. The two-channel nature of FTP makes "tunnelling" through secure transport layers particularly problematic. As a result, other protocols, such as HTTPS, tend to be used for secure file transfer, while other, lower-layer techniques (e.g. VPN) are employed for more general remote directory and file management over secure connections.		
Rights coverage	The only rights with which FTP is concerned are user access rights on the FTP server. Users are authenticated by username and password. The rights are determined by data stored on the server for each known user's account.		
Strengths	Very widely implemented		
Weaknesses	Lack of security		

Name	Government Information Locator Service (also known as Global Information Locator Service)		
ARROW type	Search	***Use in ARROW***	Not used (distributed search not implemented)
Acronym	GILS	***Reference***	FIPS 192-1
Governance	US Government Printing Office (GPO)		
URL	http://www.gils.net/		
Status		***Implementation***	Widespread use by US Federal and State Agencies
Availability	Documentation and specifications notionally freely available via http://www.gils.net/		
Description	GILS was intended as a gateway to State and Federal Government information (predominantly in the USA), made possible by widespread deployment of a specific GILS profile to the Z39.50 protocol for Search and Retrieval of information. Making use of the federated nature of a Z39.50 Search, GILS made it possible for information to be curated and disseminated at the level of individual administrative units, yet cost-effectively surfaced in searches across different Government systems. Although GILS systems such as http://www.gpoaccess.gov/gils/index.html remain operational, the effort has assumed a far lower profile in the face of very different programmes for providing access to Government information such as http://www.usa.gov/. A number of State and Federal bodies continue to automatically expose GILS metadata and support Z39.50 queries of their systems via the GILS profile, but elsewhere the data are becoming increasingly stale. To all intents and purposes, GILS would appear to be moribund. See also: Z39.50		
Rights coverage			
Strengths	Some continuing use in the US		
Weaknesses	Limited application to Government information. Apparently moribund		

Name	Interoperability of Data in Electronic Commerce Systems		
ARROW type	Conceptual model	***Use in ARROW***	Cannot be implemented directly; however <indecs> has had a considerable influence on metadata and identifier design in the commercial sector
Acronym	<indecs>	***Reference***	
Governance	<indecs> was a project and has no continuing governance		
URL	http://www.doi.org/factsheets/indecs_factsheet.html		
Status	Project complete in 2000	***Implementation***	Not applicable
Availability	The primary <indecs> deliverable --"Principles, model and data dictionary" -- can be freely downloaded from the International DOI Foundation website (see above).		
Description	<indecs> was a project part funded by the EC Info 2000 initiative and by several organisations representing the music, rights, text publishing, authors, library and other sectors in 1998-2000, it has since informed a number of metadata activities. <indecs> provided an analysis of the requirements for metadata for e-commerce of "content" (intellectual property) in the network environment, focussing on semantic interoperability. The analysis was based on a simple generic model of commerce (the "model of making"): a model of the life cycle of any kind of content or intellectual property from conception to the final physical or digital copies. Central to the analysis is the assumption that it is possible to produce a generic mechanism to handle complex metadata for all different types of content. <indecs> proposed four basic principles: • The principle of Unique Identification: every entity should be uniquely identified within an identified namespace. • The principle of Functional Granularity: it should be possible to identify an entity whenever it needs to be distinguished • The principle of Designated Authority: the author of an item of metadata should be securely identified. • The principle of Appropriate Access: everyone requires access to the metadata on which they depend, and privacy and confidentiality for their own metadata from those who are not dependent on it		
Rights coverage	<indecs> recognised that, once you are trading content on the network "every transaction is a rights transaction". Its "event based" analysis of rights and permissions has been the basis of substantive standards developments, including the ONIX family (and particularly the ONIX for Licensing Terms) and the DDEX standards used in the music industry.		
Strengths	An extremely influential analysis, on which a great deal of subsequent metadata development has depended. Still relevant 10 years on.Broadly contemporaneous with and consistent with FRBR.		
Weaknesses	Simply a conceptual model -- a considerable aid to understanding many of the issues involved in metadata creation and framework, but not in itself an "out of the box" solution to anything. As with FRBR – requires interpretation and reification. Some very minor differences with FRBR have sometime been interpreted as suggesting incompatibility between the two approaches.		

Name	Interested Party Identifier		
ARROW type	Identification	***Use in ARROW***	Not used; this is a private scheme and not available outside the rights societies – it is not widely implemented even in the IFRRO societies. However, it has informed aspects of the development of the ISNI
Acronym	IPI	***Reference***	
Governance	SUISA on behalf of CISAC		
URL	None		
Status	Internal standard	***Implementation***	Widespread among rights societies
Availability	Only available to members of CISAC		
Description	The purpose of the IPI system is the global unique identification of a natural person or a legal entity acting across multiple creation classes, roles and rights. The essential feature of the IPI system is that it distinguishes between an IP Base Number and an IP Name Number. An IP Base Number identifies an underlying entity. An IP Name Number identifies a name by which the entity is or has been known, or in which the entity participates together with others. The IPI system has been developed in the scope of the music industry to identify writers and publishers and their society of affiliation for different rights. The system supports the exchange of information between CISAC societies and aims to help improving the accuracy of information exchanged worldwide with user organisations such as radio and TV stations, and recording manufacturers. The IPI system holds a unique identifier assigned to each interested party and supporting metadata including: • Interested Party name (patronym of interested parties, modification references of interested parties, pseudonyms for natural persons and other references for legal entities) • Nationality (the linking of natural persons to countries) • Date (parameters for birth date, death date, etc) • Creation class (a class of products of human imagination and/or endeavour) • IPI right (combinations of creation classes and rights) • Membership agreement (agreements between IP's and their IPI administration societies) • Role (represents the roles of interested parties, or the functions played by interested parties) • Territory (territory of a membership agreement)		
Rights coverage	Not directly applicable, but the IPI's sole purpose is to facilitate collective rights management		
Strengths	Effective mechanism for identification of names and parties		
Weaknesses	Whatever the strengths of the IPI scheme, from an ARROW point of view its fundamental weakness – that it is not publicly available – disqualifies the IPI from consideration.		

Name	International Standard Book Number		
ARROW type	Identification	***Use in ARROW***	Identification of manifestations (where available)
Acronym	ISBN	***Reference***	ISO 2108
Governance	ISO TC46/SC9		
URL	http://www.isbn-international.org/		
Status	4th edition (2005)	***Implementation***	Nearly universal since 1970 in major international markets.
Availability	The standard is published by ISO. ISBNs are available through a network of 170 local agencies		
Description	The ISBN (International Standard Book Number) system was devised in the late 1960s. It is a unique machine-readable identification number, which marks any book unmistakably. The number has been in use now for 35 years and has revolutionised the international book-trade. 170 countries and territories are officially ISBN members. Since 1 January 2007 the number has consisted of thirteen digits. It is commonly represented in print using an EAN-13 Barcode. A different ISBN is supposed to be assigned to each edition of a book, although this is causing controversy in its application to ebooks. An ISBN can also be applied to fragments of books (eg individual chapters) for use in the supply chain.		
Rights coverage	Like all ISO TC/46/SC9 identifiers, ISBN is explicitly not an indicator of rights ownership.		
Strengths	Almost universal on books published since 1970. Well understood and completely embedded within the book trade. Arguably the most successful global supply chain identifier ever devised.		
Weaknesses	From an ARROW point of view, the fact that ISBNs are not associated with books published before 1970 is a weakness. It cannot be used to identify books published before that date (later in some countries). Although intended as a supply chain identifier, it has also been used extensively in other ways in publishers' and other systems, which has sometimes distorted its application and led to assignment errors. There are inevitably some concerns about its application to ebooks (where current practice, particularly in the United States, is inconsistent) and to digitizations of printed books (where policy remains unclear).		

Name	Actionable ISBN		
ARROW type	Identification	***Use in ARROW***	Not used (no active implementation)
Acronym	ISBN-A	***Reference***	
Governance	ISBN International and International DOI Foundation (joint agreement)		
URL	http://www.doi.org/factsheets/ISBN-A.html (an application of DOI)		
Status	Proposed (2008)	***Implementation***	Limited pilots completed. Initial implementation by mEDRA, the Multilingual European DOI Registration Agency (www.medra.org); further applications may follow.
Availability	Assigned by or on behalf of an ISBN agency. ISBN-As do not automatically exist for every ISBN; they exist only once an appropriate DOI agency has registered them in the DOI System. Several ISBN agencies are already also DOI Registration Agencies		
Description	A DOI implementation; a method for including an ISBN in a DOI syntax in a standard way. ISBN-A allows the ISBN to be expressed as a DOI and so take advantage of functionality such as multiple resolution. By definition, an ISBN-A identifies the same thing as the ISBN, and is assigned on behalf of the ISBN agency. Constructed by incorporating a complete 13 digit ISBN into the allowed DOI syntax: Example: 10.97812345/99990: Handle System DOI name prefix = "10." ISBN (GS1) Bookland prefix = "978." or "979." ISBN Publisher prefix = variable length numeric string of 2 to 8 digits Prefix/suffix divider = "/" ISBN Title enumerator and checkdigit = variable length numeric string of 8 to 2 digits (total length of "Bookland prefix", "ISBN publisher prefix" plus "ISBN Title enumerator and checkdigit" will always equal 13 digits. The check digit from the ISBN remains unchanged; the DOI system does not use check digits).		
Rights coverage	Same as ISBN, but capable of augmentation through additional services		
Strengths	Combines what is probably the most effective product identifier ever deployed with the multiple resolution capabilities of the DOI.		
Weaknesses	No active implementation		

Name	International Standard Name Identifier		
ARROW type	Identification	***Use in ARROW***	Not used (not yet implemented)
Acronym	ISNI	***Reference***	ISO 27729 (approved but not yet published)
Governance	ISO TC46/SC9. The ISNI International Agency, in the form of a consortium awaiting formal incorporation, will act as the Registration Authority for the identifier.		
URL	http://www.isni.org/		
Status	International Standard	***Implementation***	N/A
Availability	There will be multiple Registration Agencies; founding members of the consortium will initially contribute several million identities to the system (including rights managements organisations and libraries).		
Description	The scope of the International Standard Name Identifier is "the identification of Public Identities of parties: that is, the identities used publicly by parties involved throughout the media content industries in the creation, production, management, and content distribution chains." In other words, the ISNI identifies names rather than underlying parties (both individual and corporate). This approach means that only limited metadata must be made publicly available through the ISNI system (metadata which is already in the public realm) -- and confidentiality and privacy can be properly protected. Different names used by the same party (for example, pseudonyms) can be linked but only where the relationship is in the public realm (the most frequently quoted example is Ruth Rendell and Barbara Vine). According to the website "the ISNI has been designed as a “bridge” identifier, allowing various industry partners to exchange information relating to a Party without the need to disclose confidential information. To that extent the ISNI only maintains the minimum metadata set needed to differentiate (disambiguate) two Public Identities. All other relevant information remains in proprietary databases secured by conditional access." The ISNI is a 16 digit numeric string (the final digit is a check digit) ISNI 1422 4586 3573 0476. There have already been encouraging test results from sharing library (VIAF) and rights management information (ALCS, ADAMI) data.		
Rights coverage	The management of identity is critical to rights management, but there is no direct relationship between rights ownership and the ISNI. Rights management organisations are likely to be among the primary users.		
Strengths	A standard that is recognized as much needed in the media for the exchange of information. Strong support from a disparate community of interests (including libraries and publishers).		
Weaknesses	Not yet implemented; now likely to be available in 2011.		

Name	Information and documentation - Format for information exchange		
ARROW type	Metadata - library	***Use in ARROW***	Format which is the basis of the Machine Readable Cataloguing (MARC) format used for supporting exchange of bibliographic information in the library domain
Acronym	ISO 2709	***Reference***	ISO 2709
Governance	ISO TC 46 SC4		
URL	http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=41319		
Status	International Standard, revised 2008	***Implementation***	Reference standard for the structure of the Machine Readable Cataloguing (MARC) formats, the Common Communication Format (CCF), etc.
Availability	Available for purchase from ISO.		
Description	ISO 2709 specifies the requirements for a generalized exchange format which will hold records describing all forms of material capable of bibliographic description as well as other types of records. ISO 2709 describes a generalized structure, a framework designed specially for communications between data processing systems and not for use as a processing format within systems. According to this standard, the general structure of a record is the following: a label, a directory, variable fields (each field is composed of one three-digit tag, two one-digit indicators, subfields and field separator) and a record terminator. This standard does not define the length or the content of individual records and does not assign any meaning to tags, indicators or identifiers, these specifications being the functions of an implementation format (for example MARC21 or UNIMARC). See also: marcXchange, MARC 21, UNIMARC		
Rights coverage	N/A		
Strengths	ISO 2709 allows to manage variable length information without breaking, repeatable or non repeatable information, optional or mandatory information. Extremely widespread implementation, especially as MARC formats: billions of MARC records held by libraries worldwide with millions of new MARC records being created every year. While many reasons may exist for its replacement, in practice the embedded nature of MARC in library practice will mean that MARC records persist in use for a long time		
Weaknesses	This standard is technically dating back to a time when the data was physically encoded on library cards. Today, in the context of the web environment, as libraries become ever-more involved in partnerships with organisations very different from themselves marcXchange is a better option.		

