

# Preparing for the Future: National Library of Medicine's Project to Add MeSH<sup>®</sup> RDF URIs to its Bibliographic and Authority Records

---

A presentation to CNI

December 12, 2017

Diane Boehr, Head, Cataloging & Metadata Management Section

Barbara Bushman, Assistant Head, Cataloging & Metadata Management Section

National Library of Medicine

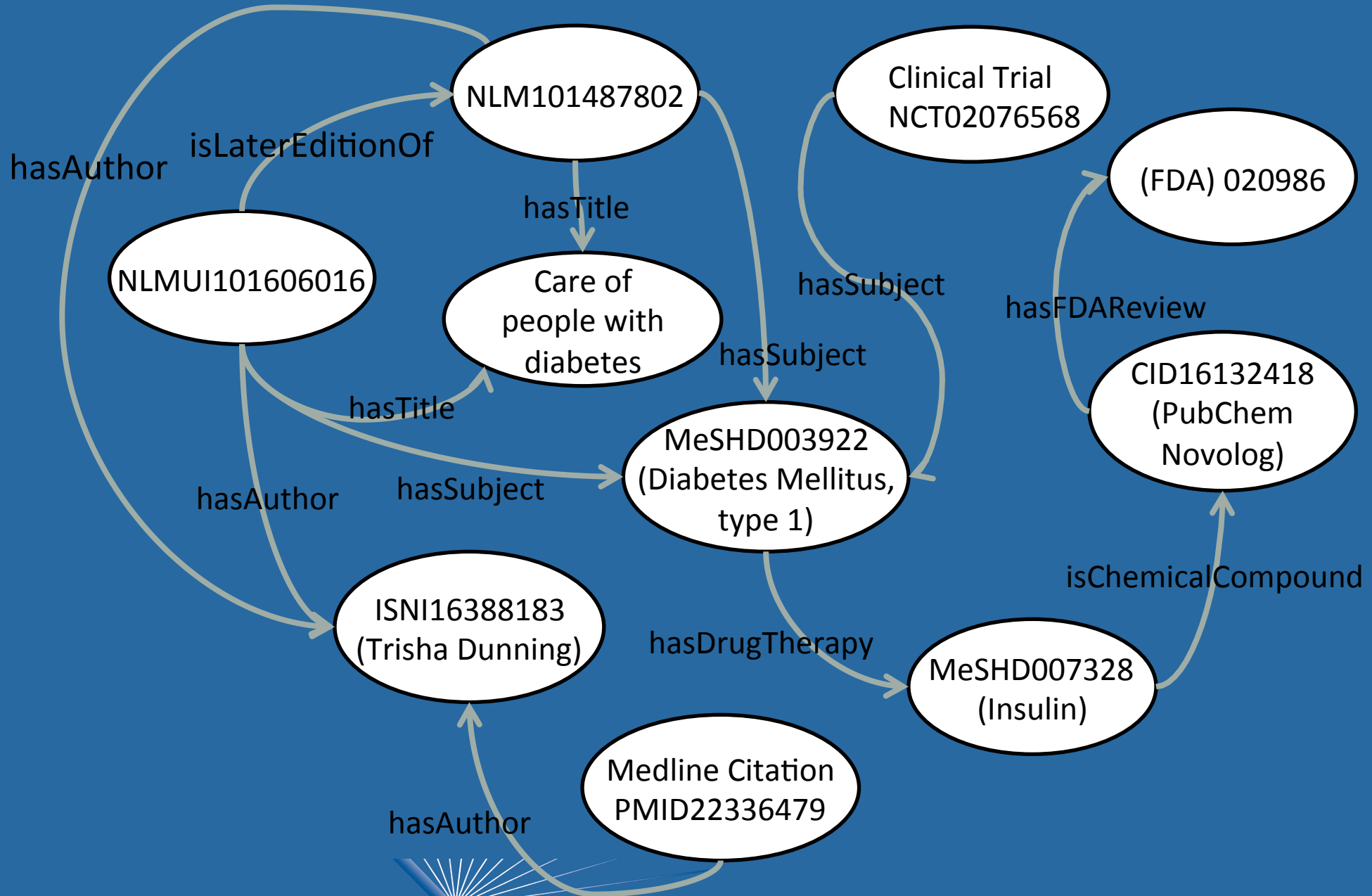
National Institutes of Health, Dept. of Health and Human Services



