

# Linked Data Initiatives at NLM

---

Barbara Bushman & Nancy Fallgren

Technical Services Division

National Library of Medicine

National Institutes of Health

U.S. Department of Health and Human Services

ALCTS Metadata Interest Group

ALA Midwinter

February 1, 2015



# Agenda

- Background
- NLM Linked Data Infrastructure Working Group
- MeSH (Medical Subject Headings) RDF Pilot
- Next Steps
- Lessons Learned



# Background

## Existing NLM Linked Data Initiatives

- PubChem RDF
- BIBFRAME
- MESH RDF Prototype

## Existing 3<sup>rd</sup> party RDF versions of NLM datasets

- MeSH (6 different versions)
- LinkedCT (clinical trials data)



# NLM Linked Data Infrastructure Working Group

- Broad collaboration across NLM divisions
- Develop and build infrastructure for transforming, storing and publishing NLM linked data
- Research best practices in publishing linked data
- Recommend NLM-wide policies and guidelines for linked data publishing
- Document guidance for maintaining the established linked data infrastructure
- Recommend processes for future data linking projects
- Prioritize NLM datasets for publication as linked data



# NLM Linked Data WG Process

## Shared working environment

- SharePoint for administrative documentation
- GitHub private site for development

## Develop a common level of understanding

## Review existing linked data initiatives

- PubChem RDF
- MeSH RDF prototype



# Pilot Project: MeSH RDF

## Community impact

- Widely used in the health and medical community
- Ability to relate many disparate health and medical resources

## Community interest evidenced by

- Multiple 3<sup>rd</sup> party versions published
- Requests stemming from BIBFRAME experimentation

## Existing MeSH RDF prototype



# Decisions

- URI ([id.nlm.nih.gov](http://id.nlm.nih.gov))
- RDF vocabulary/Predicates  
(create our own vs. use existing)
- License
- Consultants



# How to Provide the Linked Data

## FTP

- XML, XSLT, RDF

## SPARQL endpoint

- MeSH RDF files loaded into a graph
- Stored in Virtuoso triple store
- Accessible via Lodestar interface