Name	Information Technology – Metadata registries		
ARROW type	Metadata - generic	***Use in ARROW***	Not directly applicable
Acronym	ISO/IEC 11179	***Reference***	ISO/IEC 11179
Governance	ISO SC32 WG2		
URL	http://metadata-stds.org/11179/		
Status	ISO standard in six parts with different update cycles	***Implementation***	
Availability	The six separate sections of ISO 11179 are freely available for download from ISO; http://metadata-stds.org/11179/		
Description	The ISO 11179 set of standards provide guidance on defining and representing metadata in a consistent fashion, and underpin the formalisation process behind metadata specifications such as the Dublin Core. This consistency in definition is intended to aid the process of evaluating overlap between apparently similar metadata elements in different metadata schemas, and leads to less ambiguous definition than might otherwise be the case. ISO 11179 also describes 'the roles and requirements for the registration process in a metadata registry,' and there has been some enthusiasm for registering metadata specifications and 'application profiles' in conforming registries in order to reduce redundancy and encourage greater interoperability. Part 1 of ISO 11179 provides a 'Framework' explaining the scope and purpose of the standard, and placing the remaining five sections in context.		
Rights coverage	n/a		
Strengths			
Weaknesses	No direct application from the ARROW perspective		

Name	International Standard Text Code		
ARROW type	Identification	***Use in ARROW***	Not used (not yet implemented)
Acronym	ISTC	***Reference***	ISO 21047
Governance	ISO/TC46/SC9; The International ISTC Agency is not for profit company limited by guarantee and registered in England & Wales; it was formed by a consortium comprising Bowker, CISAC, IFRRO & Nielsen		
URL	http://www.istc-international.org/		
Status	Published 2009	***Implementation***	Pilot only
Availability	The standard itself is available from ISO (priced). Currently, as part of the pilot, ISTC Registration Agencies (Nielsen & Bowker) are issuing ISTCs without charge; long term commercial models and terms are not known. The recent licensing of a third Registration Agency – MvB – should aid implementation.		
Description	The International Standard Text Code (ISTC) system is a global identification system for textual works. It is intended for use by publishers, bibliographic services, retailers, libraries and rights management agencies. Each ISTC is a unique identifier assigned by a centralised registration system to a textual work, when a metadata record for that work is entered into the system. If another, identical (or near identical) metadata record has already been registered (perhaps, in the case of an out of copyright work, by another publisher), the system will assume the new ISTC request refers to the same work and will output the ISTC of the identical (or near identical) metadata record already held on the system. An ISTC does not “belong” to a single author/publisher; rather, it “belongs” to the work it identifies. This means that the same ISTC number should be used to identify the same content even when it is being published by a different publisher and/or in a different publication format. The ISTC is intended to allow both collocation and disambiguation of manifestations of textual works. This is the case even though some manifestations with different content might have very similar or even identical names, and even though some products containing the desired content have entirely different names. The ISTC is not intended for identifying manifestations of a textual work, including any physical products (e.g. a printed article) or electronic formats (e.g. an electronic book). Manifestations of textual works are the subject of separate identification systems.		
Rights coverage	"The allocation of an ISTC to a work shall have no meaning or value as legal evidence regarding the copyright status of, or any intellectual property rights in, the work."		
Strengths	When it comes to specialist textual work identifiers, ISTC is "the only show in town" (although other more generalised identifiers such as DOI or URI could be used in theory). Fulfils a critical requirement for different stakeholder groups.		
Weaknesses	As yet, little implemented. Requires a substantial incentive to see it widely implemented; it is possible that rights registries may create that incentive. However, there are some real challenges over granularity requirements in different stakeholder communities that could have an impact on how quickly the ISTC makes an impact.		

Name	Library of Congress number		
ARROW type	Identification	***Use in ARROW***	Not used (not relevant in Europe)
Acronym	LCCN	***Reference***	n/a
Governance	Library of Congress (USA)		
URL	http://www.loc.gov/marc/lccn_structure.html		
Status	De facto standard (since 1898)	***Implementation***	LoC and hence libraries worldwide
Availability	The Library of Congress assigns a number while a book is being cataloged but as there is a backlog in this process a number can be assigned before the book is published: a Preassigned Control Number (PCN) is a Library of Congress Control Number which has been assigned prior to the work's publication; this accompanies Cataloging-in-Publication (CIP) Data.		
Description	Library of Congress number (aka LCCN = Library of Congress Card Number = Library of Congress Control Number) is a unique identification number that the Library of Congress assigns to the catalog record created for each book in its cataloged collections, or expected to become so (see PCN above). Strictly speaking the LCCN is the control number for the bibliographic record, not the book. Librarians use it to locate a specific Library of Congress catalog record in the national databases and to order catalog cards from the Library of Congress or from commercial suppliers. The basic control number has fixed length 12 characters; this may be extended to identify revised versions etc. Numbers assigned 1898 to 2001 had only 2 characters for year ("structure A"); post 2001 the location of element parts was altered to accommodate a 4 digit year ("structure B"). Under each structure, the prefix, year, and serial number are the basic elements required to make an LCCN unique. The prefixes have limited semantic content.		
Rights coverage	No specific rights information carried.		
Strengths	A very long standing scheme for the identification of catalogue records, which may be useful for the identification of books before the application of the ISBN in 1970.		
Weaknesses	Only covers books catalogued by the LoC; does not identify the book but the catalogue record (a nice but an important distinction); although in many library records, not widely used elsewhere.		

Name	MARC 21		
ARROW type	Metadata - library	***Use in ARROW***	MARC21 Format for Bibliographic Data (expressed in XML) is used to exchange bibliographic information in the Library Domain
Acronym	MARC21	***Reference***	
Governance	Network Development and MARC Standards Office, Library of Congress, Washington		
URL	http://www.loc.gov/marc/marcdocz.html		
Status	MARC21 Update 11 Jan 2010 See also Change Announcements : http://www.loc.gov/marc/marcginf.html#intro	***Implementation***	Widespread implementation in major library institutions through the world (especially, but not only, in the English-speaking countries)
Availability	Specifications freely downloadable from the MARC Standards Office web site; no licence required.		
Description	MARC 21 was designed in 1998 to redefine the original MARC record format for the 21st century and to make it more accessible to the international community. MARC 21 is a result of the combination of the United States and Canadian MARC formats (USMARC and CAN/MARC). MARC21 is based on ISO 2709has formats for the following five types of data: Bibliographic Format, Authority Format, Holdings Format, Community Format, and Classification Data Format. MARC 21 allows the use of two character sets, either MARC-8 or Unicode encoded as UTF-8. MARC-8 is based on ISO 2022 and allows the use of Hebrew, Cyrillic, Arabic, Greek, and East Asian scripts. MARC 21 in UTF-8 format allows all the languages supported by Unicode. The MARC 21 format is maintained by the Network Development and MARC Standards Office at the Library of Congress and the Standards and the Support Office at the Library and Archives Canada. Input for development is provided by MARC 21 users from around the world, including libraries, library networks and utilities, and library system vendors. The content of the records is defined by a separate set of rules such as the Anglo-American Cataloguing Rules (AACR2) and, from 2010 by Resource Description and Access (RDA) [although RDA is declared independent of any technical platform]. See also: ISO 2709, MARC XML, AACR2, RDA		
Rights coverage	Various pieces of information relevant to determining Rights may be encoded within a MARC21 record. For example, fields 506 (Restrictions on Access), 540 (Terms Governing Use and Reproduction) and 521 (Information Relating to Copyright Status)		
Strengths	MARC21 records are used around the world, predominantly by libraries, to enable the consistent description and communication of bibliographic data between computers. See also : ISO 2709		
Weaknesses	Technically dated. See also : ISO 2709		

Name	marcXchange		
ARROW type	Metadata - library	***Use in ARROW***	Not used directly. The adoption of MARC21 as the preferred MARC format for ARROW could allow a consistent use of marcXchange in the future, if necessary
Acronym	marcXchange	***Reference***	ISO 25577
Governance	ISO TC46/ SC4		
URL	http://www.iso.org/iso/catalogue_detail.htm?csnumber=43005		
Status	ISO Standard (2008)	***Implementation***	
Availability	Available for purchase from ISO		

Description

marcXchange defines a generalised schema suitable for representing any ISO 2709-based representation of a MARC record in XML. marcXchange is heavily influenced by MARCXML, Library of Congress' XML Schema tied to the MARC21 format.

"The international exchange of records uses very few internationally recognized formats. MarcXchange is mainly intended for regional usage or as a framework for making regional schemas. Experience has shown that there is a need for regional deviations — even if MARC 21 or UNIMARC is chosen as the regional format. This Schema provides a specification for the development of local simple schemas, ensuring compatibility."

MarcXchange has a number of uses, including; representation of individual MARC records and groups of MARC records in XML; as an extension to METS; for transfer of MARC records in web services like SRW; to represent metadata for harvesting, for example via OAI-PMH.

See also: MARC, UNIMARC, MARC XML, METS

Rights coverage

This XML Schema is capable of encoding Rights information that already exists within a MARC record.

Strengths

marcXchange is expressed in XML and this partly solves ISO 27.09 weaknesses from the point of view of syntax. marcXchange can be used as a bridge between ISO 2709 Format of information exchange and other ways to structure data.

Weaknesses

See comments on Format of information exchange (ISO 2709).

Name	MARC XML		
ARROW type	Metadata - library	***Use in ARROW***	Elements of ARROW messaging are based on the MARC XML expression of MARC21
Acronym	MARC XML	***Reference***	
Governance	Network Development and MARC Standards Office, Library of Congress		
URL	http://www.loc.gov/standards/marcxml/		
Status	v1.1 (2003)	***Implementation***	
Availability	The XML Schema is freely available for download from the Library of Congress.		
Description	"This schema supports XML markup of MARC21 records as specified in the MARC documentation (see www.loc.gov). It allows tags with alphabetics and subfield codes that are symbols, neither of which are as yet used in the MARC 21 communications formats, but are allowed by MARC 21 for local data. The schema accommodates all types of MARC 21 records: bibliographic, holdings, bibliographic with embedded holdings, authority, classification, and community information." Developed by the Library of Congress in collaboration with OCLC and RLG, MARC XML supports the encoding and exchange of MARC 21 records in the XML format widely used across the Web. Library of Congress maintains this Schema, as well as providing software to ensure lossless conversion to and from MARC 21 records encoded in the traditional ISO 2709 structure. In the mid 1990's, Library of Congress developed two SGML DTDs that supported the conversion of cataloging data from the MARC data structure to SGML (and back) without loss of data. These DTDs are deprecated in favour of MARCXML. See also: MARC21, MARC, marcXchange		
Rights coverage	This XML Schema is capable of encoding Rights information that already exists within MARC 21 records.		
Strengths	See comments on MARC; MARC XML is expressed in XML and this partly solves ISO 27.09 weaknesses from the point of view of syntax. The MARC XML can be used as a bridge between MARC and other ways to structure data		
Weaknesses	See comments on MARC; MARC XML is an expression only of MARC21 – this has not so far been a problem from an ARROW standpoint, but it might be necessary to consider marcXchange in future.		

