

ELAG 2010

Workshop on FRBR and
Identifiers



Programme



- Session 1 Duration: 1:30
 - Introductions of the participants
 - Major questions overview
 - Manifestations
- Session 2 Duration: 1:00
 - Work Ids
 - Work relationships (manifs, reproductions)
 - Expression?, Collection
- Session 3 Duration 1:15
 - Author/ creators Ids
 - Author relationships
 - Linked data requirements
- Session 4 Duration 1:15
 - Summary of conclusions
 - Creation of Presentation

The Major Questions

Workshop on FRBR and
identifiers



Why Identifiers?



- Shorthand for metadata
 - XXX elements are required to establish uniqueness
 - Allows description in multiple forms
- For creating and discovering associations
 - Linked data expresses links; it doesn't create them
 - Identifiers make better links
 - There are good and bad links

Selected Covers



Plautus, Titus Maccius 🇺🇸 🇩🇪 🇺🇸 🇯🇵 🇯🇵

Plaute 0254-0184 av. J.-C. 🇫🇷

Plauto, Tito Maccio 🇪🇸

Plautus, Titus Maccius fl.ca. 254 v. Chr.-184 v. Chr. 🇩🇪

Plautus, Titus Maccius, m. 184 a.C. 🇯🇵

Plautus, Titus Maccius, ca 250-184 f.Kr. 🇸🇪

Plauto, ca 254-184 a.C. 🇵🇹

Plautus, ca 251-ca 184 př. Kr. 🇨🇪

Plautus, Titus Maccius (ca 250-ca 184 a.C.). 🇯🇵

VIAF ID:66462281

Preferred Forms

🇯🇵 [100 1a Plautus, Titus Maccius id \(ca 250-ca 184 a.C.\).](#) 🇯🇵

🇺🇸 [100 1a Plautus, Titus Maccius](#) 🇯🇵

🇸🇪 [100 1a Plautus, Titus Maccius, id ca 250-184 f.Kr.](#) 🇯🇵

🇵🇹 [200 1a Plauto, if ca 254-184 a.C.](#) 🇯🇵

🇩🇪 [100 1a Plautus, Titus Maccius id fl.ca. 254 v. Chr.-184 v. Chr.](#) 🇯🇵

🇨🇪 [100 1a Plautus, id ca 251-ca 184 př. Kr.](#) 🇯🇵

🇯🇵 [100 1a Plautus, Titus Maccius](#) 🇯🇵

🇺🇸 [100 1a Plautus, Titus Maccius](#) 🇯🇵

🇩🇪 [100 1a Plautus, Titus Maccius](#) 🇯🇵

🇪🇸 [100 1a Plauto, Tito Maccio](#) 🇯🇵

🇫🇷 [200 1a Plaute if 0254-0184 av. J.-C.](#) 🇯🇵