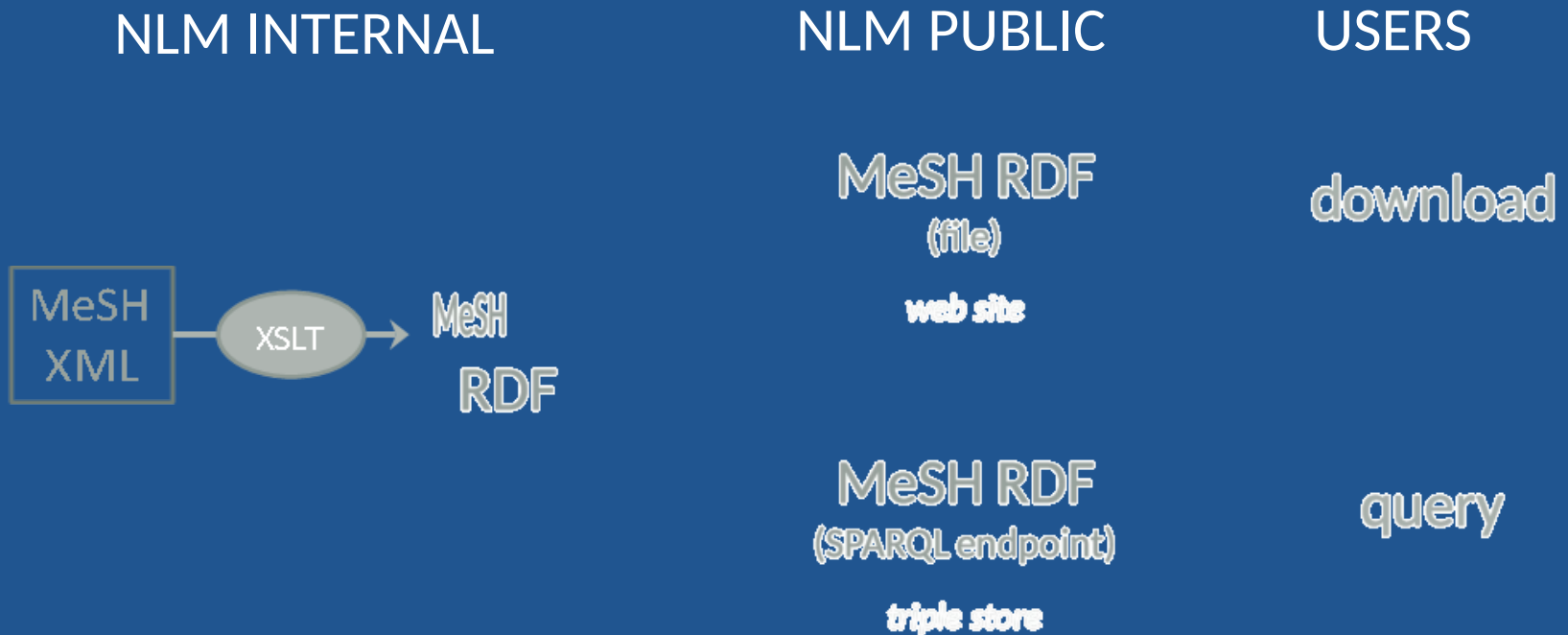
# Creating MeSH RDF

---



# Creating MeSH RDF

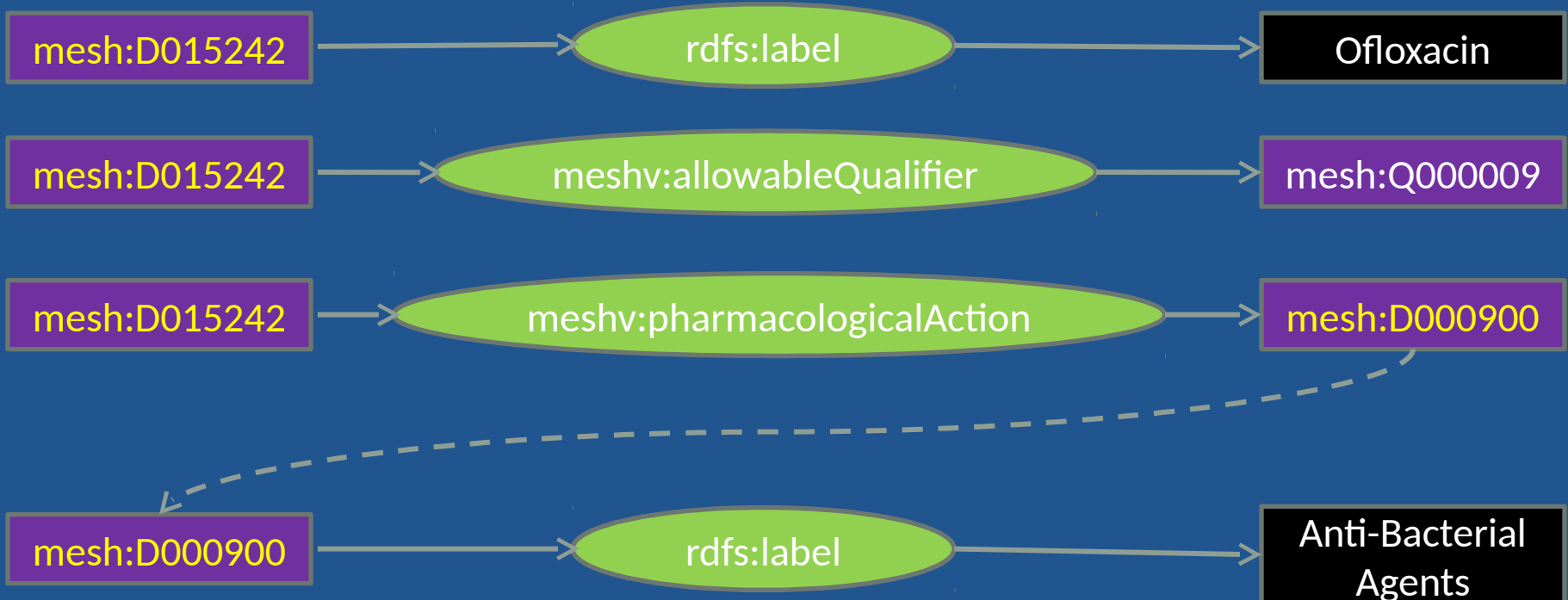
## Transformation of MeSH XML to MeSH RDF