Name	Metadata Encoding & Transmission Standard		
ARROW type	Metadata - library	***Use in ARROW***	Not used (not applicable)
Acronym	METS	***Reference***	
Governance	Network Development and MARC Standards Office, Library of Congress		
URL	http://www.loc.gov/standards/mets/		
Status	Schema version 1.9 (February 2010)	***Implementation***	http://www.loc.gov/standards/mets/mets-registry.html
Availability	The METS XML schema is freely available for download from the Library of Congress web site.		
Description	METS is a specification for encoding descriptive, administrative and structural metadata about objects within a digital library, expressed by means of an XML Schema. METS seeks to capture the metadata necessary for management of digital objects within a repository and for exchange of those objects between repositories. A METS document consists of seven major sections, including a Header, Descriptive Metadata, Administrative Metadata, File Section, Structural Map, Structural Links, and Behaviour. Depending upon its purpose, a METS document may be used as packaging information for a Submission Information Package (SIP), Archival Information Package (AIP), or Dissemination Information Package (DIP) in compliance with the Open Archival Information System (OAIS) Reference Model. Approved external Schemas such as the Dublin Core, MODS, MARC XML, PREMIS and VRA Core are available to increase consistency within METS metadata descriptions. A set of extensible Profiles are used to cover the specific needs of particular materials such as musical scores, printed monographs, etc.		
Rights coverage	The 'Administrative Metadata' section of a METS record provides scope for recording intellectual property data, specifically within the <rightsMD> and <sourceMD> sub-elements.		
Strengths	Very broad scope of metadata coverage for description of digital objects within a library context.		
Weaknesses	No relevant application in ARROW		

Name	METSRights		
ARROW type	Metadata – rights/permissions	***Use in ARROW***	Not used (not relevant)
Acronym	METSRights	***Reference***	
Governance	Network Development and MARC Standards Office, Library of Congress		
URL	www.loc.gov/standards/rights/METSRights.xsd (undocumented schema)		
Status	A draft was published in 2006, but we understand a revision is currently in progress	***Implementation***	Has been implemented by the Library of Congress; UC Berkeley Library; Geisel Library, UC San Diego; the National Library of Australia; Ex Libris
Availability	Appears to be freely available		

Description

Developed as an extension to METS in between 2004 & 2006 in response to a requirement for a simple Rights schema that the METS community could use while other more comprehensive Rights Expression Language (REL) schemas such as XrML, ODRL, and others are being developed, and debated. The focus of the simple Rights schema is to simply declare or document some basic facts about the digital collections being created and/or included in institutional digital repositories.

This rights declaration schema focuses upon:

- digital resources owned or controlled by the digital repository rather than e-resources accessed remotely, formally licensed and subscribed to by an organization (the area covered by the DLF ERMI group)
- declaring the rights holders and rights associated with the digital resources mentioned above rather than trying to fully express all rights as would a REL designed to be used with a Digital Rights Management system or product
- simplifying the declaration as much as possible given the fact that the whole DRM & REL scene is changing so rapidly

This Rights Declaration schema has 3 main elements:

- A simple declaration of type of rights (copyrighted, licensed, public domain, contractual, other) and the public statement of that Rights Declaration,
- The naming of the Rights Holder(s) with appropriate contact information,
- The Context(s) for the rights declaration based on type of users who have a set of permissions for a digital object or part of a digital object. If there are any constraints to the permissions, those are also expressed within the context by listing the constraints and explaining them in a constraint description element.

Rights coverage	See above
Strengths	Simplicity; some significant implementations
Weaknesses	Not widely known; undocumented; provides only limited “rights declarations” rather than a fully formed set of rights and permissions expressions

Name	Multipurpose Internet Mail Extensions		
ARROW type	Technical protocol	***Use in ARROW***	Not explicitly referenced
Acronym	MIME	***Reference***	IETF RFC 2045
Governance	Internet Society / Internet Engineering Task Force		
URL	http://tools.ietf.org/html/rfc2045		
Status	DRAFT STANDARD	***Implementation***	Very widely implemented
Availability	All IETF RFCs ("Requests for Comment") are freely available.		

Description	Published as a draft standard in November 1996, MIME defines how to package data of all kinds for Internet message exchange. MIME has been adopted and adapted in several key Internet protocols, mostly notably in HTTP. Strictly speaking, RFC 2045 is only applicable to Internet mail (email), but the term MIME is still used informally in other applications, such as in HTTP, which differs from strict MIME in several important respects. What follows deals with the strict form of MIME as defined by RFC 2045. MIME defines three things: (1) how to include in a message text that uses a character set other than US-ASCII; (2) how to include non-text data in a message; (3) how to mix different types of content in a single message. MIME crucially introduces the concepts of "content type" (sometimes informally referred to as "MIME type") and "content transfer encoding", and with respect to the latter defines the "Base64" method for encoding data of all kinds. MIME defines five header fields: MIME-Version; Content-Type; Content-Transfer-Encoding; Content-ID; and Content-Description. Of the latter two, Content-ID is most frequently used to label different content parts in a multipart message, for example to distinguish between alternative representations of the same email content (e.g. plain text, rich text, HTML). Content-Description is not widely used. RFC 2045, which defines the MIME format, is Part 1 of a multi-part series of RFCs, which includes RFC 2046 (media types), RFC 2047 (extensions allowing non-US-ASCII characters in header fields) and RFC 2048 (IANA registration procedures)
Rights coverage	MIME is not concerned with the semantics of a message. Its header fields are designed to facilitate message handling and not to convey business data.
Strengths	Universally used to support network communications
Weaknesses	None relevant to ARROW

Name	Metadata Object Description Schema		
ARROW type	Metadata - library	***Use in ARROW***	Not used (not applicable)
Acronym	MODS	***Reference***	
Governance	Network Development and MARC Standards Office, Library of Congress		
URL	http://www.loc.gov/standards/mods/		
Status	Schema v3.4 (2010)	***Implementation***	
Availability	The MODS Schema is freely available for download from the Library of Congress.		
Description	Developed by Library of Congress, the MODS XML Schema was intended to offer a compromise between the complexity of the MARC format and the perceived simplicity of the Dublin Core. The Schema defines a core set of 20 bibliographic elements (plus two structural 'root' elements), particularly relevant to digital library applications. "The MODS record has been designed to carry key data elements from the MARC record but does not define all of the MARC fields and does not use the field and subfield tagging from the MARC standard. There are data elements in MODS that are not compatible with the MARC record so there is some loss translating from MARC to MODS and from MODS to MARC. There is no commitment on the part of the Library of Congress to maintain compatibility between the two metadata formats beyond what is convenient to the community of MODS users." The MODS Implementation Registry at Library of Congress lists just 29 implementors, heavily skewed toward national and research libraries in the USA and United Kingdom. See also: Dublin Core, MARC21		
Rights coverage	The accessCondition element is capable of expressing information on restrictions relating to access, use, and reproduction of resources.		
Strengths	Expressed in XML and this partly solves the ISO 2709 weaknesses from the point of view of syntax. In addition MODS is simpler than MARC and is "human readable". MODS should facilitate the transition from MARC towards other way to structure data		
Weaknesses	Not widely implemented – and as a result of no direct relevance to ARROW		

Name	MPEG-21 Digital Item Declaration		
ARROW type	Published content	***Use in ARROW***	Not used (not applicable)
Acronym	MPEG-21 DID	***Reference***	ISO/IEC 21000-2
Governance	ISO/IEC JTC 1/SC 29/WG 11 (MPEG - Moving Pictures Expert Group)		
URL	http://mpeg.chiariglione.org/standards/mpeg-21/mpeg-21.htm		
Status	IS; 1st Amd to 2nd Ed: WD	***Implementation***	Limited
Availability	Available from <http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html>. Essential patents are claimed by Matsushita Electric Industrial Co., Ltd and Mitsubishi Electric Corp. However WG11 experts suggest that they are not applicable.		
Description	A Digital Item is a structured digital object with a standard representation, identification and metadata within the MPEG-21 framework. This entity is the fundamental unit of distribution and transaction within this framework. Declaring a Digital Item involves specifying the resources, metadata, and their interrelationships for a Digital Item. ISO/IEC 21000-2 defines a set of abstract terms and concepts to form a useful model for declaring Digital Items. The goal of this model is to be as flexible and general as possible, while providing for the “hooks” that enable higher level functionality. This, in turn, allows the model to serve as a key foundation in the building of higher level models in other MPEG-21 elements (including Digital Item Identification or the inclusion of rights expressions or descriptive metadata). A Digital Item is the digital representation of “a work” (the use of the term "work" here is not to be confused with the <indecs> concept of an abstraction), and as such, it is the thing that is acted upon (managed, described, exchanged, collected, etc.) within the model. The goal of this model is to be as flexible and general as possible, while providing for the “hooks” that enable higher level functionality. The DID model also provides a common set of abstract concepts and terms that can be used to define schemas for their representation, or to perform mappings between existing schemes. ISO/IEC 21000-2 is a powerful standard, which has experienced comparatively little uptake. This is believed to have one main reason: its own — and ISO/IEC 21000 generally — flexibility which makes the DID model applicable to a wide range of application. Each application will only require a small subset of the features offered by DID; thus making the standard appear to be bloated.		
Rights coverage	DIDs allow the inclusion of Identifiers — via ISO/IEC 21000-3 Digital Item Identification — or any rights metadata such as Rights expressions		
Strengths			
Weaknesses			

Name	MPEG-21 Rights Data Dictionary		
ARROW type	Metadata – rights/permissions	***Use in ARROW***	Not used (no relevant requirement)
Acronym	MPEG-21 RDD	***Reference***	ISO/IEC 21000-6
Governance	ISO/IEC JTC 1/SC 29/WG 11 (Motion Picture Experts Group or MPEG); the Registration Authority is the International DOI Foundation.		
URL	http://mpeg.chiariglione.org/standards/mpeg-21/mpeg-21.htm		
Status	Published	***Implementation***	None known
Availability	ISO Publication		
Description	The Rights Data Dictionary (RDD) comprises a set of clear, consistent, structured, integrated and uniquely identified Terms to support the MPEG-21 Rights Expression Language. The structure of the dictionary is specified, along with a methodology for creating the dictionary. The means by which further Terms may be defined is also explained. The Dictionary is a prescriptive Dictionary, in the sense that it defines a single meaning for a Term represented by a particular RDD name (or Headword), but it is also inclusive in that it recognizes the prescription of other Headwords and definitions by other Authorities and incorporates them through mappings. The RDD also supports the circumstance that the same name may have different meanings under different Authorities. The RDD specification has audit provisions so that additions, amendments and deletions to Terms and their attributes can be tracked. RDD recognises legal definitions as and only as Terms from other Authorities that can be mapped into the RDD. Therefore Terms that are directly authorized by RDD neither define nor prescribe intellectual property rights or other legal entities. As well as providing definitions of Terms for use in the REL, the RDD specification is designed to support the mapping and transformation of metadata from the terminology of one namespace (or Authority) into that of another namespace (or Authority) in an automated or partially-automated way, with the minimum ambiguity or loss of semantic integrity. The dictionary is based on a logical model, the Context Model, which is the basis of the dictionary ontology. The model is described in detail in the specification. It is based on the use of verbs which are contextualised so that a dictionary created with it can be as extensible and granular are required. The “baseline technology” from which MPEG-21 RDD was developed was a project called “<indecs>2 rdd”, developed by a consortium of 8 companies. A patent in this technology is held by Contecs:DD.		
Rights coverage	The MPEG-21 RDD was designed to enable interoperability in rights management applications		
Strengths	A powerful structural tool for the management of semantic interoperability for rights management		
Weaknesses	Not implemented		

Name	MPEG-21: Rights Expression Language		
ARROW type	Metadata - rights/permissions	***Use in ARROW***	Not used (not applicable)
Acronym	MPEG-21 REL	***Reference***	ISO/IEC 21000-5
Governance	ISO/IEC JTC 1/SC 29/WG 11 (Motion Picture Experts Group or MPEG)		
URL	http://mpeg.chiariglione.org/standards/mpeg-21/mpeg-21.htm		
Status	Published	***Implementation***	Most known implementations are of XrML, the proprietary ContentGuard format which was the baseline technology used for the development of the MPEG REL.
Availability	ISO Publication		
Description	"A Rights Expression Language (REL) is a machine-readable language that declares rights and permissions. The MPEG REL, as defined by ISO/IEC 21000-5, provides flexible, interoperable mechanisms to support transparent and augmented use of digital resources throughout the value chain in a way that protects the digital resource and honours the rights, conditions, and fees specified for it. For instance, it provides mechanisms in support of publishing, distributing, and consuming digital content such as electronic books, digital movies, digital music, broadcast content, interactive games, computer software, and other creations in digital form. It also supports specification of access and usage controls for digital content in cases where financial exchange is not a term of use, and supports exchange of sensitive or private digital content and personal information. The standard REL can support guaranteed end-to-end interoperability, consistency, and reliability among different systems and services. To do so, it offers richness and extensibility in declaring rights, conditions, and obligations; ease and persistence in identifying and associating these with digital content; and flexibility in supporting multiple usage/business models."		
Rights coverage	Essentially provides for the expression of permissions of use, typically for a single instance of a resource, and typically bound to enforcement of those permissions through the application of DRM technology.		
Strengths	A comprehensive REL which has been standardised by ISO/IEC.		
Weaknesses	Implementation of the standard appears to be very limited		