🇯🇵 [100 1a Plautus, Titus Maccius, id m. 184 a.C.](#) 🇯🇵



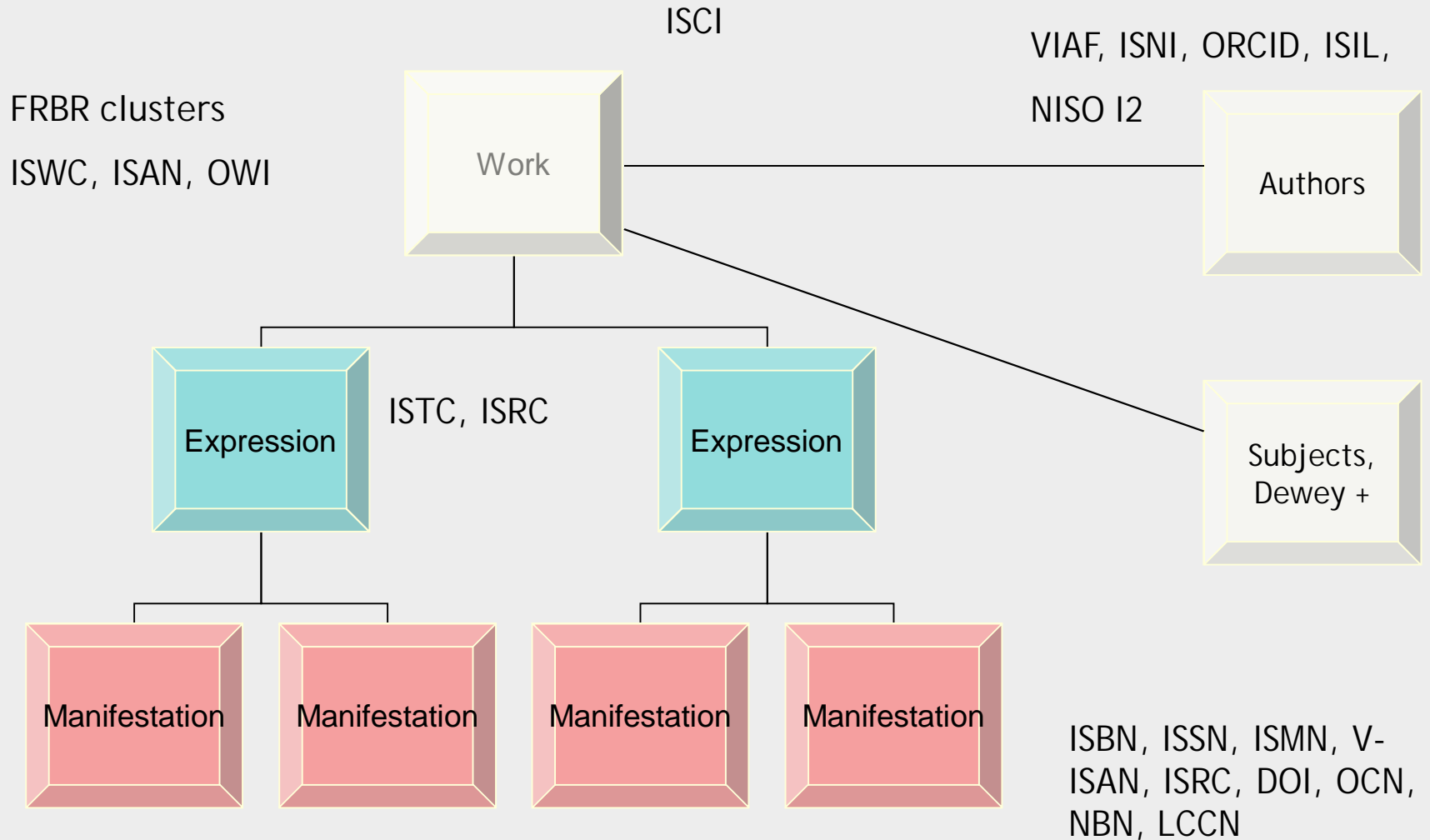
Applications



- Trade
- Discovery
- Rights Management

Are the requirements for
identifiers of these 3 domains
identical?

Family of Identifiers



FRBR questions



How well do existing identifiers adhere to the FRBR model?

Are there requirements not met by the model?

Are there any gaps in identifiers?

Links among identifiers



How well are associations represented in library catalogues?

How well are associations represented in web sites, e.g. Amazon?

What role should libraries play in the maintenance of associations?

More identifier questions



- What is involved in identifier maintenance?
 - Sharing the maintenance burden?
- How well are libraries doing?
 - Using existing identifiers
 - Maintaining and sharing identifiers

Identifiers and Discovery



- Search and browse protocols, e.g. SRU
- Web services, e.g. xISBN, XISSN, xOCLCNUM
- Harvested data
- Linked data

What is the inter-relationship of these and where do identifiers fit?

How well are library identifiers used outside?

How are existing identifiers used in linked data?

Manifestation identifiers

Workshop on FRBR and
identifiers

9-11 June 2010



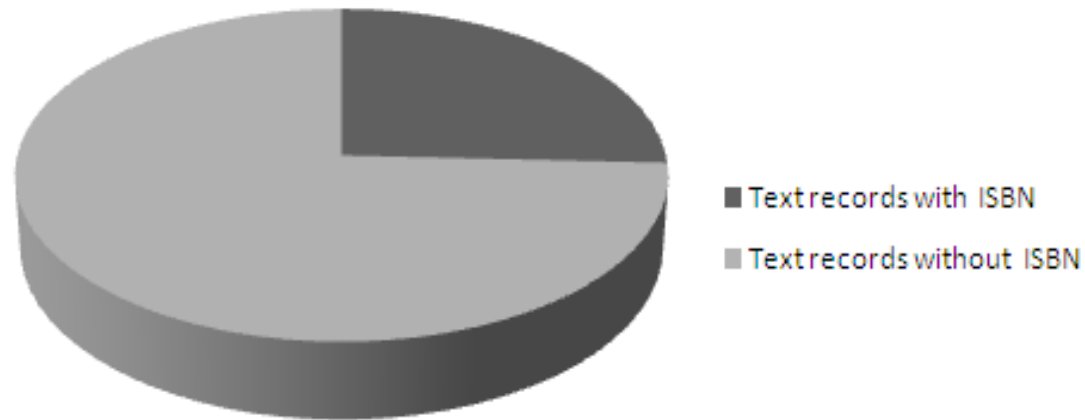
Meeting new user expectations

Management of manifestation identifiers



- Trade
 - Publishers and booksellers need IDs for trading
 - Registration handled by Bowker and Nielsen
- National libraries in some countries
 - Manage distribution of identifiers
- Libraries in general focus on the metadata record identifier - LCCN, OCN, national bibliography identifier

ISBN analysis in WorldCat



- 135 million records in sample
- 104,102,310 text records (008 = "am")
- 77,335,336 records without ISBN (020)