# Why Should We Add URIs to MARC Data?

- No current impact in MARC record
- Benefit will be in future transition of MARC to a linked data format
  - Identifiers rather than text strings
- Provide a test set of data for NLM and/or others to test conversion programs





# Augmenting Catalog Data with MeSH URIs

- Reasons for starting with MeSH:



- Break down the larger goal of preparing for linked data into manageable smaller tasks
- MeSH is NLM's own medical subject thesaurus for cataloging and indexing. We control its accuracy, stability, and currency



# Background Statistics on NLM's Catalog Data

- Contains
  - ca. 1.5 million bibliographic records
  - ca. 1.5 million authority records
    - 850,000 MeSH authority records
- On an average week NLM creates or updates ca. 350 bibliographic records
- Records are distributed to subscribers in MARC and XML formats
  - Files are available at <https://www.nlm.nih.gov/databases/download/catalog.html>



# MeSH RDF URIs

- Dereferenceable URIs were minted for each Descriptor and Qualifier by appending its unique identifier (UI) to base URI of <https://id.nlm.nih.gov/mesh>

## Example:

- **UI for Descriptor** Acquired Immunodeficiency Syndrome: D000163
- **URI for Descriptor** Acquired Immunodeficiency Syndrome : <https://id.nlm.nih.gov/mesh/D000163>



# Example MeSH Headings and URIs

MARC 21	Subject Heading	RDF URI
650	\$a Acquired Immunodeficiency Syndrome \$x ethnology	<a href="http://id.nlm.nih.gov/mesh/D000163Q000208">http://id.nlm.nih.gov/mesh/D000163Q000208</a>
650	\$a Cross Cultural Comparison	<a href="http://id.nlm.nih.gov/mesh/D003431">http://id.nlm.nih.gov/mesh/D003431</a>
650	\$a Health Policy	<a href="http://id.nlm.nih.gov/mesh/D006291">http://id.nlm.nih.gov/mesh/D006291</a>
651	\$a Africa \$x ethnology	<a href="http://id.nlm.nih.gov/mesh/D000349Q000208">http://id.nlm.nih.gov/mesh/D000349Q000208</a>
651	\$a Caribbean Region \$x ethnology	<a href="http://id.nlm.nih.gov/mesh/D017691Q000208">http://id.nlm.nih.gov/mesh/D017691Q000208</a>
655	\$a Congresses	<a href="http://id.nlm.nih.gov/mesh/D016423">http://id.nlm.nih.gov/mesh/D016423</a>

# Recording the URI in MARC Bibliographic Records

- Subfield \$0 was added to the MARC bibliographic format for subject and name fields in the mid-2000s
  - Defined as an authority control or standard number
  - 2016 defined its use for recording URIs which represent objects in RDF triple statements
- Program for Cooperative Cataloging (PCC) has not yet developed best practices for its members
- NLM granted permission to include MeSH URIs in subfield \$0 on PCC records





# Method for Updating the Bibliographic Records

- How to add the URIs
  - Manually versus programmatically
- If programmatically, what to use?
  - Existing software?
  - Custom NLM programming?
- How to generate and store URI
  - Locally store in MeSH MARC authority record
  - Programmatically add to bibliographic records



# MeSH MARC

	Tag	I1	I2	Subfield Data
→	035			‡a (DNLM)D003422 ←
	035			‡a 337422
	040			‡a DNLM ‡c DNLM
	072			‡a E2. ‡x 760. ‡x 190
	072			‡a N2. ‡x 421. ‡x 585. ‡x 190
	150			‡a Critical Care
	450			‡w nna ‡a Surgical Intensive Care
	450			‡a Intensive Care
	450			‡a Intensive Care, Surgical
	450			‡w nna ‡a Care, Critical
	450			‡w nna ‡a Care, Intensive
	450			‡w nna ‡a Care, Surgical Intensive
	550			‡a Emergencies
	680			‡i Health care provided to a critically ill patient during a medical emergency or crisis.
	688			‡a 75