Name	***Metasearch XML Gateway***		
ARROW type	Search	***Use in ARROW***	Not used (distributed search not implemented in ARROW)
Acronym	MXG	***Reference***	
Governance	Metasearch Initiative proposed by NISO		
URL	http://www.niso.org/workrooms/mi		
Status	v1.0 published Aug 2006	***Implementation***	
Availability	Specification freely available from the website. No licence required for implementation.		
Description	The NISO Metasearch XML Gateway is a low-barrier-to-entry method to expose content to metasearch services and more effectively interoperate with them. Its protocol defines a simple message and response model for allowing a metasearch service to query a content database and receive a standardized XML response. So MXG provides a mechanism for a content provider to expose its content and services to a Metasearch Service. Metasearch Services are a class of services that allow an end user to find content in multiple services with a single search. MXG is based on the NISO-registered Search and Retrive URL (SRU) protocol. The Metasearch Provider sends individual queries for each resource that uses MXG URLs via HTTP. Each Content Provider retuns an MXG compliant XML formatted response to those queries. The Metasearch is responsible for parsing, aggregating and displaying of the records retrieved from multiple sources to the end user. Three levels of implementation are defined for MXG. Each level requires increasing compliance with specifications of the SRU protocol; only the third level is fully compliant SRU. Level 1 defines a standard URL which will accommodate ANY query language; Level 2 extends Level 1 by adding the requirement that servers provide an SRU EXPLAIN record to define the capabilities of the server; Level 3 extends Level 2 by adding the requirement that servers support a standard query grammar: CQL. Concerning the XML schemas to utilize for records, MXG required a minimum of one schema although multiple ones may be supported for different Metasearch Provider. Any schema is allowable, even custom created one. From standard schema, some choices could be DC (Dublin Core), MODS (Metadata Object Description Standard), LOM (Learning Object Metadata), and so on.		
Rights coverage	This protocol doesn't provide a way for defining rights content but a way for exposing and querying them.		
Strengths			
Weaknesses	Not relevant to ARROW		

Name	National Bibliography Number		
ARROW type	Identification	***Use in ARROW***	A potential alternative manifestation identifier if no ISBN is available
Acronym	NBN	***Reference***	IETF RFC 3188*
Governance	None formal; Conference of Directors of National Libraries (CDNL) has oversight of any common development.		
URL	ftp://ftp.rfc-editor.org/in-notes/pdfrfc/rfc3188.txt.pdf *		
Status	IETF RFC (2001)*	***Implementation***	Limited
Availability	National local policies may limit the NBN usage. NBNs are not in broad use and the application to commercial content is small. Currently, only a small number of National Libraries (e.g., Finland, Sweden, Germany, Hungary) provide services to resolve links using NBNs. Decisions on assignment made across countries are not co-ordinated and the resolution approach may vary from one country to another. Scope may be limited in certain cases. A National Libraries Resolver Discovery Service has been proposed but has not yet been developed or deployed: the German National Library is starting a project to establish this service, but its timeline and costs are currently unknown.		
Description	Generic name referring to a group of identifier systems used by national libraries for identification of deposited publications lacking an identifier, to associate descriptive metadata (cataloging) that describes the resources. NBNs can be seen as a fall-back mechanism: if no other, better established identifier such as ISBN can be given, an NBN is assigned. There is no common syntax specification or global authority; hence NBNs are unique only on national level. *An attempt has been made to make the system internet-usable through a specification (http://www.ietf.org/rfc/rfc3188.txt) to represent NBN as URN, which adds a controlled prefix (ISO country code but additional sub-domains can be included), to ensure global uniqueness. Its registration authority is the Library of Congress; only national libraries may register NBNs. This has patchy support: In general the URN (Universal Resource Name) infrastructure, which the proposed NBN expansion is based on, has not really taken off, but some individual national libraries have invested effort in providing a service.		
Rights coverage	Holdings and access information may be associated with a particular NBN.		
Strengths	A fall back identifier when no other identifier is available		
Weaknesses	Limited to National Library implementation; implementation patchy and inconsistent.		

Name	Open Archives Initiative Protocol for Metadata Harvesting		
ARROW type	Technical protocol	***Use in ARROW***	Not used (metadata harvesting not implemented in ARROW)
Acronym	OAI-PMH	***Reference***	
Governance	Open Archives Initiative Steering Committee		
URL	http://www.openarchives.org/		
Status	v2.0 (2002)	***Implementation***	Widespread use by academic institutional repositories
Availability	Specification freely downloadable from the OAI web site; no licence required. Various tools freely available from http://www.openarchives.org/pmh/tools/		
Description	OAI-PMH provides an application-independent framework to support harvesting (or collection) of metadata from repositories conforming to the Protocol. Typically these repositories are archives of academic papers, often hosted by universities and similar organisations. By default, metadata is made available to harvesters in the form of a simple Dublin Core record, marked up in XML. Other record formats may be offered in addition to this. The protocol defines six simple requests, enabling harvesters to discover basic information about a repository, the metadata formats it supports, the items it contains, and additions or deletions made over time and harvest the records as a whole, or parts of them following sets defined by the data provider. Conforming repositories are not required to register their existence, although several large registries have been created for this purpose including http://roar.eprints.org and http://www.openarchives.org/Register/BrowseSites. OAI-PMH is often closely associated with the Open Access movement, although the Protocol itself is equally relevant to dissemination of data concerning closed access journals and other forms of content. See also: Dublin Core		
Rights coverage	"The OAI does not define or prescribe any rights management scheme. Issues of access restriction and management of intellectual property in exposed metadata are the responsibility of the data providers that adopt the protocol." The protocol supports the Dublin Core Metadata Element Set (DCMES) by default, which includes the capability to optionally carry a statement on Rights.		
Strengths	Widely implemented in the academic repository sector.		
Weaknesses	Little used outside the academic repository sector. No query mechanism.		

Name	Online Computer Library Centre Catalog Number		
ARROW type	Identification	***Use in ARROW***	A potential alternative manifestation identifier if no ISBN is available
Acronym	OCLC Number	***Reference***	
Governance	OCLC (WorldCat)		
URL	http://www.oclc.org/WorldCat/default.htm		
Status	De facto standard (created 1971)	***Implementation***	71,000 libraries in 112 countries
Availability	OCLC numbers are usually carried in WorldCat participating union catalogues or may be found through WorldCat searches or services using the database (eg FirstSearch). OCLC numbers can be formed into persistent internet URLs by appending 1- to 8-digit OCLC Number for the item to a Worldcat PURL service (this format is only specified when ISBN or ISSN is unavailable for the item, which if available is the preferred option for citing). The link format is http://www.worldcat.org/oclc/[item OCLC Number] OCLC numbers as WorldCat links resolve to information on participating libraries holding a given book, through "Open WorldCat" (abbreviated records from WorldCat; launched 2003). The entire database is available for search-engine harvesting.		
Description	A unique number (OCLC number: sometimes called WorldCat number) assigned to items in WorldCat, a de facto international merged catalogue (over 136 million bibliographic records that represent more than 1 billion individual holding items:) maintained by OCLC. A collocation function provides links to all the editions of an item: http://www.worldcat.org/oclc/[NUMBER]/editions. This therefore performs some of function envisaged for the ISTC service. Similarly in 2007 (still in beta –stage) "Worldcat identities" was launched to provide OCLC numbers for authors and characters and so overlaps with functions of the proposed ISNI.		
Rights coverage	Holdings and access information is associated with the Worldcat record. Some users reach WorldCat.org from links in partner search engines or bookseller Web sites.		
Strengths	Comprehensive with respect to Worldcat holdings. Provides identities for objects outside other identification schemes (for example, books published before ISBN came into use). OCLC are actively working on projects to collocate editions of the same work and by the same author.		
Weaknesses	Not widely promulgated as an identifier, so not tested outside the OCLC partner environment. Duplication among the one billion "individual holdings" is suspected but not known (de-duplication is a non-trivial task). The collocation criteria may not match ARROW requirements.		

Name	Open Digital Rights Language		
ARROW type	Metadata - rights/permissions	***Use in ARROW***	Not used (not applicable)
Acronym	ODRL	***Reference***	
Governance	ODRL International Advisory Board		
URL	http://odrl.net/		
Status	v1.1 published 2002; active work is in hand on the development of v2.0	***Implementation***	Implemented primarily through the Open Mobile Alliance REL, which is a development of ODRL
Availability	Specifications freely available from the ODRL website		
Description	"The Open Digital Rights Language (ODRL) Initiative is an international effort aimed at developing and promoting an open standard for rights expressions. ODRL is intended to provide flexible and interoperable mechanisms to support transparent and innovative use of digital content in publishing, distributing and consuming of digital media across all sectors and communities." The ODRL v2.0 "Core Model" is a formal model that uses the standard modelling language UML. One of the benefits of using UML is that it enables a formal model to be expressed graphically, which makes it easier for humans to understand than alternatives that are purely text-based, while remaining rigorous and a reliable basis for building data models and creating associated syntax bindings.		
Rights coverage	Essentially provides for the expression of permissions of use, typically for a single instance of a resource, and typically bound to enforcement of those permissions through the application of DRM technology.		
Strengths	Open source, and freely available. Strong Core Model appears to be fairly comprehensive (although an initial review suggests it may not meet all Use Cases). Dealing explicitly with interoperability problems and endorsing RDF as a tool. Widely used since 2004 through Open Mobile Alliance implementation		
Weaknesses	Not yet widely implemented outside the mobile space; some semantic imprecision (although this can be overcome through "profiles").		

Name	ONIX Standards Framework		
ARROW type	Metadata - commercial	***Use in ARROW***	Elements of the ONIX framework have been critical to the development of ARROW messaging – see individual entries
Acronym	ONIX	***Reference***	
Governance	EDItEUR		
URL	www.editeur.org		
Status	N/A	***Implementation***	In the ONIX standards
Availability	All ONIX standards are freely available under a perpetual non-exclusive cost-free licence.		
Description	The ONIX name is used by EDItEUR for a family of standards, some of which are in turn families of related standards; the primary members of this family are: ONIX for Books ONIX for Serials ONIX-SPS: Serials Products and Subscriptions ONIX-SOH: Serials Online Holdings ONIX-SRN: Serials Release Notification ONIX for Licensing Terms ONIX-PL: ONIX for Publications Licenses ONIX-DS: ONIX for Distributions ONIX-RP: ONIX for Repertoire Other ONIX specification include formats for identifier registration (ONIX for DOI Registration; ONIX for ISTC Registration) and some formats (such as ONIX for Subrights) which have been partially developed but not piloted. ONIX messages share syntactic structures and (to the extent possible) code value definitions. For all ONIX messages there is a clear separation between structure and code lists; this enables code lists to be updated as frequently as may be necessary without disturbing structure, providing essential flexibility while maintaining backward compatibility.		
Rights coverage	Various of the ONIX messages are designed to carry rights and/or permissions data appropriate to their particular function.		
Strengths	ONIX provides a widely recognised brand identity for standards, particularly in the book retail chain. ONIX for Books is very widely implemented internationally. The family of standards is designed to be internally consistent and has been designed to be flexible in its approach to meeting new requirements, including a well-structured approach to rights and permissions metadata..		
Weaknesses	Not all ONIX standards have been widely deployed. Interoperability with some library standards is imperfect, because differences in requirement have led to significant differences in form.		