Text records in WorldCat without ISBN



Decade	Text no	
	ISBN	Century
2000 to 2009	5162145	5162145
1990 to 1999	8197844	
1980 to 1989	9956354	
1970 to 1979	9675366	
1960 to 1969	8146606	
1950 to 1959	4863042	
1940 to 1949	3096027	
1930 to 1939	3619221	
1920 to 1929	3191575	
1910 to 1919	2546043	
1900 to 1909	2892532	56184610

Decade	Text no	
	ISBN	Century
1890 to 1899	1812800	
1880 to 1889	1462170	
1870 to 1879	1142666	
1860 to 1869	998529	
1850 to 1859	887630	
1840 to 1849	783051	
1830 to 1839	647781	
1820 to 1829	557528	
1810 to 1819	448374	
1800 to 1809	531816	9272345

Application of manifestation level identifiers



- Does a CrossRef DOI identify a work?

http://www.crossref.org/CrossTech/2010/02/does_a_crossref_doi_identify_a.html

- *“reader has a reasonable expectation that when they follow a citation, they will be taken to what the author saw”*

- Using identifiers to facilitate the e-book supply chain by Brian Green

http://www.niso.org/apps/group_public/download.php/3684/contentid_print.pdf

- *“ Each different format of an electronic publication (e.g. ‘.lit’, ‘.pdf’, ‘.html’, ‘.pdb’) that is published and made separately available shall be given a separate ISBN.”*



024 indicator 1	Contents	Count
blank	invalid	52
	invalid	1,604
0	ISRC International standard recording code	32,518
1	UPC	741,734
2	ISMN International standard music number	105,967
3	EAN International article number	113,481
4	SICI	1,740
7	Source specified in subfield 2	616,330
8	unspecified	6,253,128
		7,866,554
024 indicator 2		
blank	no information	7,459,460
0	no difference	949
1	difference between coded identifier and display equivalent	56
		7,460,465

Figures from WorldCat excluding ISBN and ISSN

6 of 7 million unspecified!!

WorldCat by type of material



	Mono component	Serial component	Collection	Subunit	Integrating resource	Monograph	Serial	Total
Text	3415130	2640	332870	1341666	102710	139081491	7349884	151626391
Music	14108	96	2891	36010	17	4231440	2791	4287353
Mss music	438	0	1752	5180	0	130292	0	137662
Maps	36943	30190	5658	171081	135	2584493	8922	2837422
Mss maps	1768	0	314	439	0	49401	0	51922
Projected	3395	38	19200	12748	48	4289639	2819	4327887
Sound recording non musical	1815	51	3546	1847	16	1312886	1911	1322072
Musical sound recording	5102	12	5036	19111	15	3999758	560	4029594
2d graphic computer file	21814	10	88630	133043	72	1088267	91	1331927
kit	2648	1250	849	17993	1573	1231855	77946	1334114
mixed	34	0	1621	10	1	76291	259	78216
3d obkect	818	0	899659	52126	0	15546	1	968150
mss text	264	0	456	245	0	80791	8	81764
all	4959	14	85660	55581	0	7667623	0	7813837
	3509236	34301	1448142	1847080	104587	165839773	7445192	180228311



Resource Type	Records in WorldCat	# ISO identifiers
Music	2,889, 344	105,967
Sound recording	5,351,666	32,518



- <http://xisbn.worldcat.org/xisbnadmin/xoclcnum/api.htm>
- <http://xisbn.worldcat.org/webservices/xid/oclcnum/55847258?method=getEditions&format=xml&fl=lccn,owi,isbn>
- <http://xisbn.worldcat.org/webservices/xid/lccn/2004273129?method=getEditions&format=xml&fl=lccn,isbn,owi>

