

IDENTITY MANAGEMENT ANALYSIS: AN EMPIRICAL INVESTIGATION INTO THE STATE OF LIBRARY COMMUNITY'S AUTHORITY DATA CONFORMANCE TO THE NEW STANDARD

Oksana L. Zavalina and Vyacheslav Zavalin

Department of Information Science, University of North Texas



ICKM 2017



Identity management

Contributes to KM through supporting effectiveness of:

- **Systems**
 - access to information, user friendly services
- **People:**
 - Culture of openness and sharing

NASA
knowledge
strategy
framework

Focuses on:

- **creation**
- **conversion**
- **sharing, and**
- **maintenance of identity data**

Called **authority data** in the cultural heritage institutions (including libraries)

Identity management

in the library community

Mostly
in
MARC
format

Authority data --
standardized **machine-
readable records** that
describe:

- persons 
- institutions 
- places 
- events 
- works, etc. 
- *relations between them*

Extensive and rapidly growing databases, e.g.:

- US Library of Congress
 - **Name Authority File (NAF):** over 10 M records, 22 % growth between 2011 & 2014
 - **Subject Authority File (SAF)**
 - Concepts
 - Objects
- Virtual International Authority File (VIAF), etc.

Details

Visualization

International Conference on Knowledge Management

URI(s)

- > <http://id.loc.gov/authorities/names/n2006023404>

Instance Of

- > [MADS/RDF ConferenceName](#)
- > [MADS/RDF Authority](#)
- > [SKOS Concept](#) 



Scheme Membership(s)

- > [Library of Congress Name Authority File](#)

Collection Membership(s)

- > [Names Collection - Authorized Headings](#)
- > [LC Names Collection - General Collection](#)

Variants

- >  Conference on Knowledge Management, International
- >  ICKM

Exact Matching Concepts from Other Schemes

- > <http://viaf.org/viaf/sourceID/LC%7Cn+2006023404#skos:Concept> 



Example of meeting
authority record (LC NAF)

Earlier Established Forms

[IKMS International Conference on Knowledge Management](#)

Sources

- > found: Knowledge management, c2005: t.p. (International Conference on Knowledge Management; Knowledge Management; ICKM2005; organized jointly by the Information and Knowledge Management & Technology (ASIST))
- > notfound: People, knowledge and technology, c2004: t.p. (1st IKMS International Conference on Knowledge Management)

Editorial Notes

- > [THIS 1XX FIELD CANNOT BE USED UNDER RDA UNTIL THIS RECORD HAS BEEN REVISED]

Change Notes

- > 2006-03-28: new
- > 2012-08-16: revised

Alternate Formats

- > [RDF/XML \(MADS and SKOS\)](#)
- > [N-Triples \(MADS and SKOS\)](#)
- > [JSON \(MADS/RDF and SKOS/RDF\)](#)
- > [MADS - RDF/XML](#)
- > [MADS - N-Triples](#)
- > [MADS/RDF - JSON](#)
- > [SKOS - RDF/XML](#)
- > [SKOS - N-Triples](#)
- > [SKOS - JSON](#)
- > [MADS/XML](#)
- > [MARC/XML](#)

This record
needs
revision to
conform to
RDA standard

Resource Description and Access (RDA)




- New standard for metadata, including authority data in library community
- Developed since 2008, officially implemented in 2013
 - replaced previous standard AACR
- Intended to greatly improve functionality of authority data:
 - focus on representing important attributes and relations
 - for Linked Data / Semantic Web development
- Introduces a number of new data elements in authority records, e.g.:
 - 35 new MARC fields for name and/or title authority records overall
 - 7 new MARC fields for meeting name authority data
 - 5 new Linked Data enabling MARC subfields

e.g., **370**
Associated Place

e.g., MARC
field **377**
Associated Language

e.g., **#u** *Universal Resource Identifier (URI)*

Problem Statement / Research focus

- Only 3 published studies analyzed authority data in relation to RDA guidelines
 -  • 2 focused on personal name authority records (*Moulaison, 2015; Thompson, 2016*)
 - either small sample of records or a subset of data elements
 - 1 study (*Kimura, 2015*) focused on authority data created in China, Japan, & Korea
 - over 1M of name authority records of 3 kinds: personal, corporate, and meeting
 - **mostly non-RDA-based authority data**
- **Shortage of research evaluating results of implementation of RDA in authority records**
 - **across various kinds of authority records**
 - **for the whole spectrum of data elements, and**
 - **with large samples or Big Data approach**