Name	ONIX for Books		
ARROW type	Metadata - commercial	***Use in ARROW***	ONIX for Books has provided the model and much of the semantics for ARROW messaging
Acronym	ONIX-4B	***Reference***	
Governance	EDItEUR, through the ONIX for Books International Steering Committee		
URL	http://www.editeur.org/83/Overview/		
Status	Release 2.1 rev 03 (2004) Release 3.0 (2009)	***Implementation***	Widely used within the book trade since 2000.
Availability	See ONIX		
Description	ONIX for Books is a standard XML format for the communication of product information from publishers to wholesalers and retailers, either direct or through the services offered by "data aggregators". While focused on books, it also covers other types of non-periodical publication which may be distributed through the book supply chain. In content terms, ONIX for Books carries bibliographic detail (such as product identifiers, titles, contributors, binding and format, dimensions, page extent, publisher and imprint); trade detail for different markets (distributor, availability, RRP, discount group); and promotional detail (descriptions, links to supporting material – text, audio, video – on a publisher's website or elsewhere). Release 3.0 is expected to come into general use during 2011. This is the first release in which digital products such as ebooks have been treated as a "core" element in ONIX coverage. ONIX for Books has been or is being adopted as the national standard for book trade product information in at least fourteen countries, including the UK, US, Canada (both English- and French-speaking), Australia, France, Germany, Spain, Italy, Belgium, the Netherlands, and in Scandinavia. The format is also being used to communicate metadata from publishers to enrich library catalogues and as part of national library CIP programmes. As a communications format, ONIX for Books makes it possible to deliver rich product information into the book supply chain in a standard form. By providing a template for the content and structure of a product record, ONIX has helped to stimulate the introduction of better internal information systems for publishers, capable of bringing together all the metadata needed for the description and promotion of new and backlist titles.		
Rights coverage	Can specify territorial sales rights attaching to a product, and any non-territorial sales restrictions applying either globally or (Release 3.0 only) within a designated territorial market; also distribution rights applying to a designated distributor. For digital products, provision for description of common forms of usage constraint, applied either by the characteristics of the hardware and software used. or by DRM.		
Strengths	Widely adopted by the international publishing community and the book trade; capable of communicating complex and rich product information about books and digital products		
Weaknesses	Entirely focused on commerce; no cataloguing rules. Mapping to MARC does not necessarily create valid MARC records. Using fixed standard references or vocabularies is essential for efficient business communications, but may create difficulties when and where interoperability with non-ONIX systems is needed. Work is continuing on ONIX/MARC interoperability (including at OCLC http://www.oclc.org/research/news/2010-04-09.htm and through the VMF project http://cdlr.strath.ac.uk/VMF/)		

Name	ONIX for Distribution		
ARROW type	Metadata - commercial	***Use in ARROW***	Not used (not yet applicable)
Acronym	ONIX-DS	***Reference***	
Governance	EDItEUR, managed jointly with IFRRO		
URL	http://www.editeur.org/23/ONIX-for-RROs/		
Status	v1.0 published 2008	***Implementation***	Limited
Availability	See ONIX		

Description	One of two message formats commissioned from EDItEUR by IFRRO for the management of communication between RROs (the other being ONIX-DS). A member of the ONIX for Licensing Terms family. ONIX-DS is designed to allow the sharing of "distribution" information between RROs. Distribution is the process by which revenues are allocated by an RRO. A distribution message therefore typically accompanies a payment, and informs the recipient of the elements that make up the payment. Essentially, ONIX-DS is a specialised "sales reporting message", for reporting on transactions in delegated rights. So far as we are aware, there is only one live implementation of ONIX-DS, for communication between CLA and PLS in the UK. Distribution messages can be very large indeed, and interchange between CLA and PLS is managed using FTP. It is anticipated that ONIX-DS will be widely implemented within the RRO community more quickly than ONIX-RP, and IFRRO has set a target of 50% of its members to implement. In order to facilitate this process, IFRRO has developed an "ONIX-DS compliant" Excel format to supplement the XML format. Because of the considerable diversity in systems and semantics in the RRO community worldwide, it was agreed that the core semantics included in the IFRRO namespace should initially be limited, and supplemented by local namespaces agreed between trading partners. In the UK, a substantial local "UKRRO" namespace has been developed. It is hope that the IFRRO namespace will be developed over time through terms from local namespaces being "promoted" to the IFRRO namespace.
Rights coverage	Although the rights and permissions semantics which have so far been developed are mapped to existing RRO licensing activities, expanding the semantics to new applications is relatively straightforward, because of the inherent flexibility of a standard developed within the ONIX-LT framework.
Strengths	A comprehensive and flexible messaging standard for sharing distribution information; no known competitors (although it can be characterised as a specialised sales reporting message).It is possible for trading partners to adopt a simplified profile to meet specific requirements
Weaknesses	Distribution information is inherently complex, because of the complexity of the underpinning rights and licensing information on which it depends. Implementing ONIX-DS is complex, but this would be true for any message with the same scope.

Name	ONIX for ISTC Registration		
ARROW type	Metadata - commercial	***Use in ARROW***	ARROW is piloting ISTC registration using this message suite.
Acronym	ONIX-ISTC	***Reference***	
Governance	ISTC International Agency in association with EDItEUR		
URL	http://www.editeur.org/106/ONIX-ISTC-Registration-Format/		
Status	v1.0	***Implementation***	Limited because of limited implementation of ISTC
Availability	See ONIX		
Description	ONIX for ISTC Registration is a concise XML format for handling message exchanges between ISTC registrants and ISTC registration agencies. The same format is used for both a request and a response, with coding to indicate the message status within an interchange. The metadata content is limited to that which has been determined by the ISTC International Registration Agency as required for the international ISTC registration database. The "style" of the message is similar to that of ONIX standards, and wherever possible existing ONIX elements have been used; but a substantial proportion of the format uses elements which are specialised for the ISTC registration application. The format is not intended for general use for the delivery of ONIX metadata describing a "work": the content is too limited, and the elements relating to the purpose and status of the message are too specialised.		
Rights coverage	None.		
Strengths	Fitness for purpose; shared data constructs with other ONIX messages.		
Weaknesses	Designed for a specific purpose, and application is limited to this purpose alone.		

Name	ONIX for Licensing Terms		
ARROW type	Metadata - rights/permissions	***Use in ARROW***	Provides the model and semantics for communication about rights and permissions in ARROW messaging
Acronym	ONIX-LT	***Reference***	
Governance	EDItEUR		
URL	http://www.editeur.org/85/Overview/		
Status	Unpublished framework	***Implementation***	Through specific ONIX formats
Availability	N/A		
Description	ONIX-LT is the conceptual framework within which ONIX messaging standards for communicating rights, licensing and permissions metadata are developed. It is not separately published. ONIX-LT can be seen not only in ONIX-PL, but also in ONIX-RP and ONIX-DS. It can also be seen in the ACAP semantics.		
Rights coverage	A framework designed specifically to support the communication of rights and permissions information.		
Strengths	A framework approach which should allow any fom of rights and permissions information to be communicated between trading partners.		
Weaknesses	The communication of rights and permissions information is inherently complex.		

Name	ONIX for Publication Licences		
ARROW type	Metadata - rights/permissions	***Use in ARROW***	None immediately, but available for use for the communication of licence information when required
Acronym	ONIX-PL	***Reference***	
Governance	EDItEUR		
URL	http://www.editeur.org/21/ONIX-PL/		
Status	v1.0 (December 2008)	***Implementation***	Very limited
Availability	Specification freely downloadable from the EDItEUR web site; no licence required. An open source tool (OPLE) is also available which facilitates the expression of licences in ONIX-PL.		
Description	ONIX-PL is part of the "ONIX for Licensing Terms" family of messages. It was developed specifically to support the expression of licences in the publisher to academic library supply chain, in response to the difficulty being experienced by libraries in managing very diverse licence terms for their growing collections of digital resources. Licences are encoded in accordance with a standard XML schema, with the intention that they can then be communicated within the supply chain. While it is possible to encode a complete licence in ONIX-PL, there is a clear difference between the encoding of the majority of licence clauses, and those clauses which grant (or withhold) specific usage permissions. While other clauses are simply encoded as text within a standard heading structure, usages are encoded using tightly constrained semantics and syntactic structures, which makes usage clauses machine interpretable (allowing, for example, highly simplified permissions information to be displayed to users at the point of use). The validity of the proposed approach has been demonstrated in a project, funded by JISC in the UK, RELI (http://www.lboro.ac.uk/departments/ls/disresearch/RELI/about.html). The primary challenge facing ONIX-PL, like most standards early in their lifecycle, is implementation. While there is strong support for the theory, in practice there are problems on the library side (with a lack of systems able to ingest ONIX-PL licences) and on the publisher side (with uncertainty about who should be creating the XML expressions, and where the necessary skills can be located). Widespread implementation in the supply chain is likely to take some time.		
Rights coverage	The permission semantics of ONIX-PL are specifically geared to communication in the library supply chain; however, ONIX-PL should be extensible to any similar licence application.		
Strengths	Capable of communicating any kind of publication licence, subject to extension of allowed values. Extremely flexible and infinitely extensible.		
Weaknesses	Still in exploratory phase of implementation. Difficult to get traction because of the complexity of the task and the need for both library and publisher systems to be able to support. Publishers having difficulty identifying internal responsibility for a new task.		

Name	ONIX for Repertoire		
ARROW type	Metadata - rights/permissions	***Use in ARROW***	None immediately, but expected to be used for the communication of repertoire information when required
Acronym	ONIX-RP	***Reference***	
Governance	EDItEUR, managed jointly with IFRRO		
URL	http://www.editeur.org/23/ONIX-for-RROs/		
Status	v1.0 published 2008	***Implementation***	Limited
Availability	Documentation freely available from the EDItEUR web site; no licence required.		
Description	One of two message formats commissioned from EDItEUR by IFRRO for the management of communication between RROs (the other being ONIX-DS). A member of the ONIX for Licensing Terms family. ONIX-RP is designed to allow the sharing of "repertoire" information between RROs, a repertoire being the definition of a set of resources to which a specific set of rights or permissions relate. In other words, ONIX-RP allows RROs to share with each other the mandates that they hold from rightsholders. A repertoire may be defined very broadly - " all resources published in the UK by Publisher Y can be included in any photocopying licence worldwide" or very narrowly "this specified resource is to be excluded from this specified licence". So far as we are aware, there is only one live implementation of ONIX-RP, for communication between PLS and CLA in the UK. This is a very active implementation, and messages are shared in real time using web services. Because of the considerable diversity in systems and semantics in the RRO community worldwide, it was agreed that the core semantics included in the IFRRO namespace should initially be limited, and supplemented by local namespaces agreed between trading partners. In the UK, a substantial local "UKRRO" namespace has been developed. It is hope that the IFRRO namespace will be developed over time through terms from local namespaces being "promoted" to the IFRRO namespace.		
Rights coverage	Although the rights and permissions semantics which have so far been developed are mapped to existing RRO licensing activities, expanding the semantics to new applications is relatively straightforward, because of the inherent flexibility of a standard developed within the ONIX-LT framework.		
Strengths	A comprehensive and flexible messaging standard for sharing repertoire information; no known competitors.		
Weaknesses	Repertoire information is inherently complex, and the ONIX-RP specification reflects this reality. Implementing ONIX-RP is a non-trivial challenge, which requires a mix of skills (understanding rights and permissions as well as XML). However, this would be the case for any message standard meeting the same set of requirements.		

Name	OpenSearch		
***ARROW* type**	Search	***Use in ARROW***	Not used (not applicable)
Acronym	OpenSearch	***Reference***	
Governance	the 'OpenSearch community'		
URL	http://www.opensearch.org/		
Status	1.0 (March 2005), 1.1 Draft (December 2005)	***Implementation***	Wikipedia, IE7, Firefox 2+, Windows 7...
Availability	Specifications freely downloadable from the OpenSearch web site.		

Description

Originally created by Amazon's A9.com subsidiary, OpenSearch provides a relatively straightforward means for search engines to syndicate their results for aggregation and re-use by third parties.

Search clients such as your web browser can find and use OpenSearch description documents to learn about the public interface of a search engine. These description documents contain templates that indicate how the search client should make requests. Search engines can use the OpenSearch response elements to add search metadata to results in a variety of content formats.

OpenSearch is a search protocol that primarily supports keyword searching, and is most valuable for searching across unstructured documents.

***Rights* coverage**

OpenSearch does not deal directly with Rights.