Work Level identifiers

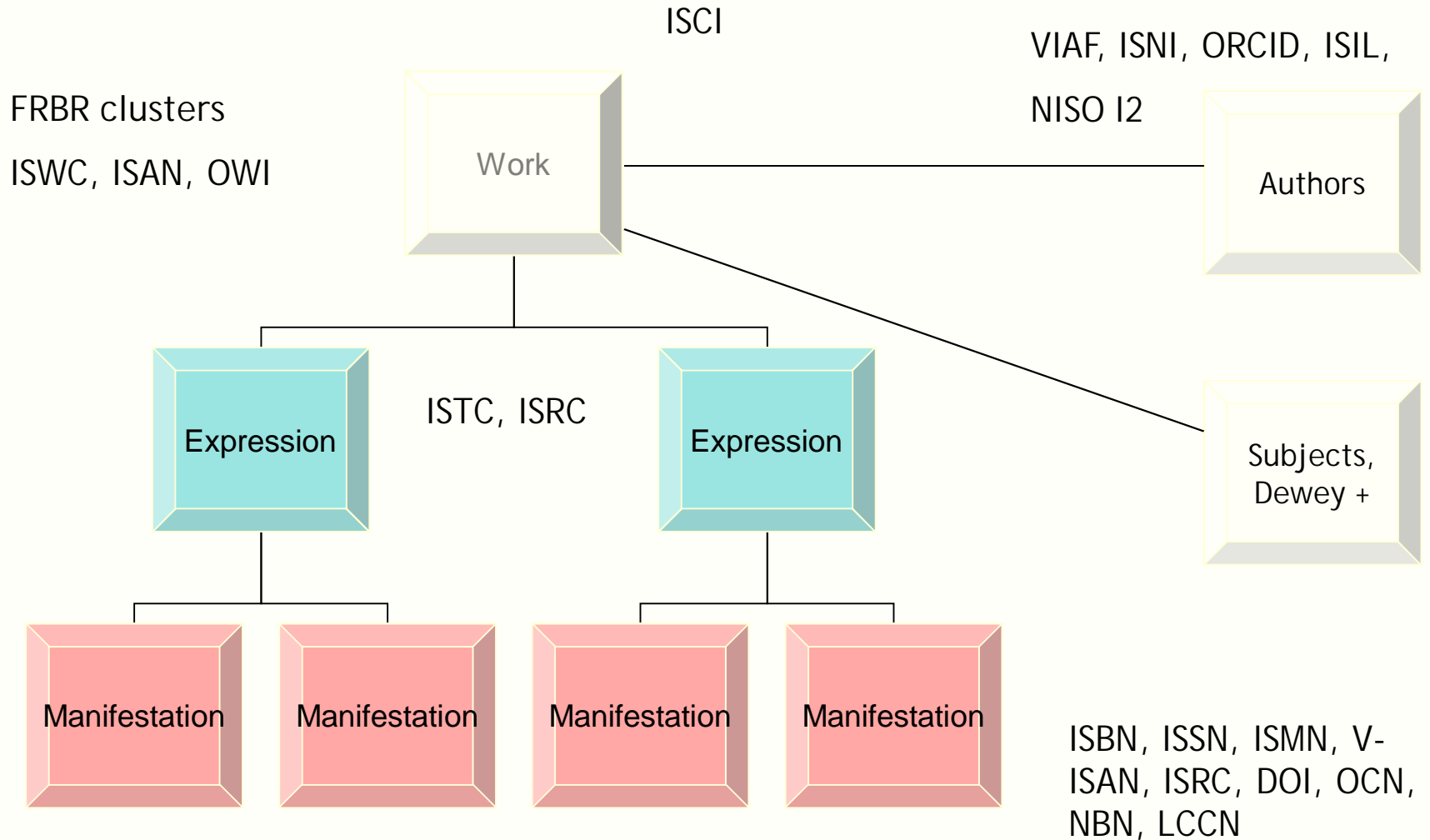
Workshop on FRBR and
identifiers

9-11 June 2010



Meeting new user expectations

Family of Identifiers



Work Level Identifiers



- Collocating identifiers
- Not easy to manage in a distributed way as manifestation identifiers (block to each publisher administered nationally)
 - Same work is published in multiple countries
 - Manifestations span time
- Assignment needs to be retrospective

Management of Work Level identifiers



- Trade
 - Value to publishers less clear
 - Useful for grouping their own versions
- Discovery and Rights Management
 - Greater need
 - Example: Arrow Project - cooperation Libraries and Rights Management organisations ISTC/ISNI

Are these represented in catalogues?



- <http://www.bisg.org/contentweb/papers/the-international-standard-text-code-istc-a-work-in-progress/>
- http://www.isan.org/portal/page?_pageid=168,1&_dad=portal&_schema=PORTAL





- OCLC exposes work identifiers July 2009

<http://community.oclc.org/metatalogue/archives/2009/07/oclc-exposes-work-identifiers.html>

- <http://xisbn.worldcat.org/webservices/xid//owi/owi718389?method=getEditions&format=xml&fl=lccn,isbn,owi>

Work / Manifestation links



How well are associations represented in library catalogues?

How well are associations represented in web sites, e.g. Amazon?

What role should libraries play in the maintenance of associations?

Do we need expression level identifiers?