	Tag	I1	I2	Subfield Data
→	035			‡a D003422Q000639 ←
	035			‡a 337433
	040			‡a DNLM ‡c DNLM
	150			‡a Critical Care ‡x trends

# Updating Legacy Data and Ongoing Processing of New Bibliographic Records



Authority Heading / Subheading	Authority Record 035\$a MeSH Unique Identifier	Authority Record 029\$0 URI	65X Field Bibliographic Record
150 __ \$a Critical Care \$x trends	035 __ \$a D003422Q000639	029 __ \$0 <u><a href="https://id.nlm.nih.gov/mesh/D003422Q000639">https://id.nlm.nih.gov/mesh/D003422Q000639</a></u>	650 12 \$a Critical Care \$x trends \$0 <u><a href="https://id.nlm.nih.gov/mesh/D003422Q000639">https://id.nlm.nih.gov/mesh/D003422Q000639</a></u>

# Resulting Authority Record

NLM decided to store the RDF URIs in the Voyager authority record in a locally defined field 029

Tag	I1	I2	Subfield Data
→ 029			⌘0 <a href="https://id.nlm.nih.gov/mesh/D003422">https://id.nlm.nih.gov/mesh/D003422</a>
035			⌘a (DNLM)D003422
040			⌘a DNLM ⌘c DNLM
072			⌘a E2. ⌘x 760. ⌘x 190
072			⌘a N2. ⌘x 421. ⌘x 585. ⌘x 190
150			⌘a Critical Care
450			⌘w nna ⌘a Surgical Intensive Care
450			⌘a Intensive Care
450			⌘a Intensive Care, Surgical
450			⌘w nna ⌘a Care, Critical
450			⌘w nna ⌘a Care, Intensive
450			⌘w nna ⌘a Care, Surgical Intensive
550			⌘a Emergencies
680			⌘i Health care provided to a critically ill patient during a medical emergency or crisis.
688			⌘a 75

Valid Values

016	National Bibliographic Agency Control Number
020	International standard book number
022	International Standard Serial Number
024	Other Standard Identifier
029	Actionable Authority Identifier
031	Musical Incipits Information
034	Coded Cartographic Mathematical Data
035	System Control Number (use 035⌘z also for NACO(www.mdd))

OK Cancel

# Resulting Bibliographic Record

060	0	0	‡a 2014 A-266
060	1	0	‡a WM 140
245	0	0	‡a Brain disorders in critical illness : ‡b mechanisms, diagnosis, and treatment / ‡c edited by Robert D. Stevens, Tarek Sharshar, E. Wesley Ely.
264		1	‡a Cambridge : ‡b Cambridge University Press, ‡c 2013.
300			‡a xv, 440 pages, 8 unnumbered pages of plates : ‡b illustrations (some color) ; ‡c 26 cm
336			‡a text ‡b txt ‡2 rdacontent
337			‡a unmediated ‡b n ‡2 rdamedia
338			‡a volume ‡b nc ‡2 rdacarrier
504			‡a Includes bibliographical references and index.
650	1	2	‡a Mental Disorders ‡0 <a href="https://id.nlm.nih.gov/mesh/D001523">https://id.nlm.nih.gov/mesh/D001523</a>
650	2	2	‡a Brain Diseases ‡0 <a href="https://id.nlm.nih.gov/mesh/D001927">https://id.nlm.nih.gov/mesh/D001927</a>
650	2	2	‡a Critical Care ‡0 <a href="https://id.nlm.nih.gov/mesh/D003422">https://id.nlm.nih.gov/mesh/D003422</a>
650	2	2	‡a Critical Illness ‡0 <a href="https://id.nlm.nih.gov/mesh/D016638">https://id.nlm.nih.gov/mesh/D016638</a>
700	1		‡a Stevens, Robert D., ‡e editor.
700	1		‡a Sharshar, Tarek, ‡e editor.
700	1		‡a Ely, E. Wesley, ‡e editor.

