



SciBite
an ELSEVIER company

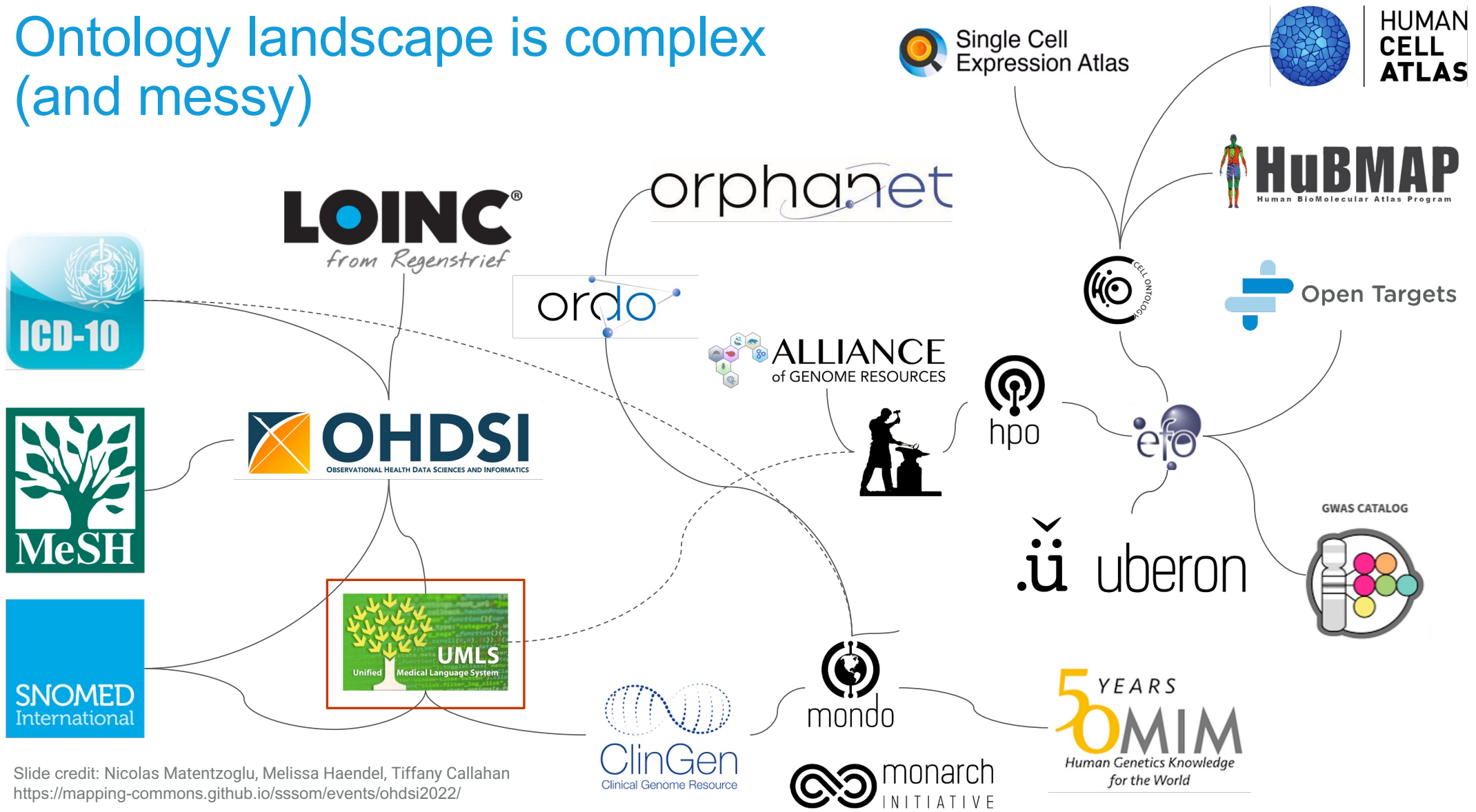
Managing mappings at scale

Simon Jupp

Head of Semantic Technology



Ontology landscape is complex (and messy)









Slide credit: Nicolas Matentzoglou, Melissa Haendel, Tiffany Callahan
<https://mapping-commons.github.io/sssom/events/ohdsi2022/>