Workshop on FRBR and identifiers

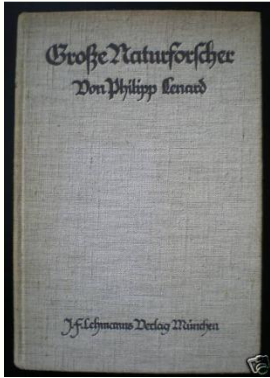


Other relationships

Workshop on FRBR and
identifiers



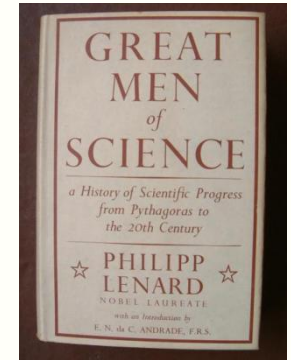
Work



Original in German

Expression

English translation

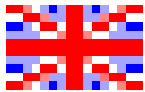


München:
Lehmann, 1930
[2nd ed]

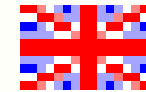
Manifestation

NY: Macmillan,
1933
[2nd ed]

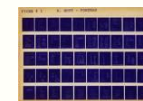
Parallel



descriptions



Reproductions



1991

2003

2009

Reprints and Reproductions



- Electronic reproductions
 - Google
 - Hathi
- Microform
- Photocopy
- Reprints that aren't new editions

- Presented to the user as possible alternatives



- International Standard Collection Identifier (ISCI) ISO DIS 207730 - reached final voting
 - Discovery
 - Collective collection management (virtual collections)

Creator identifiers

Workshop on FRBR and
identifiers



Management of Creator identifiers



- Again collocating identifiers
 - Needs centralised management of some kind
 - Authors change publishers, are associated with multiple countries

To watch



- VIAF
 - Library personal name authority files from 12+ national sources
- ISNI
 - Creators: authors, publishers, performers, composers, editors, illustrators
- ORCID
 - Researchers; receivers of research grants, authors of articles and theses
- NISO I2
 - Institutions involved in the supply chain

ISNI consortium data sets: *European resource*



- VIAF (Virtual International Authority File)
 - 14 million records, 12 + countries
 - Harvested and maintained by OCLC; matching developed over 5-6 years
- IPI (Interested Party Identifier)
 - 2.5 million records, mostly music and sound creators
 - Managed by CISAC (also involved with ISTC, ISAN, ISWC, ISRC)
- IPDA (International Performers Database Association)
 - Separate databases

3. IS YOU IS OR IS YOU AINT MA BABY	ISWC: T-070.266.534-6 Durat
Creator(s) AUSTIN BILLY JORDAN (US 1) LOUIS	Creator-Id 00001623636 00077343267
Performer(s) ANDREW SISTERS ARMY AIR FORCE BAND ATWOOD EDEN AUSTIN-JORDAN W BILLY BANJO BAND BROWN B BUTTERFIELD B BUTTERFIELD B/ORCHESTRA DAMONE V GARNER E HARRIS B HAWKINS J JACKSON J JORDAN L MAJOR MILLER G MAJOR MILLER G/ARMY AIR FORCE BAND MOORE M WILLIAMS C	

Proof of Concept: sample results



Identifier	ISNI assigned	VIAF URL	Confidence level	Matching Criteria
ALCS 1094842	0000 1000 0035 481x	http://viaf.org/60767865	Title: 0.86	Cooke, Andrew 1960 "economics of leisure & sport the"
Adami 1147	0000 1000 0000 051x	http://viaf.org/9852975	Title: 1.00	FREDERIX, FRANCOISE "le pot de terre et le pot de fer"
Bowker 993642	0000 100 0037 1425	http://viaf.org/53397582	Title:0.93	Herman, R.A. "littlest pumpkin"
Prolitteris 001569267	0000 100 0037 9929	http://viaf.org/60010769	Partial date and title: 0.60	GEISEL, SIEGLINDE 1965- « hutterer », « hutterite »
Suisa 0023610120	0000 1000 0048 7970	http://viaf.org/46976221	Date: True 1959, 2008	SILBERBAUER, NORBERT 1959-2008
Suisa 0002575946 8	0000 1000 0042 0645 8	http://viaf.org/37105079	Title: 1.00	RENIS, TONY 1938- "quando, quando, quando"

Proof of Concept: Statistics

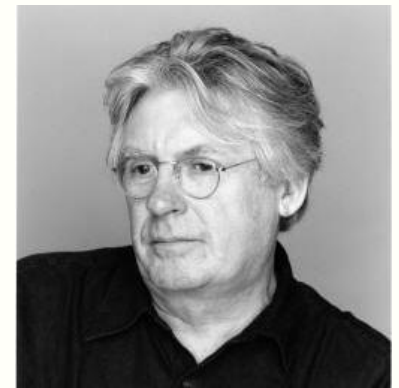


	Sample	Matched	Percentage	ISNIs assigned
IFRRO (ALCS)	311,046 records 35,088 contacts 43,436 identities	33,918 (Unmatched 9528)	78.09%	31,733
IPDA (Adami)	78,892 records 860 identities	506 (unmatched 354)	58.83%	505
Bowker	1090 identities	882	80.92%	879
IFRRO (Prolitteris)	3228 identities	2341	72.52%	2341
CISAC (Suisa)	1295 identities: (847 with titles, 347 with 2 dates)	299	23.09%	299