Purpose of the study

Explore the state of implementation of RDA standard in **all 5 kinds of authority records represented in US Library of Congress NAF**



- corporate



- geographic names



- meeting



- personal, and

- uniform (preferred) titles for works



Research Questions

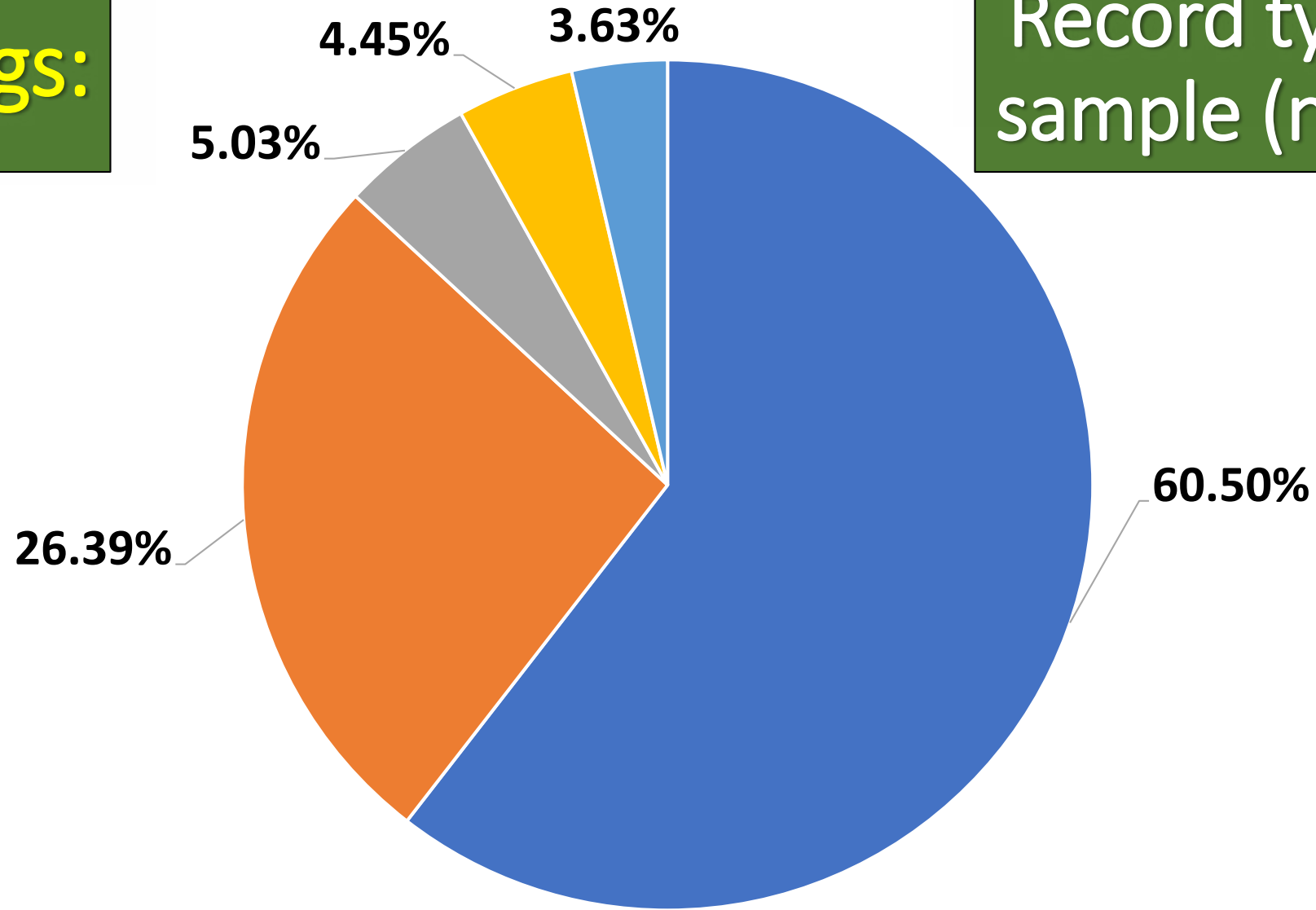
- What elements are receiving the most attention from the record creators or editors?
- How are the Linked Data enabling elements of RDA applied in the existing identity data?