Strengths

Weaknesses

Not directly relevant to ARROW

Name	Open URL Framework		
ARROW type	Identification	***Use in ARROW***	Not used (not applicable)
Acronym	Open URL	***Reference***	ANSI/NISO Z39.88
Governance	National Information Standards Organisation (USA)		
URL	http://www.niso.org/kst/reports/standards/		
Status	NISO Standard (2004)	***Implementation***	Appropriate copy resolution.
Availability	Free specification.		
Description	The OpenURL Framework for Context-Sensitive Services (usually called "OpenURL") is a mechanism for packaging and transporting metadata and identifiers over a network. It is used to reference a publication for the purpose of context-sensitive linking through a local resolver. An OpenURL link points to the copy of the resource most appropriate to the context of the request; if a different context is expressed in the query, a different copy ends up resolved to, but the change in context is predictable, and does not require the creator of the hyperlink to handcraft different URLs for different contexts An OpenURL consists of two parts: a base URL (which addresses the user's institutional link-server) plus a query string (which contains contextual data, usually bibliographic data). An OpenURL is not an identifier in the normal sense, since the same semantic content will have many such OpenURL labels, but the packages are constructs called ContextObjects: since anything may have identity, one can say that the ContextObject is identified by the OpenURL string, but this is not necessarily persistent. Open URLs may become persistent in certain applications: notably the DOI directory is OpenURL-enabled so can recognize a user with access to an OpenURL link resolver. Hence on resolving a DOI, metadata can be pulled from CrossRef to create an OpenURL targeting the current local link resolver. Such an OpenURL link that contains a DOI name is persistent; publishers who use the CrossRef system to identify their content make their products OpenURL-aware. Open URL is a widely used packaged with link resolver systems (both commercial and non-commercial), notably working with CrossRef, to provide a range of library-configured links and services. A main application is the “appropriate copy” problem: an identifier may designate an authoritative version of content at publisher-designated resources, yet a user working in an institution may be subject to a preference context (a local subscription, or an agreed local deal for access to a preferred database).		
Rights coverage	Contextual information carried in an OpenURL package normally may relate to rights information such as access rights to a local subscription copy, mediated by a commercial link resolver system (eg Ex Libris).		
Strengths	Widely implemented to solve the “appropriate copy problem”		
Weaknesses	Not directly relevant to ARROW		

Name	Portable Document Format		
ARROW type	Published content	***Use in ARROW***	Not used (not applicable)
Acronym	PDF	***Reference***	ISO 32000-1
Governance	ISO TC 171/SC 2 (Document management/application issues)		
URL	http://www.adobe.com/devnet/pdf/		
Status	Published 2008	***Implementation***	Extremely widely -- on most PCs
Availability	The standard itself is available from ISO (priced). Originally a proprietary format owned by ADOBE, which still owns various implicated patents. Now freely available to any organisation that wishes to implement applications that create or otherwise make use of PDF.		
Description	A file format developed in the early 1990s to facilitate the sharing of formatted documents between platforms, PDF now has a central role in many document workflows, not least in printing (both desktop and professional). Publishers use PDF as the format for sending files to their printers; as a result, it naturally became the format of choice in the early days of digital publishing (since it required minimal changes to workflow -- although the optimisation of a PDF for printing is not identical to that for online use, so typically publisher still creates subtly different files for these different applications. PDF retains the appearance of the printed page, allowing the creator to stay in control of the visual impact of the page, which may be critical for some types of publication (complex text books, for example). However, this does not always make PDF as easy to read on screen as it might be. If the underlying file is appropriately tagged, there are applications which allow PDF to "reflow" to fit the available screen. Many scanning projects create PDF files. Often these take the form of a "layered" file, where what is presented to the user is a page image (in an image format) with text file behind it created using OCR. This has the advantage of making the text searchable (the OCR file is directly associated with the scanned image) while hiding the results of the OCR process itself (which without human intervention can create files with a fairly high error rate).		
Rights coverage	Metadata (including rights data) can be embedded in or associated with PDF files, but this is dependent on external metadata standards, not PDF itself.		
Strengths	Extremely widely implemented, cross platform format for sharing documents. Allows visual appearance of document to be retained on different viewing and printing platforms.		
Weaknesses	If PDF can be described as having a weakness, it is the obverse of one of its strengths - dependence on the "printed page" paradigm. While "tagged PDF" can be reflowed, PDF is best suited to applications where page format needs to be retained.		

Name	Preservation Metadata: Implementation Strategies		
ARROW type	Metadata – library	***Use in ARROW***	Not used (not applicable)
Acronym	PREMIS	***Reference***	
Governance	Network Development and MARC Standards Office, Library of Congress		
URL	http://www.loc.gov/standards/premis/		
Status	Data Dictionary & Schema v2.0 (2008)	***Implementation***	The PREMIS Implementation Registry at Library of Congress lists just 13 implementing projects.
Availability	The PREMIS Data Dictionary and Schema are freely available for download from the Library of Congress.		
Description	Arising from an OCLC/RLG working group that operated from 2003-2005, PREMIS is concerned with the metadata associated with preserving digital library resources. PREMIS consists of a Data Dictionary and an associated XML Schema for its expression. "The PREMIS Data Dictionary defines a core set of semantic units that repositories should know in order to perform their preservation functions. Preservation functions can vary from one repository to another, but will generally include actions to ensure that digital objects remain viable (i.e., can be read from media) and renderable (i.e., can be displayed, played or otherwise interpreted by application software), as well as to ensure that digital objects in the repository are not inadvertently altered, and that legitimate changes to objects are documented. The Data Dictionary is not intended to define all possible preservation metadata elements, only those that most repositories will need to know most of the time. Several categories of metadata are excluded as out of scope, including: format-specific metadata, implementation-specific metadata and descriptive metadata."		
Rights coverage	The Rights entity aggregates information about rights and permissions that are directly relevant to preserving objects in the repository. Each PREMIS rights statement asserts two things: acts that the repository has a right to perform, and the basis for claiming that right.		
Strengths	Focus on preservation		
Weaknesses	From the ARROW point of view, the focus on preservation		

Name	Publishing Requirements for Industry Standard Metadata		
ARROW type	Metadata - commercial	***Use in ARROW***	Not used (ARROW is entirely focused on books)
Acronym	PRISM	***Reference***	
Governance	IDEAlliance (International Digital Enterprise Alliance)		
URL	http://www.idealliance.org/industry_resources/intelligent_content_informed_workflow/prism		
Status	v.2.1 (2009)	***Implementation***	Primarily in the US magazine publishing industry, where it is believed to be quite widely used.
Availability	Appears to be available without licence, but the website is not explicit on this point.		
Description	The Publishing Requirements for Industry Standard Metadata (PRISM) specification defines an XML metadata vocabulary for managing, aggregating, post-processing, multi-purposing and aggregating magazine, news, catalogue, book, and mainstream journal content. PRISM recommends the use of certain existing standards, such as XML, RDF, the Dublin Core, and various ISO specifications for locations, languages, and date/time formats. In addition PRISM provides a framework for the interchange and preservation of content and metadata, a collection of elements to describe that content, and a set of controlled vocabularies listing the values for those elements. Metadata is an exceedingly broad category of information covering everything from an article's country of origin to the fonts used in its layout. PRISM's scope is driven by the needs of publishers to receive, track, and deliver multi-part content. The focus is on additional uses for the content, so metadata concerning the content's appearance is outside PRISM's scope. PRISM focused on metadata for: • General-purpose description of resources as a whole • Specification of a resource's relationships to other resources • Definition of intellectual property rights and permissions • Expressing inline metadata (that is, markup within the resource itself).		
Rights coverage	Explicitly covers Usage Rights		
Strengths	Widely implemented for magazines, and some implementation for journals		
Weaknesses	From an ARROW point of view, the major weakness is that (so far as we know) it is not used anywhere for book metadata.		

Name	Resource Description and Access		
ARROW type	Metadata - library	***Use in ARROW***	Not used (not yet widely deployed)
Acronym	RDA	***Reference***	
Governance	AACR Committee of Principals		
URL	http://www.rdatoolkit.org/		
Status	Published June 2010	***Implementation***	n/a
Availability	Available online since June 2010 (priced after August 31, 2010)		
Description	"RDA is the new cataloguing standard that will replace AACR2 in 2009. RDA goes beyond earlier cataloguing codes in that it provides guidelines on cataloguing digital resources and a stronger emphasis on helping users find, identify, select and obtain the information they want." "The Joint Steering Committee for Development of RDA is responsible for developing RDA. The JSC consists of representatives from six major Anglo-American cataloguing communities. These include the American Library Association (ALA), the Australian Committee on Cataloguing (ACOC), the British Library (BL), the Canadian Committee on Cataloguing (CCC), the Chartered Institute of Library and Information Professionals (CILIP), and the Library of Congress (LC)." RDA is built on two conceptual models developed by IFLA; Functional Requirements for Bibliographic Records (FRBR) and Functional Requirements for Authority Data (FRAD). The development process recognises that libraries operate in a digital, web based environment and that they wish to exploit strengthening relationships with data creators and users outside the library sector. See also: AACR2, FRBR		
Rights coverage			
Strengths	Built on sound conceptual model foundations		
Weaknesses	In pilot phase. Regarded with some scepticism within the library community perhaps partly because of the time in development and partly because it represents such a seismic shift from the past.		

Name	Resource Description Framework		
ARROW type	Technical protocol	***Use in ARROW***	Not used (not applicable)
Acronym	RDF	***Reference***	N/A
Governance	World Wide Web Consortium		
URL	http://www.w3.org/RDF/		
Status	W3C Recommendations	***Implementation***	N/A
Availability	All W3C standards are freely available.		

Description	Published in its current form in 2004, RDF was developed as a language for representing information about resources on the web. RDF is defined in a series of six W3C Recommendations: Primer; Concepts and Abstract Syntax; Semantics; Vocabulary Description Language (RDF Schema); RDF/XML Syntax Specification; and Test Cases. The idea is to represent information with a semantic graph, which forms a network of connected and identified entities. RDF defines both an abstract language, with defined semantics, and the means to express that language in XML. Other, non-XML forms of expression have been proposed (e.g. Notation3, Turtle) but these have not been standardised. The XML syntax has been criticised for being too verbose. RDF is in fact founded upon an extremely simple idea: that information about a resource can be represented by one or more statements, each containing just three components: a "subject", representing the resource in question; a "predicate", representing a property of the resource; and an "object", representing the value of the property. Since both subjects and objects are resources, complex statements can be built up of sequences of these "triples". One of the simplest applications of RDF is in RDFa, a W3C standard for embedding metadata in XHTML pages. RDFa is used by Creative Commons to embed rights-related information in web pages. Adobe Systems Inc based their Extensible Metadata Platform (XMP) on a subset of RDF; XMP is used for embedding metadata in PDF and other non-text files, and the PLUS Coalition's License Data Format employs XMP to embed license information in photographs. RDF underlies ontological languages such as OWL and SKOS, as well as much of the W3C Semantic Web Activity, and is the basis of the RSS 1.0 web syndication feed language.
Rights coverage	RDF in itself is independent of any application, such as rights expression, but as indicated above, RDF is designed to enable representation of information about web resources, and has been adopted in various forms for the representation of rights-related information.
Strengths	Expresses an extremely powerful model for the representation and manipulation of metadata of all kinds, when linked with other “semantic web” technologies. Basis of continuing developments for vocabulary mapping and interoperability mechanisms
Weaknesses	Still at an early stage of adoption. Technology to exploit the potential of RDF is still emergent.

Name	Representational State Transfer		
ARROW type	Technical protocol	***Use in ARROW***	Not used (SOAP preferred)
Acronym	REST	***Reference***	
Governance	None – not a standard		
URL	http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm		
Status	Not a standard	***Implementation***	Widespread
Availability	No licence required for use		

Description

REST is a general design style or software architecture used when designing distributed software systems. More specifically, it is often viewed as an alternative to the SOAP/WS-* approach when specifying web services, where messages are passed between computers in a request/response 'conversation'.

REST revolves around the concept of a 'resource'. In a computer network, this might be a single fixed document, or a service (maybe a bookings database) that you can interact with.

Resources are associated with verbs and nouns – what do you want to do, and what do you want to do it to? Any resource comes with a set of available verbs and nouns.

With SOAP, an XML document describing the request or response is piggy-backed on top of (usually) an HTTP web request. REST-style requests are embodied in the HTTP request itself, making full use of the features of HTTP. For example, the REST style would use an HTTP GET request to retrieve data from a server, and an HTTP POST request to change that data or delete it. The 'verb' is a pre-defined part of HTTP itself, and the noun – which identifies which data (or 'resource') to retrieve or change – is identified by the remainder of the URI. In a SOAP message, POST or GET is irrelevant, and both the 'verb' and the 'noun' are embedded in the attached XML document.

Technical differences aside, REST is a set of design principles rather than a specific protocol or implementation.