David Harsent & public identities



- David Harsent, born 9.12.1942
 - Jack Curtis Name id 2016649
 - Oliver Dalton Name id 2915469
 - Francis Greig Name id 2106650
 - David Lawrence Name id 2915470
 - Sam Lawrence Name id 1021214
 - David Pascoe Name id 2117783





Cold kill	ALCS	LC	DNB		OCLC
Dead sit round in a ring	ALCS	LC	DNB		OCLC
Der kreis der toten			DNB		OCLC
Down into darkness	ALCS	LC			OCLC
Geruch des todes	ALCS		DNB		OCLC
Nothing like the night	ALCS	LC	DNB		OCLC
Quatre morts assis en rond	ALCS			BNF	
Vier doden in een kring	ALCS				OCLC

- VIAF reveals titles and translations to ALCS for which perhaps it may be missing collection of royalties, e.g. *Der kreis der toten*.
- ALCS reveals translations, e.g. *Vier doden in een kring*



BNF Record for 20041125afrey0103 ba0

Fixed		
<u>001</u>		BNF FRBNF145845587 09:15:26.0
<u>100</u>		20041125afrey0103 ba0
<u>101</u>		eng
<u>102</u>		GB
<u>103</u>		19421209
<u>105</u>		a
<u>106</u>		1
<u>120</u>		b
<u>152</u>		AFNOR Z 44-061 ‡c 2
<u>200</u>		‡7 ba0yba0y ‡8 fre ‡9 0 ‡a Pascoe ‡b David ‡f 1942-....
<u>300</u>	0	Romancier. - Scénariste pour la télévision et le cinéma
<u>801</u>		FR ‡b BnF ‡c 20041125
<u>810</u>		Fugitive / David Pascoe, 2004 ←
<u>810</u>		LC authorities - http://authorities.loc.gov (2004-11-25)
<u>901</u>		222615388 ‡9 1
<u>901</u>		229034974 ‡9 1
<u>910</u>		fugitive ‡9 1

ALCS data will assist in correctly linking these 3 records & splitting the DnB authority

ORCID (Open Researcher Contribution Identifier)



- Initiative of Nature and Thomson Reuters
- 6.5 million article creators; 55,000 from Thomson Reuters
- Talking about 2,000,000 USD to start up
- More than 60 institutions: American Institute of Physics, American Psychological Association, Association for Computing Machinery, British Library, CrossRef, Elsevier, European Molecular Biology Organisation, Hindawi, INSPIRE (project of CERN, DESY, Fermilab, SLAC), Massachusetts Institute of Technology Libraries, Nature Publishing Group, Public Library of Science, ProQuest, SAGE Publications Inc., Springer, Thomson Reuters, University College London, University of Manchester (JISC Names Project), University of Vienna, Wellcome Trust and Wiley-Blackwell.
- Will include user interface for corrections, additions

ISNI and ORCID



- Model 1 ORCID is an ISNI data source and Registration Agency
 - ORCID and ISNI synergetic with ORCID providing user front end and ISNI providing identifier and Cross domain linking
- Model 2 ORCID is at arms length of ISNI
 - VIAF is a reference file for both ORCID and ISNI

Conservative estimates:
ISNI market 25 million; ORCID 7 million
(30-40% of ORCID is covered by JISC names)

Creator ID questions



How well are library IDs used outside?

Roles of libraries in maintenance of creator IDs

Are the requirements for identifiers of trade, discovery, rights identical?



How well are associations represented in library catalogues?

How well are associations represented in web sites, e.g. Amazon?

What role should libraries play in the maintenance of associations?

Closer Look at Identifiers and Linked Data

Workshop on FRBR and identifiers



Linked data expresses links



- Links must first be created
 - Better with shared identifiers
- Example VIAF / Wikipedia links
 - Made by harvesting Wikipedia then searching for authors
 - Metadata comparison is done to establish a link
 - Link is then encoded and rendered as linked data
 - Process would be simpler and more reliable with common identifiers

Linked data examples: VIAF



- <http://viaf.org/viaf/99366184> is a `viaf:NameAuthorityCluster`,
- <http://viaf.org/viaf/99366184/#skos:Concept> is a SKOS Concept
- and <http://viaf.org/viaf/99366184/#foaf:Person> is a FOAF Person.