# Timing of the Process



Several factors weighed in determining the timing of the process:

- Other demands on our system and system development staff
- Potential impact on our users, subscribers, and other NLM products that use our MARC data
- System downtime for our Voyager ILS and coordination with other NLM projects impacting the ILS and our programming staff

NLM selected the weekend leading up to July 4 to update the legacy data



# Internal Impacts

- Stakeholders had to be notified of upcoming system downtime
- NLM Catalog DTD had to be updated
- MARC to XML transformations had to be checked and adjusted for the new DTD attribute
- Update documentation and instruct staff on how to work with records containing URIs



# Impacts on the Community

- Notifications to subscribers of NLM data
  - Formation of a new listserv
- Consultation with LC about impact on CIP (Cataloging-in-Publication)
  - Required a change to our MARCOUT program
- Consultation with OCLC





# Cleaning Up Data

- The first test run found ca. 4100 invalid MeSH in bibliographic records
  - Most of these errors were spacing problems, capitalization, punctuation, or incorrect MARC tagging
  - About 3300 of these were able to be fixed in an automated fashion using the MarcEdit program, the remaining 800 were corrected manually by CaMMS staff over a period of two days
- Only a few errors were discovered in the authority records



# Outcomes

- The July update was a success
- Weekly updates ensure our MeSH data remains clean and accurate
- In November we successfully reloaded the updated MeSH file with URIs into the authority and bib files



# Lessons Learned

Adding RDF URIs to bibliographic data is a complex task with many moving parts. It requires:

- Understanding the data and the implications of changing that data
- Planning and coordination with staff throughout the library to determine the impact on other systems and processes
- For NLM, it also involves coordination with our cataloging partners and the larger library community



# Next Steps



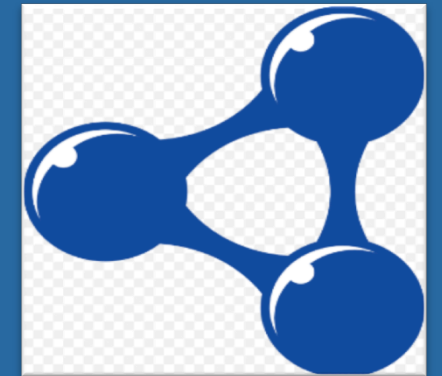
- Collect feedback from consumers of our enhanced data
- Investigate adding RDF URIs to other MARC fields
- Continue to work with developers of MarcEdit, the Authority Toolkit, and possibly other tools
- Work within NLM to ensure that MeSH RDF URIs are distributed in all outputs of MeSH data
- Work with vendors to publish some NLM bibliographic data as linked data and examine interaction with other data sources
- Look at other NLM products and services and coordinate linked data efforts to further connect more of our data together



# Conclusion



- Although not yet linked data, adding MeSH RDF URIs to NLM's bibliographic and authority records is the first step in preparing MARC data to better position ourselves for the transition to a linked data environment



# Additional Information

- For more details on:
  - MeSH RDF: <https://id.nlm.nih.gov/mesh/>
  - Downloading NLM bibliographic data: <https://www.nlm.nih.gov/databases/download/catalog.html>
  - This project see the upcoming *Cataloging and Classification Quarterly* article entitled: “*Preparing for the Future: National Library of Medicine’s<sup>®</sup> Project to Add MeSH<sup>®</sup> RDF URIs to its Bibliographic and Authority Records*”  
<https://doi.org/10.1080/01639374.2017.1382642>



# Contact Information

- Diane Boehr – [boehrd@mail.nlm.nih.gov](mailto:boehrd@mail.nlm.nih.gov)
- Barbara Bushman – [bushmanb@mail.nlm.nih.gov](mailto:bushmanb@mail.nlm.nih.gov)



# Questions?