Rights coverage	No direct relevance to rights
Strengths	Succinct, a good match to the overall architecture of the WWW, often allows gradual elaboration of web services without breaking existing client implementations, simple to apply to relatively straightforward services
Weaknesses	Informal, lacks the associated standard practices for implementing security (for example) that accompany SOAP. Extremely difficult to apply properly to complex domains. SOAP vs REST often has the character of a philosophical debate, but it may simply be a matter of scale or maturity

Name	Schematron		
ARROW type	Technical protocol	***Use in ARROW***	Message validation
Acronym	-	***Reference***	
Governance	ISO standard 19575 part 3		
URL	http://www.schematron.com/		
Status	ISO Standard 2006	***Implementation***	implemented in some XML software tools (often in pre-ISO v1.5 form)
Availability	No licence required for use. Specification freely available from ISO		

Description

Schematron is an XML schema definition language, like XSD or RELAX NG. However if differs in that it is pattern- and rule-based rather than constructed around the grammar of an XML document:

- DTDs define the grammar of a class of XML documents – what markup tags can or must be used, their order, and what attributes they can carry
- XSD schemas additionally constrain the data types and content of particular markup elements (so a markup element can be defined as containing a date, or a two digit number, or free text less that 100 characters)
- Schematron validation can also be used to check the data content, but in a way that allows business rules such as 'if that date is in the future, this number must be less than 50'. It provides for checks where there are inter-dependencies between markup or data elements in the XML, which XSD cannot check

Schematron is usually viewed as an adjunct to plain XSL validation of an XML document, as it tends to be verbose when used to define the basic grammar of a class of XML documents. In many implementations, the Schematron rules are first automatically transformed into a long series of XSLT instructions. The XSLT is then applied to the XML document that needs validating: each XSLT instruction checks the document against a specific rule and then outputs an error if it fails.

Schematron validation may also be used to validate particular (rare) types of XML where the document structure is 'non-deterministic', which cannot be validated against standard W3C XSD schemas.

Rights coverage	No direct relevance to rights		
ARROW type	Data representation	***Use in ARROW***	
Strengths	Can express business rules and validate XML documents against those rules		
Weaknesses	Verbose		

Name	SOAP – (formerly Simple Object Access Protocol)		
ARROW type	Technical protocol	***Use in ARROW***	Web services
Acronym	SOAP	***Reference***	
Governance	W3C XML Protocol Working Group		
URL	http://www.w3.org/TR/2007/REC-soap12-part0-20070427/		
Status	V1.2 published 2007	***Implementation***	Widely implemented in common development languages and frameworks
Availability	No licence required for use. Specification and other documents freely available from W3C		
Description	SOAP is a protocol for defining the format for a document (a 'message') passed between two computers, a sender and a receiver. It's also a communication protocol used to exchange such messages electronically. A SOAP 'web service' is an application-specific protocol for exchanging data between two computers using one or more messages defined using SOAP. A SOAP message is an XML document, passed from sender to receiver using (usually) HTTP. For example, the document could contain parameters for a search or the complex details of a purchase order. Another XML document (passed back as a reply to the first) could contain the search result information, confirmation that the purchase order has been accepted, or a error message if the receiver couldn't process the sender's request properly. The required content and structure of the request and response documents – the exact XML elements used to convey the search parameters or the returned data – are defined using SOAP.. Complete web service exchanges can be defined using an allied standard called WSDL. SOAP can be considered a small part of a much larger suite of technical standards for web services termed WS-*. Complete web service exchanges can be defined using an allied standard called WSDL, and there is a suite of other WS-* standards and practices covering for example, security, transaction control etc. Most of these allied standards and practices are administered by OASIS (see http://www.oasis-open.org/committees/tc_cat.php?cat=ws) See also Web Services, REST		
Rights coverage	No direct relevance to rights		
Strengths	Easily implemented using most programming frameworks, and (because it usually uses HTTP or HTTPS for message transfer) can be reliable and secure		
Weaknesses	Perceived as needlessly technical and verbose by supporters of the competing REST approach. Some concerns over performance arise from the use of XML. SOAP vs REST often has the character of a religious debate, but it may simply be a matter of scale or maturity – SOAP is more often associated with enterprise-scale web services with more complex support requirements		

Name	Search and Retrieval via URL/ Search and Retrieve Web service		
ARROW type	Search	***Use in ARROW***	Not used (distributed search not implemented)
Acronym	SRU/SRW	***Reference***	
Governance	SRU Editorial Board, hosted by Library of Congress		
URL	http://www.loc.gov/standards/sru/		
Status	v1.2 (2007). v2.0 draft available from OASIS	***Implementation***	Some experimental implementation
Availability	Specifications freely downloadable from the Library of Congress; no licence required. Some conforming tools available at http://www.loc.gov/standards/sru/resources/tools.html.		
Description	SRU/SRW were originally conceived in 2000 as a pair of query protocols under the aegis of a project from the Z39.50 Maintenance Agency; 'Z39.50 Next Generation.' The intention was to preserve some of Z39.50's abstract query capabilities whilst substituting HTTP, SOAP and other Web protocols (in other words, Web Services technologies) for Z39.50's own communications protocol. SRW is no longer presented as a separate protocol, and is now considered simply a variant of SRU. The Search Web Services Technical Committee of OASIS is currently working on a major revision of SRU 1.2 (and its Contextual Query Language, CQL); http://www.loc.gov/standards/sru/oasis.html. See also: Z39.50		
Rights coverage	SRU does not directly address Rights.		
Strengths	Preserves the rich functionality of Z39.50 in a more lightweight implementation		
Weaknesses	Adoption is still limited even with the library community due to limited support in Integrated Library Management Systems		

Name	Transport Layer Security/Secure Sockets Layer		
ARROW type	Technical protocol	***Use in ARROW***	Messaging security
Acronym	TLS, SSL	***Reference***	
Governance	IETF Network Working Group		
URL	http://tools.ietf.org/html/rfc5246		
Status	v1.2 published in 2008	***Implementation***	Widely implemented in common browser, e-mail, VOIP, e-commerce *etc* software
Availability	No licence required for use. Small cost to acquire an X.509 certificate to authenticate your identity		
Description	TLS and its predecessor SSL are cryptographic protocols that can be built into applications that communicate over the internet, to ensure trusted, secure communication without the risk of interception or modification of the data while in transit. The most familiar implementation of TLS is within the HTTPS protocol, the secure version of HTTP that's used within a web browser. TLS is an application of *public key cryptography*. Cryptographic keys are used to encrypt and decrypt messages. Some types of key come in mathematically-related pairs, where one is used to encrypt and a different key is used to decrypt. Public key cryptography relies on pairs of keys where it is very difficult to work out what the 'other' key is if you have just one. So one of a pair of keys can safely be made public. Now to prove who you are, pick a phrase like your name. Encrypt it using the private key, and tell everyone what the resulting encrypted text is. Anyone wanting to check your identity can use your well-known public key to decrypt the encrypted text – and if their decrypted result is your name, they can be confident you are who you claim to be. Because only you know the private key that matches your public key, only you can construct a text that will decrypt to give your name. TLS uses similar principles to establish the authenticity of at least one of the communicating parties (typically the server). It uses an X.509 Certificate instead of a simple name, but it similarly authenticates the identity of the server. TLS then allows the server and client to exchange a random number that is used to encrypt and decrypt the remainder of the data that is exchanged in a session. X.509 Certificates are issued by Certificate Authorities such as VeriSign, trusted third-party organisations which are the ultimate arbiters of identity.		
Rights coverage	No direct relevance to rights		
Strengths	Almost universally applied, highly secure		
Weaknesses	None of concern to ARROW		

Name	Topic Maps		
ARROW type	Metadata - generic	***Use in ARROW***	Not used (not relevant)
Acronym	N/A	***Reference***	ISO/IEC 13250
Governance	ISO/IEC JTC 1/SC 34 – Document Description and Processing Languages		
URL	http://www.iso.org/ http://www.isotopicmaps.org/		
Status	International Standard	***Implementation***	See http://www.topicmap.com/
Availability	International Standards may be purchased from ISO, Geneva, or through many national standards bodies.		
Description	Topic Maps is a knowledge representation technology. First published as an International Standard in 2000, a second edition was published in 2002. The standard was originally based on SGML. Following publication of the second edition in 2002 it was agreed that the Topic Maps standard should be completely reorganised and re-written as a multi-part standard, and the original SGML-based syntax (HyTM) replaced with an XML-based syntax (XTM). Work on seven parts of the new standard is in progress, with three parts published so far. Work is also in progress on two related standards: a Topic Maps Query Language (TMQL) and a Topic Maps Constraint Language (TMCL). A Technical Report is also being prepared, to show how Dublin Core metadata can be expressed using Topics Maps. A topic map represents knowledge as a collection of statements about topics, which are labels representing abstract subjects. Statements may contain associations between topics and may identify occurrences of these topics in actual resources. Statements about topics may be scoped in order to define the limits of their validity. Topics, associations and occurrences may all be typed. A subject may have an identifier, which should be a URI that enables the subject to be unambiguously identified. There are a number of commercial and open source implementations of Topic Maps and systems using the technology are in live use within businesses in several countries. While the technology has not matured and spread as fast as other knowledge representation technologies, such as RDF, Topic Maps continues to have its devotees among implementers and users, especially in Norway and Germany.		
Rights coverage	Metadata, including rights-related metadata, can be expressed using Topic Maps, but the standard does not specify any particular way in which this should be done. The Technical Report on expression of Dublin Core metadata using Topic Maps is likely to shed some light on this.		
Strengths	ISO Standard with some implementation.		
Weaknesses	Limited implementation. Not directly relevant to ARROW		

Name	UNIMARC		
ARROW type	Metadata - library	***Use in ARROW***	Bibliographic format (expressed in XML) used by some library partners to submit their catalogue to The European Library (TEL) where it is converted into MARC XML
Acronym	UNIMARC	***Reference***	
Governance	IFLA Permanent UNIMARC Committee (PUC)		
URL	http://www.unimarc.net/		
Status	3rd edition 2008	***Implementation***	33 countries use UNIMARC as an exchange format
Availability	Available in print for purchase from Saur Verlag.		

Description

The Universal MARC format, UNIMARC, was created by IFLA in 1977 'with the primary purpose of facilitating the international exchange of bibliographic data in machine-readable form.' Although intended for international exchange, the format has actually been adopted as the national format in a number of countries including France, Italy and Russia.

Like MARC21 and other variants of MARC, the UNIMARC record structure is an implementation of ISO 2709, still expressed using an opaque set of short codes.

The core Bibliographic capabilities of UNIMARC are supplemented by three further UNIMARC formats to handle Authorities information (UNIMARC/A), Classification and Holdings. Although drafts of the Classifications and Holdings formats were disseminated in 2000 and 1999 respectively, they have yet to be formalised.

See also: ISO 2709, MARC 21

Rights coverage

Various pieces of information relevant to determining Rights may be encoded within a MARC record. For example, fields 310 Notes pertaining to binding and availability, 314 Notes pertaining to responsibility, 324 Original version note, 327 Contents Note, etc.