MeSH Heading	Ofloxacin
Tree Number	<a href="#">D03.438.810.835.322.500</a>
Scope Note	A synthetic fluoroquinolone antibacterial agent that inhibits the supercoiling activity of bacterial <a href="#">DNA GYRASE</a> , halting <a href="#">DNA REPLICATION</a> .
Entry Term	DL-8280
Entry Term	DR-3355
Entry Term	Hoe-280
Entry Term	Ofloxacin Hydrochloride
Entry Term	Ofloxacin
Entry Term	ORF-28489
Entry Term	Ru-43280
Entry Term	Tarivid
Allowable Qualifiers	<a href="#">AA AD AE AG AI AN BL CF CH CL CS CT DU EC HI IM IP ME PD PK PO RE SD ST TO TU UR</a>
Pharm. Action	<a href="#">Anti-Bacterial Agents</a>
Pharm. Action	<a href="#">Anti-Infective Agents, Urinary</a>
Pharm. Action	<a href="#">Topoisomerase II Inhibitors</a>
CAS Type 1 Name	(+/-)-9-Fluoro-2,3-dihydro-3-methyl-10-(4-methyl-1-piperazinyl)-7-oxo-7H-pyrido(1,2,3-de)-1,4-benzoxazine-6-carboxylic acid
Registry Number	A4P49JAZ9H
Related Number	100986-86-5 ((R)-isomer)
Related Number	118120-51-7 (HCl)
Related Number	82419-36-1 (Ofloxacin)
Related Number	86784-41-0 (cpd w/o isomeric designation)
Related Number	I2UWV315WA
Previous Indexing	<a href="#">Anti-Infective Agents</a> (1981-1988)
Previous Indexing	<a href="#">Anti-Infective Agents, Urinary</a> (1981-1988)
Previous Indexing	<a href="#">Oxazines</a> (1981-1988)
History Note	89
Date of Entry	19880516
Unique ID	D015242

# MeSH in RDF

Subject	Predicate	Object
D015242	MeSH Heading	Ofloxacin
D015242	Allowable Qualifiers	<a href="#">AA</a> <a href="#">AD</a> <b><a href="#">AE</a></b> <a href="#">AG</a> <a href="#">AI</a> <a href="#">AN</a> <a href="#">BL</a> <a href="#">CF</a> <a href="#">CH</a> <a href="#">CL</a> <a href="#">CS</a> <a href="#">CT</a> <a href="#">DU</a> <a href="#">EC</a> <a href="#">HI</a> <a href="#">IM</a> <a href="#">IP</a> <a href="#">ME</a> <a href="#">PD</a> <a href="#">PK</a> <a href="#">PO</a> <a href="#">RE</a> <a href="#">SD</a> <a href="#">ST</a> <a href="#">TO</a> <a href="#">TU</a> <a href="#">UR</a>
D015242	Pharm. Action	<a href="#">Anti-Bacterial Agents</a>



# XML2RDF Modeling Issues

- Descriptor/Qualifier pairs
  - Not in MeSH XML
  - ‘Illegal’ descriptor/qualifier combinations
- Hierarchical relationships are not identified in MeSH XML
- Transitive relationships are not always true between descriptors in multiple tree nodes



# MeSH Trees for Eye

## Face [A01.456.505]

Cheek [A01.456.505.173]

Chin [A01.456.505.259]

▶ Eye [A01.456.505.420]

→ Eyebrows [A01.456.505.420.338]

Eyelids [A01.456.505.420.504] +

Forehead [A01.456.505.580]

Mouth [A01.456.505.631] +

Nasolabial Fold [A01.456.505.682]