Methods

- Quantitative content analysis of authority records
- Intended to apply Big Data analytics approach and collect all RDA-based authority records in NAF
 - A total of **1.2 M** (as of March 2016)
- Record harvesting software limitations resulted in collecting **408.5 K** records
 - **34%** of the dataset
 - Large random sample
 - Representative sample: all 5 types of NAF authority records harvested

Findings:

Record types in the sample (n=408,551)



■ meeting name ■ personal name ■ corporate name ■ uniform title ■ geographic name

Example of RDA-based name authority record in MARC

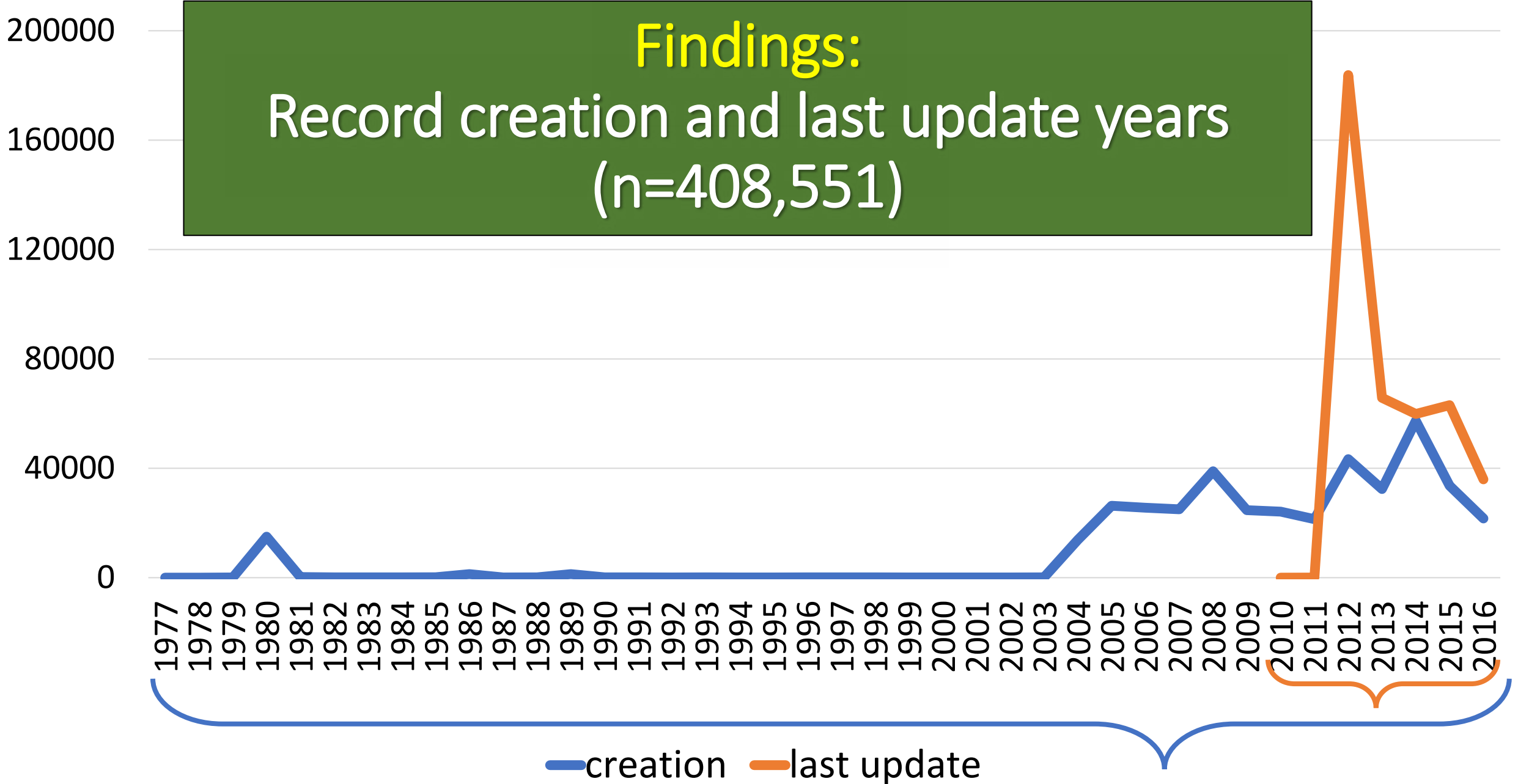
This meeting name record was originally created in AACR standard, edited later to conform to RDA