Strengths

Has been widely implemented in some countries, mainly in Europe. See also ISO 2709

Weaknesses

Technically dated. See also ISO 2709

Name	Uniform Resource Identifier / Locator / Name		
ARROW type	Identification	***Use in ARROW***	It is expected that all ARROW internal identifiers will be URI compliant
Acronym	URI, URL, URN	***References***	IETF RFC 1738, 3986, 3305
Governance	IETF, IANA, ICAAN		
URL	ftp://ftp.rfc-editor.org/in-notes/pdfrfc/rfc3986.rfc.pdf ftp://ftp.rfc-editor.org/in-notes/pdfrfc/rfc1739.rfc.pdf		
Status	Stable specifications	***Implementation***	Ubiquitous on the internet
Availability	No licence required for use.		
Description	The terms URL and URI are used – often interchangeably – to describe identifiers associated with resources (usually on the internet). Resources for example can be documents (text, images etc), a source of information such as a network-accessible database, or some other 'service' that provides or acts upon data. The Uniform Resource Identifier is a text string which includes an initial 'scheme' that controls the syntax of the remainder of the URI. For example http and isbn are both acceptable scheme names. The remainder of the URI identifies the resource, in whatever terms are specified by the scheme. So if the scheme is http, the URI can contain a server address (specified via the IP or DNS address of the server), TCP port number, a path name that may end in a filename, and query or fragment sections (starting with ? and # respectively). The URI when used with the familiar http scheme thus combines both identification of a particular resource (often by filename), and a definition of how to access it (ie via the http network protocol). In contrast, a URI such as urn:isbn:978-0-00-729012-3 identifies a particular book, but does not provide any indication of its location or the location of any metadata. This distinction between URIs that define means of access to something, and URIs that simply identify something is relatively clear – at least in principle. The former are termed URLs (universal resource locators), the latter URNs, Uniform Resource Names, and each is a type of URI. But the terminology used is somewhat muddled because of a lack of precision in identification, and confusion over the nature of a 'resource'. It is difficult to explain what an URI such as http://news.bbc.co.uk/ actually identifies, except in fairly abstract terms; it is difficult to view a path/filename incorporated into a URI such as http://upload.wikimedia.org/wikipedia/commons/2/28/EAN-13-ISBN-13.svg) as a reliable identifier when it is so mutable. And it is difficult to explain that http://193.128.166.228:3000/categories/1278.html and http://193.128.166.228:3000/categories/1278.xml are two different URIs that identify the same resource (albeit differing representations of the same resource). In practice, and excepting the most technical discussions, the terms URL and URI are used almost interchangeably. The most commonly encountered URIs (which schemes such as http and ftp) are also URLs.		
Rights coverage	No direct relevance to rights		
Strengths	Ubiquity, resolvability		
Weaknesses	Lack of semantic precision and persistence		

Name	The Virtual International Authority File		
ARROW type	Identification	***Use in ARROW***	Piloting use in ARROW for disambiguating authors/contributors
Acronym	VIAF	***Reference***	
Governance	OCLC in a collaborative project with LoC, DNB and BNF		
URL	http://www.oclc.org/research/projects/viaf/ and http://viaf.org/		
Status	Research project in progress	***Implementation***	N/A
Availability	Access to search the VIAF in beta is available at http://viaf.org/		

Description

From the OCLC website:

"The Deutsche Nationalbibliothek, the Library of Congress, the Bibliothèque nationale de France, and OCLC are jointly conducting a project to match and link the authority records for personal names in the retrospective personal name authority files of the Deutsche Nationalbibliothek (dnb), the Library of Congress (LC), and the Bibliothèque nationale de France (BnF).

• OCLC has proven software for matching and linking authority records for personal names.

• That software will be used to match the authority records from The Deutsche Nationalbibliothek and the Bibliothèque nationale de France to the corresponding authority records from the Library of Congress.

• Once the existing authority records are linked, shared OAI servers will be established to maintain the authority files and to provide user access to the files."

While VIAF is not in itself a standard, nor a project designed to create a standard, it provides the potential underpinning for a standard name identifier (such as ISNI). Some experiments have already been undertaken to match rights management records (from ALCS) with VIAF, and these have proved to be promising in delivering a high proportion of matches.

Rights coverage

Authority files have only limited direct bearing on rights issues, although to the extent that they authoritatively identify the death date of an author, this can provide guidance on whether a specific work is beyond its copyright term.

Strengths

The largest available international resource for disambiguating authors/contributors

Weaknesses

Not a standard (closely involved in the development of ISNI)

Name	Web Services		
ARROW type	Technical protocol	***Use in ARROW***	SOAP-protocol web services used for communication between partners
Acronym	WS, SOAP, REST	***Reference***	See Description
Governance	W3C, OASIS, IETF		
URL	http://www.w3.org/2002/ws/ – http://www.oasis-open.org/specs/		
Status	Various	***Implementation***	N/A
Availability	All W3C Recommendations, OASIS Standards and IETF RFCs are freely available.		
Description	"Web Services" is a portmanteau term for a collection of standards and less formal specifications that define the use the web communication protocol HTTP to enable automated access to processes running on remote servers anywhere on the Internet. Web services are an example of "client-server" computing in which an automated client sends a request message to a remote server and receives a response message in return. There are two competing approaches to the delivery of web services. The first, represented by most of the formal standards in this area, involves the use of XML messages in accordance with the W3C Recommendation for the Simple Object Access Protocol (SOAP). Building on the basic message formats for service requests and response defined by SOAP, W3C and OASIS have between them defined a large number of supporting standards for diverse range of add-on services, including: service definition (WSDL); service discovery (UDDI); service security (WS-Security); service distributed management (WSDM); reliable messaging. All these standards build upon W3C SOAP and the W3C XML technology stack. The second approach is much less formal, and is frequently referred to as REST (acronym for REpresentational State Transfer). Advocates of this approach claim that there is no need for a complex messaging format for requests and responses, and that the same results can be achieved using pre-existing Internet messaging and security standards such as HTTP, HTTPS and SSL. An important principle of REST is that the client making the request should not need to know anything about the internal details of the service in order to make a request. Both approaches are widely implemented in e-commerce and other distributed applications.		
Rights coverage	Web service standards do not specifically cover rights communication, but a web service can deliver data of all kinds, including rights metadata.		
Strengths	Increasingly widely implemented set of protocols for the management of machine to machine communication using Web protocols and the internet as a carrier.		
Weaknesses	None relevant		

Name	Extensible Markup Language		
ARROW type	Technical protocol	***Use in ARROW***	Message syntax
Acronym	XML	***Reference***	XML 1.0 (Fifth Edition)
Governance	World Wide Web Consortium		
URL	http://www.w3.org/TR/xml/		
Status	W3C Recommendation	***Implementation***	N/A
Availability	All W3C standards are freely available.		
Description	XML is a highly successful standard for representing structured data in a serial, plain text-based format. It was developed during the mid-90s as a simplified dialect of ISO 8879:1986 Standard Generalized Markup Language (SGML), specifically for use in web applications, although it is now used much more widely. XML is the basis of many other domain-specific standards for representing structured data of all kinds, from e-commerce transaction formats (e.g. UBL, EDItX) and product metadata (e.g. ONIX) to web syndication feeds (e.g. Atom), complex scientific and technical text and drawings (e.g. MathML, SVG), communication protocols (e.g. SOAP), office file formats (e.g. ODF and OOXML) and programming languages (e.g. XSLT). The XML standard effectively defines two things: basic rules of syntax for the construction and use of markup tags, that all XML applications must folllow; and a Document Type Definition (DTD) language for specifying schemas for sets of XML tags for specific applications. The DTD language is slowly being superseded by other schema languages, but is still widely used, especially in publishing applications. The current (fifth) edition of XML 1.0 is intended to replace both the fourth edition of XML 1.0 and XML 1.1. XML 1.1 aimed to provide better support for Unicode in XML, especially for Chinese, Japanese and Korean applications, but was widely criticised for being technically flawed. There is still controversy about publication of the fifth edition, because it contains new material which many implementers consider to be "breaking changes", i.e. existing software implementations won't be able to process correctly all Fifth Edition documents.		
Rights coverage	XML is a generic data representation language and does not specify or recommend any particular approach to the expression of rights information.		
Strengths	Universal application		
Weaknesses	None relevant		

Name	XSL Extensible Stylesheet Language		
ARROW type	Technical protocol	***Use in ARROW***	Metadata transformation
Acronym	XSL, XSLT and XSL-FO	***Reference***	
Governance	W3C XSL Working Group		
URL	http://www.w3.org/TR/2007/REC-xslt20-20070123/ and http://www.w3.org/TR/2006/REC-xsl11-20061205/		
Status	XSLT v2.0 published 2007. XSL (XSL-FO) v1.1 published 2006	***Implementation***	XSLT v1.0 widely implemented in common XML tools, v2.0 increasingly accepted
Availability	No licence required for use. Specification and other documents freely available from W3C		
Description	Note: XSL consists of two separate but interrelated standards, XSLT for processing and transformation of XML documents, and XSL-FO for XML documents suitable for rendering into other forms for viewing and printing. An XSLT (XSL Transformations) 'stylesheet' is a set of instructions for an XSLT processor, software for transforming an XML document into a different (not necessarily XML) document. The stylesheet consists of a set of templates. Templates contain patterns that are used to select particular elements within the XML document and instructions that act on and modify the markup and content of those elements. Although XSLT can be thought of as a general XML-aware programming language, it is used most commonly for converting data held in an XML document into HTML for output as a web page, or into XSL-FO for further rendering and output (*eg* as printed pages). An XSL-FO (XSL-Formatting Objects) document is an XML document which contains a specification for how it should be presented (primarily in print). Each content element in the XML document has (or inherits) presentational attributes that such as the typeface, size, colour, spacing and so on – the attributes available are similar to those available in the familiar CSS for web pages. In addition, the XSL-FO document includes master page designs with header, footer, margin geometry and so on. XSL-FO rendering software can then be used to print the document on paper. Note that the original XML markup is likely to provide the contextual information to aid interpretation of the data (*eg* where the content is a date, the XML element within which the date is contained tells you that date represents the publication date). This context is lost when the XML document is converted (via XSLT) to an XSL-FO document, and is replaced at least conceptually with information about how publication dates should be printed.		
Rights coverage	No direct relevance to rights.		
Strengths	XSLT widely used for XML to HTML conversion		
Weaknesses	XSL-FO not as accessible as CSS for presentation		

Name	XML Schema Definition		
ARROW type	Technical protocol	***Use in ARROW***	Message validation
Acronym	XSD, WXS	***Reference***	
Governance	W3C XML Schema working group		
URL	http://www.w3.org/TR/2004/REC-xmlschema-0-20041028/		
Status	v1.0 published 2004, v1.1 is available as a draft	***Implementation***	v1.0 is widely implemented in common XML software tools
Availability	No licence required for use. Specification and other documents freely available from W3C		
Description	Note: XSD can apply to both the original W3C Schema definition language (filename rider .xsd), and to schema definition languages in general (eg RELAX-NG, Schematron). Generally, an XML schema is a way of defining formal constraints on the structure and content of an XML document – the data elements that can or must occur in the document, ordering, repetition and nesting of the elements, and the XML attributes that may be attached to particular elements. An XML document can be validated against a particular schema to ensure that it is structurally correct. In addition, a schema can also validate the data type and content of particular data elements, to ensure they are syntactically correct (eg that an element that is supposed to contain a date does contain a date). Schemas also incorporate the concept of XML namespaces. A schema can define a type of XML document based on references to data elements that are themselves defined in other schemas. Namespaces avoid the issue that would arise if data elements in those other schemas used identical names (eg two schemas both used an element called <date>). See also Schematron		
Rights coverage	No direct relevance to rights, but several XML standards referred to in this document are formally defined by XSD or other schema definitions.		
Strengths	Richer way of defining XML documents than the original DTD language defined within the XML standard. DTDs constrain the structure of the XML markup, whereas schemas can also constrain the nature of the content of the XML document		
Weaknesses	Cannot express constraints on the interrelationship of data content between multiple elements in an XML document (which might be termed the 'business rules' of a particular XML application)		

Name	Information Retrieval : Application Service Definition & Protocol Specification		
ARROW type	Distributed search	***Use in ARROW***	
Acronym	Z39.50	***Reference***	ANSI/NISO Z39.50
Governance	Z39.50 Maintenance Agency, % The Library of Congress		
URL	http://www.loc.gov/z3950/agency/		
Status	Z39.50:2003 defines v2 and 3 of the protocol.	***Implementation***	Widespread use in commercial library systems. Limited adoption elsewhere.
Availability	Specifications freely downloadable from NISO and the Maintenance Agency. Functionally equivalent ISO 23950 available to purchase from ISO.		
Description	Z39.50 defines a pre-Web client-server protocol for search and retrieval of information held in remote databases. Most widely used in querying library systems, there has also been some limited adoption in Government and the Environmental community. Z39.50 permits complex queries across diverse underlying databases via an abstracted query syntax that removes the requirement for searchers to understand the structure of the target databases. Significant variations in vendor implementation of Z39.50, combined with ambiguity in the mappings between database indices and Z39.50's abstract terms make searching multiple sources less straightforward than the Protocol's authors intended. Efforts such as the Bath Profile (maintained by Library & Archives Canada) seek to remove some ambiguity by explicitly defining a limited set of common bibliographic queries and the manner in which conformant systems should handle them. SRU/SRW replace Z39.50's own communications protocol with HTTP, and seek to provide some of Z39.50's power in a manner more suited to the Web environment. See also: SRU/SRW, GILS		
Rights coverage	Z39.50 does not deal directly with Rights.		
Strengths	Widely implemented in the library community		
Weaknesses	Complexity of implementation. Little implemented outside Integrated Library Management Systems (ILMS); technically now superseded by SRU/SRW		