Nose [A01.456.505.733]

Parotid Region [A01.456.505.750]

## Sense Organs [A09]

Ear [A09.246] +

▶ Eye [A09.371]

Anterior Eye Segment [A09.371.060] +

Anterior Capsule of the Lens [A09.371.061]

Axial Length, Eye [A09.371.199]

Eyelids [A09.371.337] +

Lacrimal Apparatus [A09.371.463] +

Oculomotor Muscles [A09.371.613]

Pigment Epithelium of Eye [A09.371.670] +

Posterior Eye Segment [A09.371.714] +

Retina [A09.371.729] +

Sclera [A09.371.784]

Tenon Capsule [A09.371.839]

Uvea [A09.371.894] +

Nose [A09.531] +

Taste Buds [A09.846]



# RDF Statements Must Always Be True

<Face> <has narrower term> <Eye>

<A01.456.505> <has narrower term> <A01.456.505.420>


<Eye> <has narrower term> <Eyebrows>

<A01.456.505.420> <has narrower term> <A01.456.505.420.338>

<Sense Organs> <has narrower term> <Eye>

<A09> <has narrower term> <A09.371>

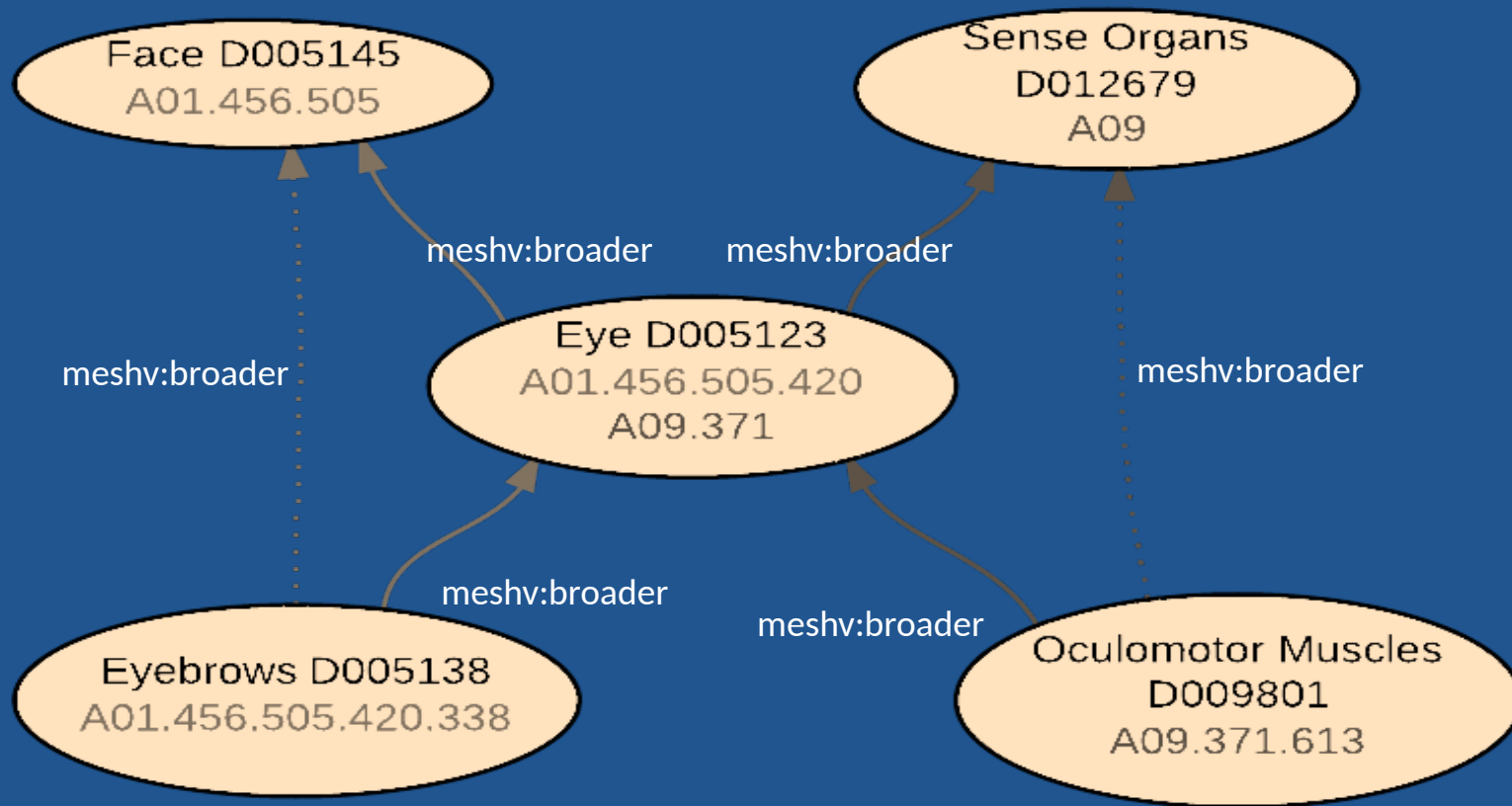
<Eye> <has narrower term> <Eyebrows>

 <A09.371> <has narrower term> <A01.456.505.420.338>



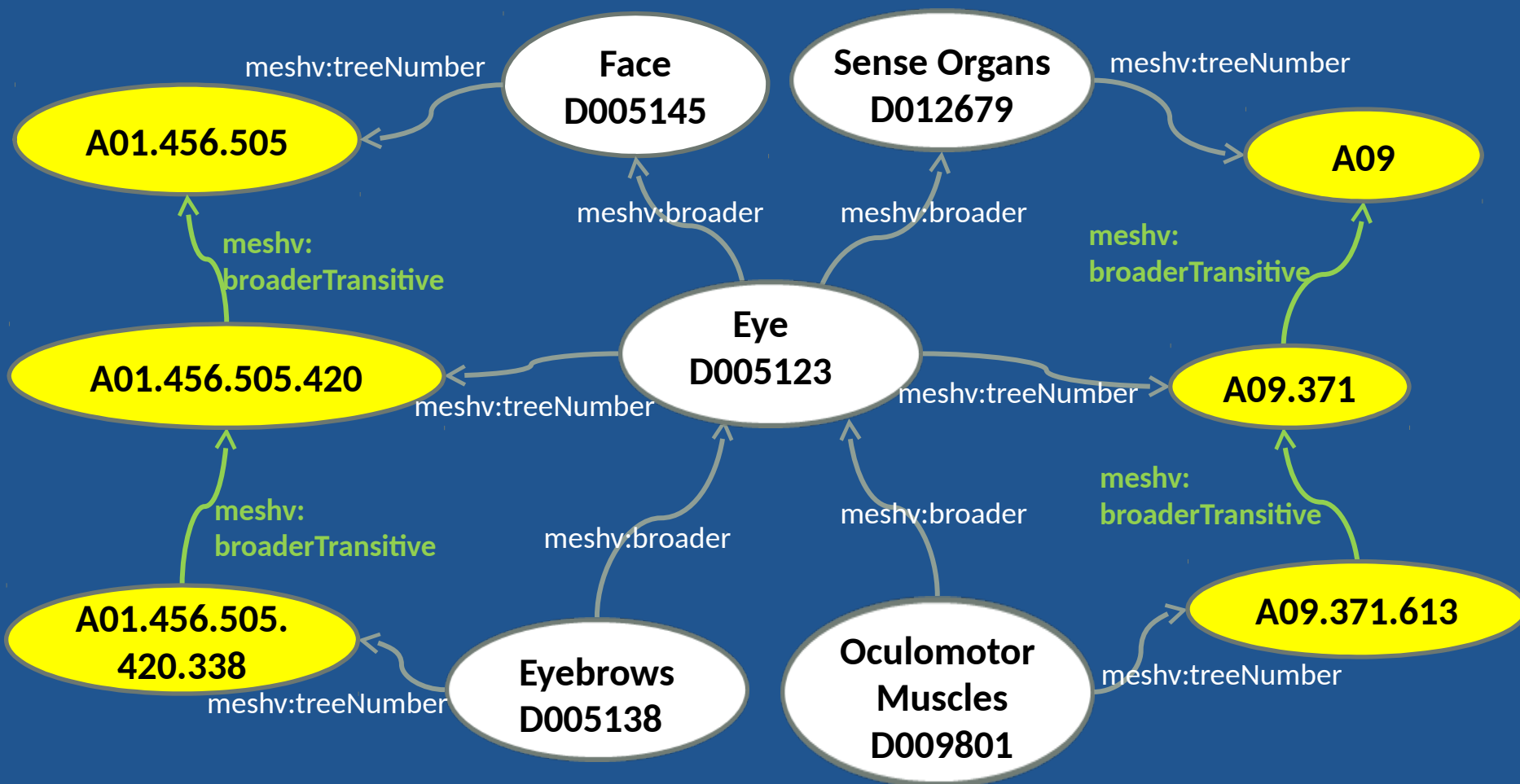
# Using Only Transitive Relationships

Are Sense Organs really a broader term for Eyebrows?





# Using Transitive and Non-Transitive Relationships



# (Soft) Beta Launch

<http://id.nlm.nih.gov>

- Launched Nov. 17, 2014 at the American Medical Informatics Association conference

Work in progress

- Still tweaking model and documentation

No public news announcements

No press release

No direct link on NLM home page