Mappings from public sources


- More than 10 million extracted
- Often have weak semantics – many use *has_dbxref*
- Prefixes of CURIEs ambiguous – need normalization (HT Bioregistry and <https://identifiers.org/>)
- Hard to judge quality



A1BG-AS1 **HGNC:37133**   


Primary ID: https://www.genenames.org/data/gene-symbol-report/#!/hgnc_id/HGNC:37133   

Annotations Relationships Graph History



Labels 

Label 



- A1BG-AS1  



Synonyms 

IAO_0000115 








- A1BG antisense RNA 1  

Alias_symbol 

- FLJ23569  

Mappings  

HasDbXref 

- ENA:BC040926  
- ENSEMBL:ENSG00000268895  
Source HGNC:37133
- RefSeq:NR_015380  
- UCSC:uc002qse.3  



Available ontology mappings

- Many mappings available in the public domain provided by terminologies e.g.
 - MedDRA->SNOMED CT (<https://www.meddra.org/software-packages>)
 - MedDRA->ICD10 (<https://www.meddra.org/software-packages>)
 - SNOMED CT->ICD10 (<https://www.snomed.org/maps>)
 - SNOMED CT to ICD-O (<https://www.snomed.org/maps>)

The screenshot shows the NCI Term Browser interface. At the top, it displays the NIH National Cancer Institute logo and the URL www.cancer.gov. Below the header, there are navigation tabs for Terminologies, Value Sets, and Mappings. A search bar is present with a search button and a help icon. Below the search bar, there are radio buttons for search criteria: Contains, Exact Match, Begins With, Name, Code, Property, and Relationship. A 'Sources' dropdown menu and a 'Quick Links' dropdown menu are also visible. The main content area is titled 'Select a mapping data set to search here, or click on a mapping data set name to go to its home page and browse or search from there. It is not yet possible to search more than one mapping data set at a time.' Under the 'Mappings:' section, there is a list of mappings with radio buttons next to each item:

- [GO to NCI Mapping; GO to NCI Mapping \(February 2020\)](#)
- [MA to NCI Mapping; MA to NCI Mapping \(November 2011\)](#)
- [NCI to ChEBI Mapping; NCI to ChEBI Mapping \(August 2023\)](#)
- [NCI to HGNC Mapping; NCI to HGNC Mapping \(September 2023\)](#)
- [PDQ to NCI Mapping; PDQ to NCI Mapping \(201607\)](#)
- [SNOMEDCT_US to ICD-10-CM Mapping; SNOMED CT US Edition to ICD-10-CM Mapping \(2022_09_01\)](#)
- [SNOMEDCT_US to ICD10 Mapping; SNOMED CT US Edition to ICD-10 Mapping \(2022_09_01\)](#)

Below this list is a 'Search' button. Under the 'Other mappings available for download:' section, there is a list of download links:

- [NCI to SwissProt Mapping; NCI to SwissProt Mapping \(2023-09-29\)](#)
- [ICD-O-3.1 to NCI Topography Mapping; ICD-O-3.1 to NCI Topography Mapping \(2017-12-21\)](#)
- [ICD-O-3.1 to NCI Morphology Mapping; ICD-O-3.1 to NCI Morphology Mapping \(2017-12-21\)](#)
- [ICD-O-3.1 to NCI Axis Mappings; ICD-O-3.1 to NCI Axis Mappings \(2017-12-21\)](#)
- [NCI Maps To GDC \(23.09d\)](#)
- [NCI Maps To ICD10_2016 \(23.09d\)](#)
- [NCI Maps To ICD10CM_2017 \(23.09d\)](#)
- [NCI Maps To ICD9CM_2014 \(23.09d\)](#)
- [NCI Maps To ICDQ3 \(23.09d\)](#)
- [NCI Maps To MedDRA \(23.09d\)](#)

At the bottom of the page, there is a footer with links for NCI Home, Contact Us, Policies, Accessibility, FOIA, HHS Vulnerability Disclosure, U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute, and USA.gov.