ARN	6904209				
Rec stat	c	Entered	20060404	Replaced	20170304073758.0
Type	z	Upd status	a	Enc lvl	n
Roman		Ref status	a	Mod rec	
Govt agn		Auth status	a	Subj	a
Series	n	Auth/ref	a	Geo subd	n
Ser num	n	Name	n	Subdiv tp	n
010	nb2006008174				
040	Uk †b eng †e rda †c Uk †d DLC †d IEN †d WaU				
046	†s 20120727 †t 20120812				
111	2 Olympic Games †n (30th : †d 2012 : †c London, England)				
368	Multi-sport event †a Sporting event				
368	Sports tournaments †2 lesh				
370	†c Great Britain †e London (England) †2 naf				
377	eng				
411	2 Games of the XXX Olympiad †d (2012 : †c London, England)				
411	2 Games of the 30th Olympiad †d (2012 : †c London, England)				
411	2 Games of the Olympiad †n (30th : †d 2012 : †c London, England)				
411	2 Olympiad, Games of the †n (30th : †d 2012 : †c London, England)				
411	2 London 2012 (Olympics) †d (2012 : †c London, England)				
411	2 Summer Olympics †n (30th : †d 2012 : †c London, England)				
411	2 London Olympic Games †d (2012 : †c London, England)				
411	2 London Summer Olympics †d (2012 : †c London, England)				