# Beta Evaluation

## Feedback from partners and others

- Public GitHub site (<https://github.com/HHS/meshrdf> )
- Customer service
- Social media

## Analytics

- Log files
- WebTrends



# MeSH RDF Next Steps

Next release of MeSH RDF ca. May 2015

- Update to 2015 MeSH
- Resolve outstanding issues raised during beta

Updating/versioning

Review MeSH RDF elements

Contribute to revising MeSH XML



# Lessons Learned

- Have a flexible timeframe
- Collaborate broadly
- Document everything
- Ask for help
- Understand expectations and anticipated outcomes
- Create an evaluation plan
- Value community collaboration



# MeSH RDF Beta

## Demo

- Landing page
- Technical documentation
- GitHub
- Sample SPARQL query





# Medical Subject Headings (MeSH) RDF Linked Data (beta)

The National Library of Medicine (NLM) is now offering a beta version of the Medical Subject Headings (MeSH®) data in RDF (Resource Description Framework). RDF is a well-known standard for representing structured data on the Web. Systems that use RDF are often called Linked Data because of RDF emphasis on well-described links between resources.

During this beta release, NLM is seeking stakeholder input and feedback as part of a broader effort to evaluate the creation of an NLM Linked Data Service. NLM hopes that users will help us refine MeSH RDF.