Mapping challenges



Locating mappings



Normalizing mappings



Provenance/quality



Custom mappings

Governance
Versioning
Updates



Overview of Semi-Automated Ontology matching approaches

Rule-based

- Lexical mapping e.g., Agreement Maker Light (AML)
 - Looking at label and synonym string similarity
- Structural/logic based e.g., LogMap
 - Exploit the ontology structure and relationships to infer mappings

Machine learning

- Word embeddings e.g., word2vec
- Language models e.g., BERTMap, MapperGPT



Ontology matching for life sciences

- Generally combining approaches work best
- Existing tools can generate large amounts of “likely” mappings

Challenges

- Difficult setup and installation of tools
- Poor performance over large terminologies
- Lack of emphasis on curation, versioning and maintenance
 - Lack a good user interface
- Subtly and complexity of life-science vocabulary e.g. indication vs phenotype vs adverse event
 - Semantics of what constitutes a match is often context dependent



SciBite mapping service

- Normalize mappings derived from ontologies
 - Add reciprocal mappings
 - Provide SSSOM-compliant metadata so they can be filtered
- CURIE-index based on Bioregistry
- Supplemented with value public sets
- SciBite-curated sets to fill gaps
- Access via SciBite tools – CENtree, SciBiteSearch
- Supplement with custom mappings from WorkBench

SciBite Workbench



- Platform for generating mappings at scale
- Utilizing SciBite's NER technology, public sources and SciBite curated vocabularies to predict and score mappings
- Allows curators to review and validate mappings prior to publishing
- Easy to use and manage versions and provenance of mappings