DOI and linked data



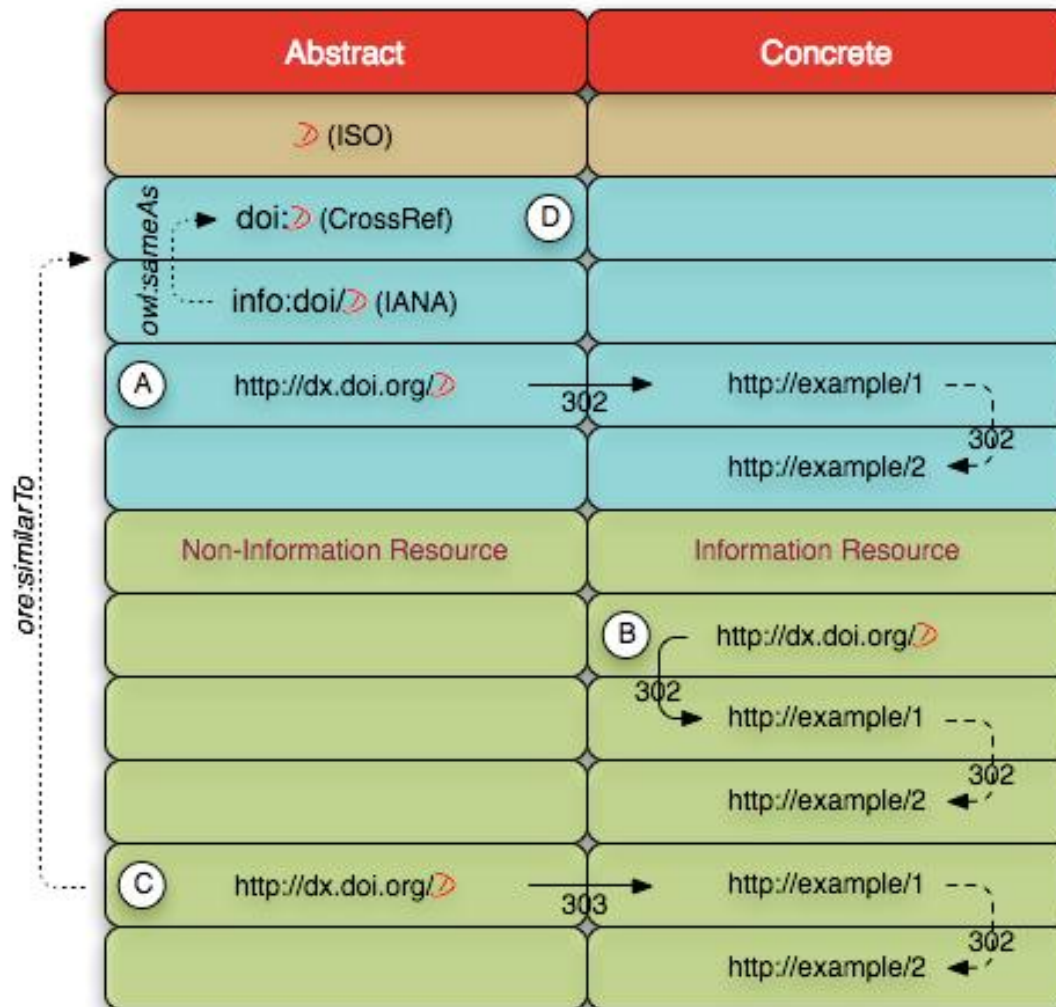
DOI:
What
Do
We
Get?

Ⓓ = DOI

Generic
Data

Web
Data

Linked
Data





- http://www.crossref.org/CrossTech/2010/02/doi_what_do_we_got.html

Linked data examples: Hungarian National Library



- OPAC records (RDFDC):

<http://nektar.oszk.hu/resource/manifestation/2645471>

Digital Library record (RDFDC):

<http://oszkdk.oszk.hu/resource/DRJ/404> Name authority

(FOAF): <http://nektar.oszk.hu/resource/auth/33589>

Subject term (SKOS):

http://nektar.oszk.hu/resource/auth/magyar_irodalom

Identifiers and Discovery



- Search and browse protocols, e.g. SRU
- Web services, e.g. xISBN, XISSN, xOCLCNUM
- Harvested data
- Linked data

What is the inter-relationship of these and where do identifiers fit?

How well are library identifiers used outside?

How are existing identifiers used in linked data?

Workshop on FRBR and Identifiers

Plenary Presentation 11.06.2010



Programme



- Session 1 Duration: 1:30
 - Introductions of the participants
 - Major questions overview
 - Manifestations
- Session 2 Duration: 1:00
 - Work Ids
 - Work relationships (manifs, reproductions)
 - Expression?, Collection
- Session 3 Duration 1:15
 - Author/ creators Ids
 - Author relationships
 - Linked data requirements
- Session 4 Duration 1:15
 - Summary of conclusions
 - Creation of Presentation

Why Identifiers?