RDA-specific MARC fields

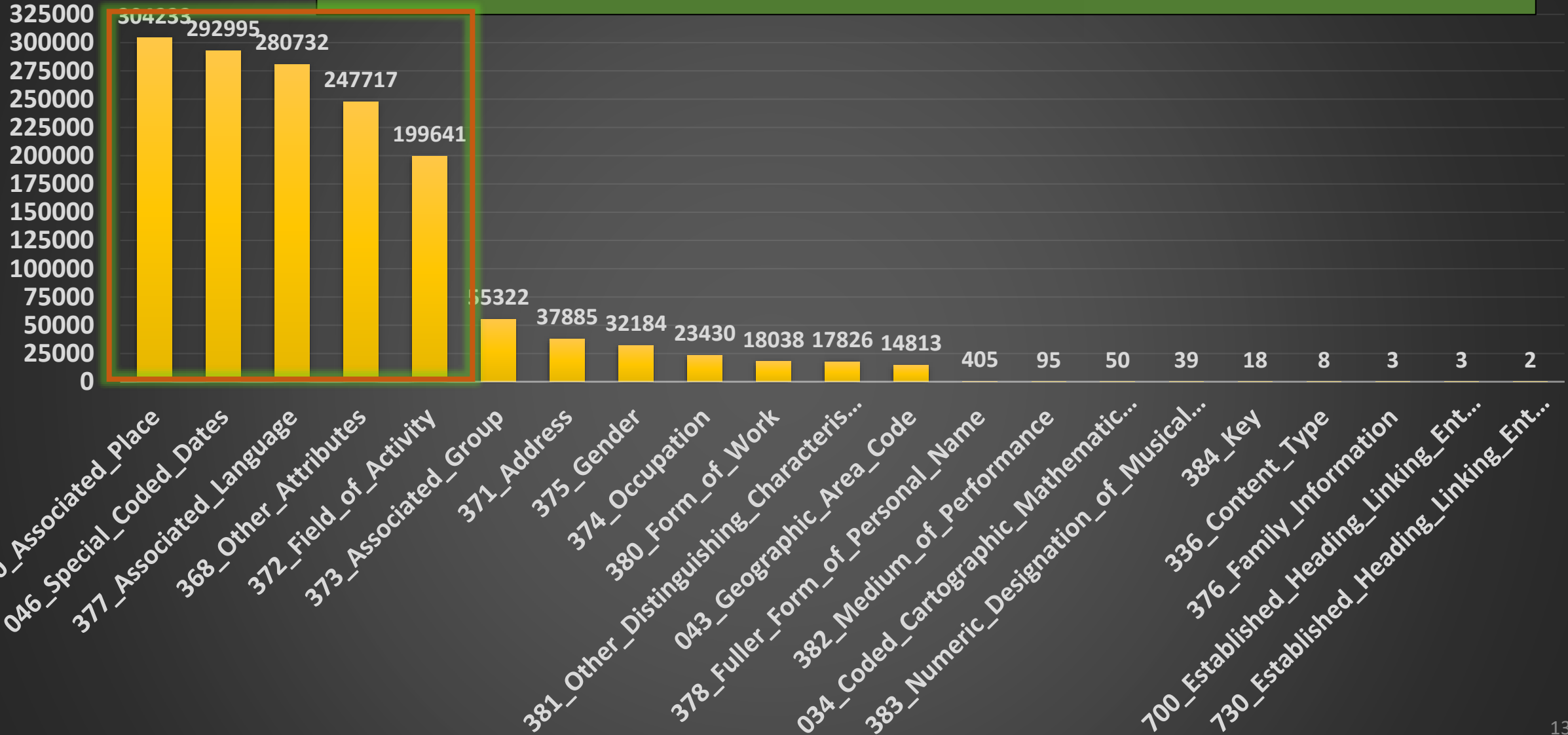
670	Olympic games impact study, 2005.
670	Olympic organisation website, 4 Apr. 2006 †b London page (London 2012 Games of the XXX Olympiad to be held 27 July to 12 Aug. 2012)
670	Programme procurement in construction, learning from London 2012, 2013: †b galley (The Games of the 30th Olympiad held in London during the summer of 2012)
670	Wikipedia, Feb. 17, 2013 †b (2012 Summer Olympics; The 2012 Summer Olympics, officially the Games of the XXX Olympiad and commonly known as London 2012, was a major international multi-sport event celebrated in the tradition of the Olympic Games, as governed by the International Olympic Committee (IOC). It took place in London, United Kingdom, from 27 July to 12 August 2012. The first event, the group stage in women's football, began two days earlier, on 25 July)
670	Olympic Games 27 July-12 August official London 2012 website, viewed Feb. 17, 2013: †b home page (Olympic Games; London 2012 Olympic and Paralympic Games) †u http://www.london2012.com/
670	Olympic.org website, Feb. 17, 2013: †b Olympic Games > London 2012 (London 2012; Olympic Games; Opening date: 27 July 2012; Closing date: 12 August 2012; Country of the host city: Great Britain; HTML headoer: London 2012 Summer Olympics) †u http://www.olympic.org/london-2012-summer-olympics