Once beta testing is finished, NLM will release the authoritative, consistent, and permanent MeSH RDF data, which can be incorporated into systems, products, and the broader Web of Linked Data. NLM will continue to develop tools and services that provide MeSH data based on feedback from the beta period.

## Why MeSH as Linked Data?

The MeSH thesaurus is a controlled vocabulary produced by NLM since 1960. NLM uses MeSH in our products and systems for indexing, cataloging, and

# MeSH Linked Data (beta) Technical Documentation

## Descriptors

A Descriptor is a class in MeSH RDF with the name [meshv:Descriptor](#). Also known as Main Headings or MeSH Headings, Descriptors are used to index citations in the NLM MEDLINE database and to describe the subjects for NLM Catalog records. Descriptors are searchable in PubMed and NLM Catalog with the search tag [MH]. Most Descriptors indicate the subject of a resource (including geographic terms). Some indicate publication types (what a resource is rather than what it is about; for example: Randomized Controlled Trial or Letter). For more information about Descriptors, visit the NLM [MeSH Record Types page](#).

[\[ jump to descriptor properties or descriptor relations \]](#)

## Class Information

In MeSH RDF, the subclasses of [meshv:Descriptor](#) are:

- [meshv:TopicalDescriptor](#)
- [meshv:PublicationType](#)
- [meshv:CheckTag](#)
- [meshv:GeographicalDescriptor](#)

The chart below displays the properties of the [meshv:TopicalDescriptor D015242](#), 'Ofloxacin'.