- Shorthand for metadata
 - XXX elements are required to establish uniqueness
 - Allows description in multiple forms
- Discovery:
 - Linked data, Harvesting, Web services, protocol enquiry
- Trade: Publishers, aggregators, sellers
- Rights Management:
- Long term preservation, especially of digital content



Conclusions: Identifiers and FRBR



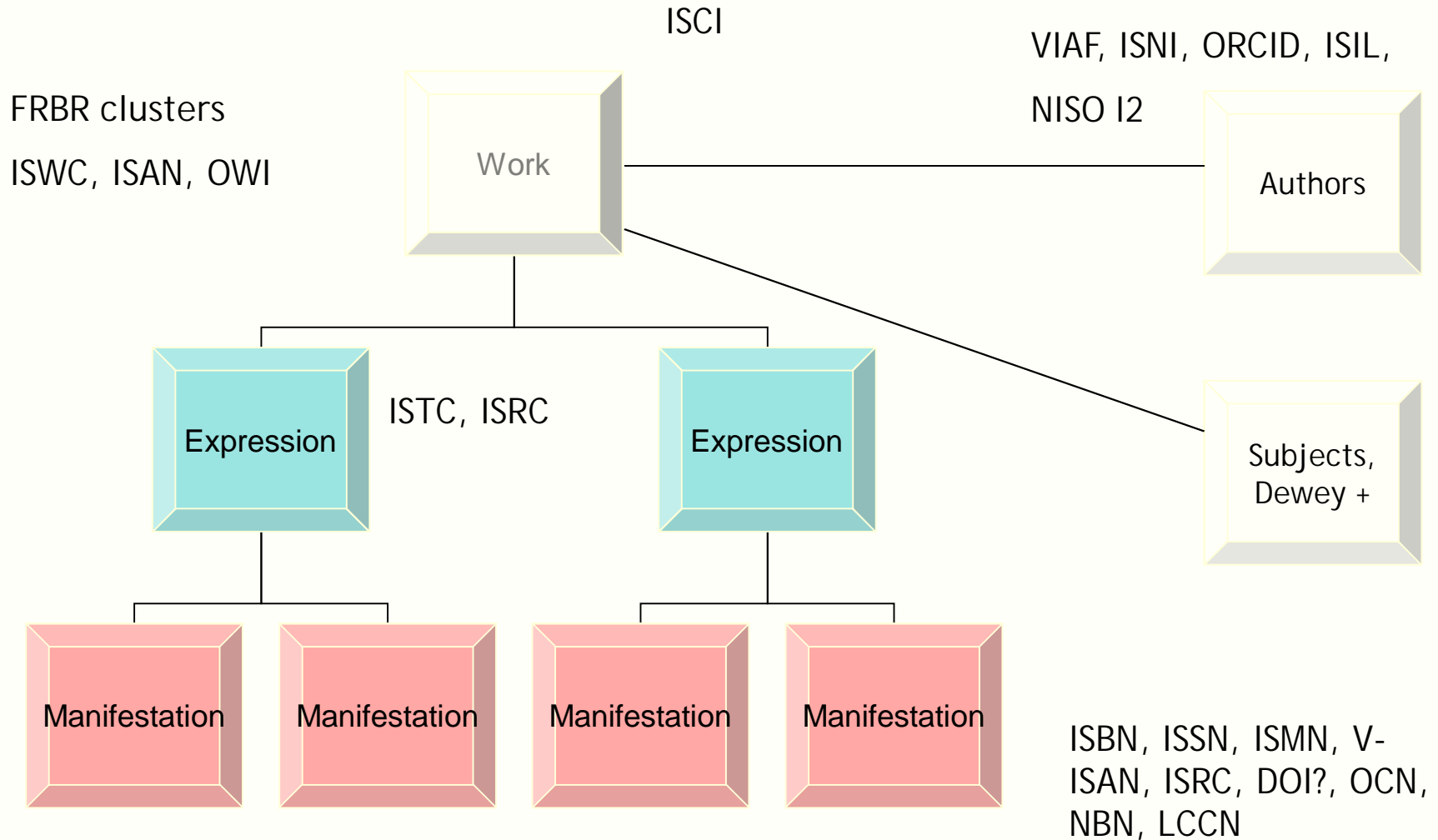
- We have or will soon have appropriate identifiers for most of FRBR entities
 - Need for item identifiers ? Need for electronic content format negotiation? Needs further study
 - Need system/s to relate, e.g. creators to works, items to manifestations
 - Some manifestations, e.g. still images
- They don't all fit neatly to one entity (ISTC is work/expression; ISBN not always manifestation; DOI can be assigned at any level because ...)
- Some overlap

Conclusions: Identifiers and FRBR



- Can work better with what we currently have
 - Libraries using ISBN & ISSN consistently better than other identifiers
- Need for library identifiers as well as trade
 - 74% monographic texts lack ISBN
 - Libraries can contribute well for collocating identifiers - works and creators
- Need some additional groupings that are not explicit in FRBR - reproduction and reprint to original

Family of Identifiers



Conclusions: Identifiers and Linked Data



- Linked data expresses links; it doesn't create them
 - Shared Identifiers make better links
 - There are good and bad links
- Need to understand the relationship of identifiers within linked data and how they serve interoperation among
 - Linked data
 - Harvested data
 - Search protocols and APIs

Further investigation



- How to get IDs into Wikipedia
- How to diffuse IDs into:
 - Library catalogues, especially ISO IDs that are not ISBN
 - Embed in electronic publications
 - Workflow optimisation