Findings:
Record creation and last update years
(n=408,551)



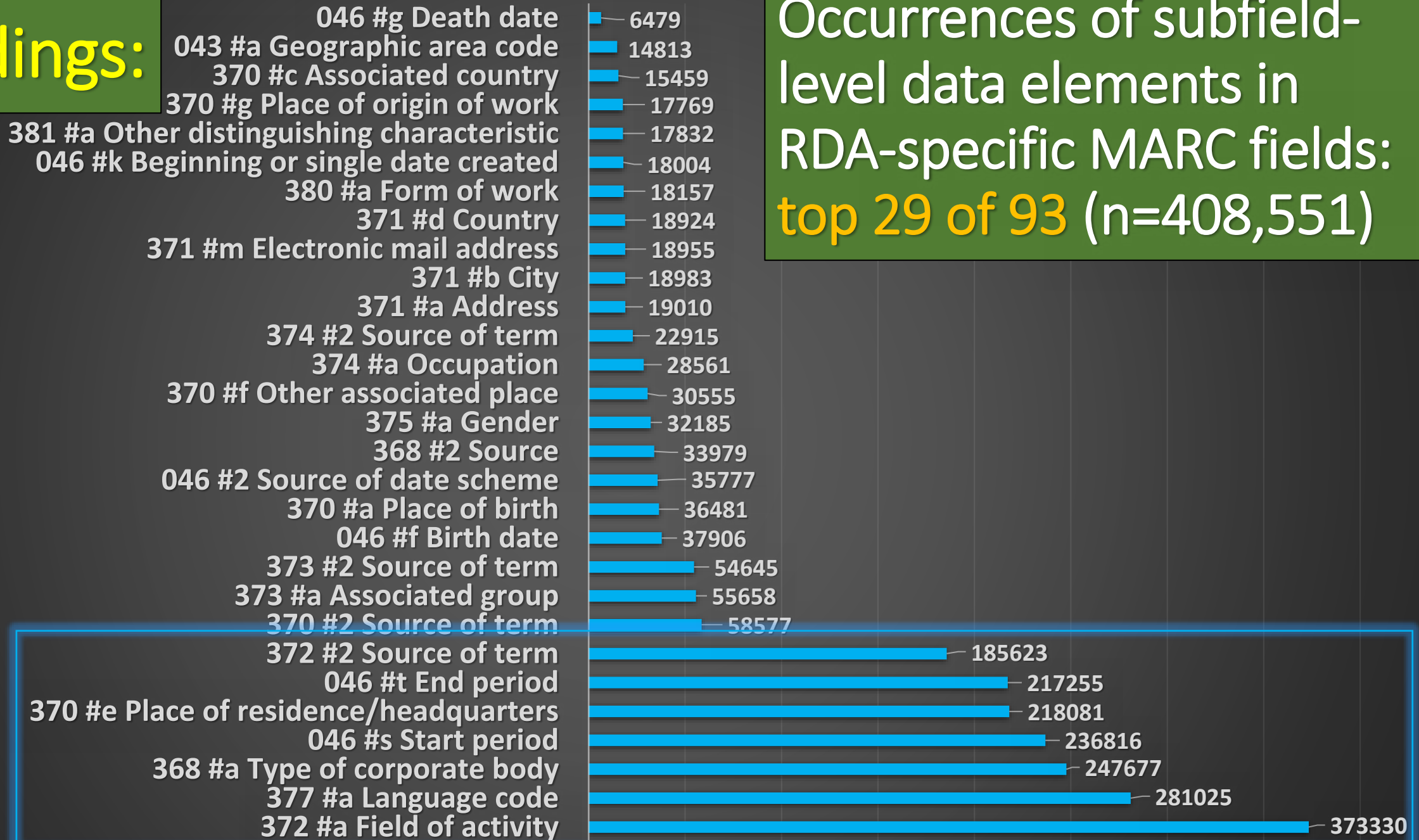
Findings:

Occurrences of RDA-specific MARC fields: observed 21 out of 35 fields



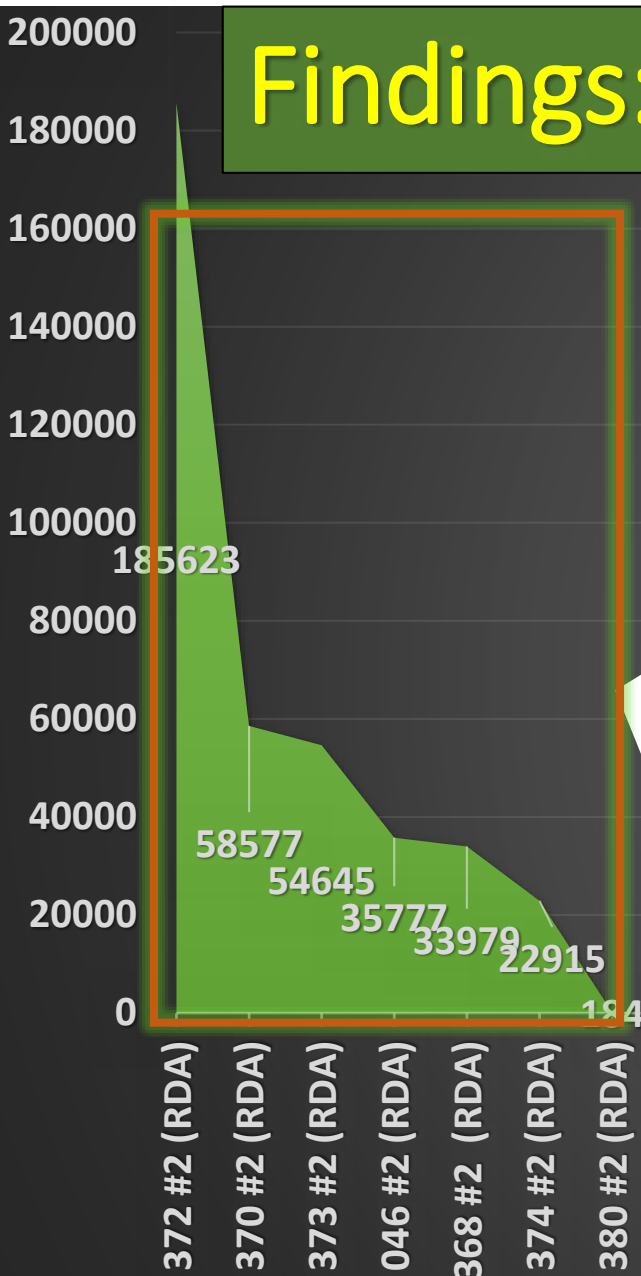
Findings:

Occurrences of subfield-level data elements in RDA-specific MARC fields: **top 29 of 93** (n=408,551)



Findings:

Occurrences of Linked Data enabling data elements (n=408,551)



#2 Source of term
 for various RDA-specific MARC fields:
 046 Special Coded Dates
 368 Other Attributes
 370 Associated place
372 Field of Activity
 373 Associated Group
 374 Occupation

#i Relationship information
 for one non-RDA-specific MARC field:
 500 Related personal name heading

Conclusions & Future Research

- Developing understanding of the impact of RDA-based record creation on the quality and functionality of name authority records.
- Supplementing quantitative analysis of a large dataset by in-depth analysis of its subsets
 - e.g., samples of records from each of the 5 types of authority records in **NAF**
- Expanding research by including subject authority records from **SAF**
- Monitoring the change in authority records over time
 - identify the trends in enhancing RDA-based authority data through increasing level of use of data elements important to facilitating information access
 - levels of application of data elements that provide Linked Data functionality (e.g., #0 Authority record control number or standard number, #2 Source of term, #4 Relationship, #i Relationship information, #u Uniform Resource Identifier, etc.).

THANK YOU!

Questions?

Comments?

Ideas?