GitHub, Inc. (US)

https://github.com/HHS/meshrdf/issues/

Google

☆

↓

🏠

This repository

Search

ExploreGistBlogHelp

fallgrennj

+

⚙️

HHS / meshrdf

👁️ Unwatch ▾33

★ Star2

🍴 Fork1

IssuesPull requestsLabelsMilestones

Filters ▾🔍 is:issue is:open

New issue

📄

🚨 3 Open✔️ 0 Closed

Author ▾Labels ▾Milestones ▾Assignee ▾Sort ▾

📄

🚨 URIs for Tree numbers in SPARQL resolve to JSON 

Bug

#9 opened 5 days ago by bushmanb

0

📄

🚨 SPARQL endpoint and browser compatibility

#8 opened 5 days ago by bushmanb

0

📄

🚨 Schema incomplete 

Data Model Suggestion

#7 opened 5 days ago by bushmanb

0

🔗

📊

🔧

NIH

U.S. National Library of Medicine

25

# Medical Subject Headings (MeSH) RDF Linked Data (beta)

Enter SPARQL Query

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX meshv: <http://id.nlm.nih.gov/mesh/vocab#>
PREFIX mesh: <http://id.nlm.nih.gov/mesh/>
```

```
SELECT *
FROM <http://id.nlm.nih.gov/mesh2014>
WHERE {
  mesh:D015242 meshv:pharmacologicalAction ?pa .
  ?pa rdfs:label ?paLabel .
}
```

## Example Queries

- [MeSH Linked Data Predicates](#)

Retrieve the list of distinct predicates in MeSH RDF.

- [Ofloxacin Pharmacological Actions](#)

The Pharmacological Actions of Ofloxacin and their labels.

- [Allowable Qualifiers](#)

Any MeSH descriptor that has an allowable qualifier of 'adverse effects'.

- [String search on 'infection'](#)

Any MeSH term ('D' or 'M') that has 'infection' as part of its name. (inference required)

RDFS inference? ☐

Output: HTML

Results per page: 50

Submit Query

Reset

Previous 50 results per page (offset 0)

Next

pa	paLabel
<a href="#">mesh:D000892</a>	Anti-Infective Agents, Urinary
<a href="#">mesh:D000900</a>	Anti-Bacterial Agents
<a href="#">mesh:D059005</a>	Topoisomerase II Inhibitors



# Medical Subject Headings (MeSH) RDF Linked Data (beta)

[SPARQL endpoint](#)[Technical Docs](#)[MeSH RDF Home](#)[MeSH Home](#)

**About:** [Anti-Infective Agents, Urinary](#) 

<http://id.nlm.nih.gov/mesh/D000892>



**Type:** [MeSH TopicalDescriptor](#)

Topical Descriptors indicate the subject of an indexed item such as a journal article. See D063926 (Drug Hypersensitivity Syndrome) for an example.  
[more types...](#)

## Related to

**concept** (MeSH Concept)

- [Antiseptics, Urinary](#)

**preferredConcept** (MeSH Concept)

- [Anti-Infective Agents, Urinary](#)

**broader** (MeSH TopicalDescriptor)

- [Anti-Infective Agents](#)

**treeNumber** (MeSH TreeNumber)

- [D27.505.954.613.056](#)
- [D27.505.954.122.237](#)

**allowableQualifier** (MeSH Qualifier)

- [contraindications](#)
- [metabolism](#)
- [antagonists & inhibitors](#)
- [history](#)
- [toxicity](#)
- [chemical synthesis](#)
- [pharmacokinetics](#)
- [adverse effects](#)
- [blood](#)
- [economics](#)
- [more...](#)



# Questions/Comments

Barbara Bushman

bushmanb@mail.nlm.nih.gov

Nancy Fallgren

fallgrennj@mail.nlm.nih.gov

Beta MeSH RDF

<http://id.nlm.nih.gov/mesh/>