Workbench Curate and annotate

Mouse (MA) to Human (NCIt) Anatomy

Created: Oct 16, 2023, Modified: Oct 16, 2023

MAPPINGS CONFIGURATION

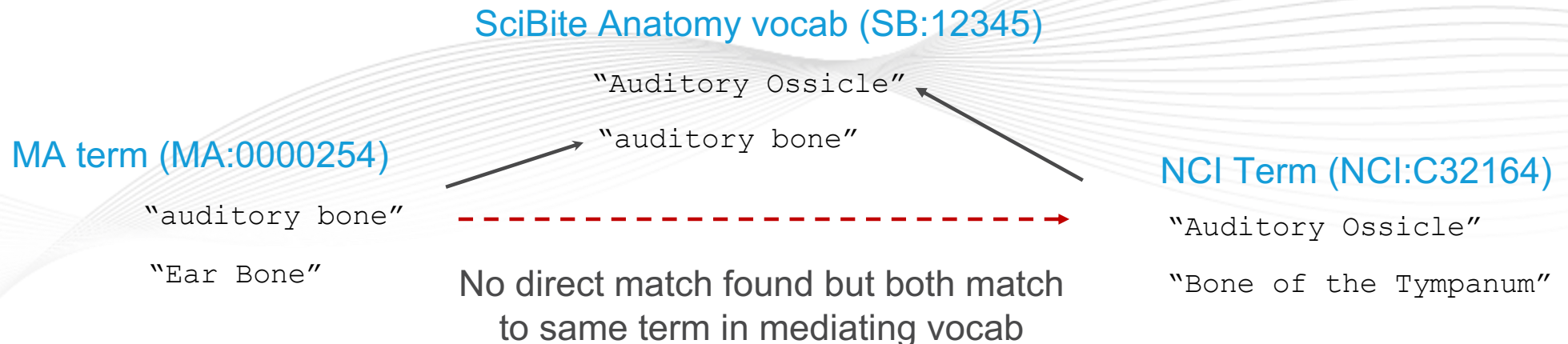
Search Mapping Relation Type

Source Term	Target Term	Mapping Relation	Type	Score	Matched
<input checked="" type="checkbox"/> abdomen MA_0000029	Abdominal_Cavity NCL_C12664	EXACT	MANUAL_VERIFIED	0.98	abdomen
<input checked="" type="checkbox"/> abdomen skin MA_0000523	Abdominal_Skin NCL_C52758	EXACT	MANUAL_VERIFIED	0.98	abdomen skin
<input checked="" type="checkbox"/> abdominal aorta MA_0000474	Abdominal_Aorta NCL_C32038	EXACT	MANUAL_VERIFIED	0.98	abdominal aorta
<input checked="" type="checkbox"/> abductor pollicis longus MA_0002266	Abductor_Pollicis_Longus NCL_C52888	EXACT	MANUAL_VERIFIED	0.98	abductor pollicis longus
<input checked="" type="checkbox"/> accumbens nucleus MA_0000892	Accumbens_Nucleus NCL_C52733	EXACT	MANUAL_VERIFIED	0.98	accumbens nucleus
<input checked="" type="checkbox"/> acetabulum MA_0001335	Acetabulum NCL_C32042	EXACT	MANUAL_VERIFIED	0.98	acetabulum
<input checked="" type="checkbox"/> adenohypophysis MA_0000177	Anterior_Lobe_of_the_Pituitary_Gland NCL_C12772	EXACT	MANUAL_VERIFIED	0.98	adenohypophysis
<input type="checkbox"/> adipose tissue MA_0000009	Adipose_Tissue NCL_C12472	EXACT	AUTOMATED	0.98	adipose tissue
<input type="checkbox"/> adrenal gland MA_0000116	Adrenal_Gland NCL_C12666	EXACT	AUTOMATED	0.98	adrenal gland
<input type="checkbox"/> adrenal gland capsule MA_0000117	Adrenal_Gland_Capsule NCL_C32050	EXACT	AUTOMATED	0.98	adrenal gland capsule
<input type="checkbox"/> alveolar artery MA_0001913	Alveolar_Artery NCL_C52848	EXACT	AUTOMATED	0.98	alveolar artery
<input type="checkbox"/> alveolar duct MA_0000417	Alveolar_Duct NCL_C32056	EXACT	AUTOMATED	0.98	alveolar duct
<input type="checkbox"/> alveolar sac MA_0000418	Alveolar_Sac NCL_C32057	EXACT	AUTOMATED	0.98	alveolar sac



Utilising SciBite's Mediating Vocabularies

- Utilising SciBite extensive curated vocabularies and public sources to act as mediators when generating mappings between source and target ontologies that have very little metadata (synonyms on terms)
- You can supply a mediating vocab to help find more mappings
- E.g., When mapping two anatomy ontologies e.g., MA and NCI anatomy, we can use SciBite Anatomy vocab to help discover potential matches



Creating an ontology alignment in SciBite Workbench

Workbench identifies potential masking terms

Workbench will identify SciBite vocabularies that will act as mediators

Add New Mapping

Mapping Name
Mouse (MA) to Human (NCIt) Anatomy

Source Ontology
MOUSE_MAPPING 34 / 100

Target Ontology
HUMAN_MAPPING

Mask Terms

vein (180) × artery (166) × muscle (149) × nerve (107) × nucleus (106) ×

intestine (36) × spinal (36) × ganglion (36) × hand (36) × lower (37) ×

foot (37) × renal (41) × cortex (41) × vertebra (44) × left (44) × skin (45) ×

dorsal (46) × respiratory (47) × digit (48) × connective (48) × right (49) ×

smooth (51) × layer (52) × duct (54) × blood (54) × vessel (59) ×

tissue (66) × epithelium (85) × system (87) × bone (94) × gland (97) ×

SUGGEST MASK TERMS

Best Hits only

Exact Match only

Cutoff score
0.7

Mediating Ontology
Anatomy

SUGGEST MEDIATING ONTOLOGY

CANCEL GENERATE MAPPING



Mouse (MA) to Human (NCIt) Anatomy

- Created: Oct 16, 2023,
- Modified: Oct 16, 2023,

MAPPINGS CONFIGURATION



Search

<input type="checkbox"/>	Source Term	Target Term	Mapping Relation	Type	Score	Matched
<input type="checkbox"/>	middle phalanx of foot MA_0001386	Phalanx_of_the_Foot NCI_C25774	BROADER	AUTOMATED	0.88	phalanx of foot
<input type="checkbox"/>	coat hair bulb MA_0000161	Hair_Bulb NCI_C32706	BROADER	AUTOMATED	0.83	hair bulb
<input type="checkbox"/>	cranial/facial muscle MA_0000					
<input type="checkbox"/>	efferent duct epithelium MA_0000					
<input type="checkbox"/>	extrinsic tongue muscle MA_0000					
<input type="checkbox"/>	eye anterior chamber MA_0000					

Can quickly filter on Exact, Broader or No match

Mouse (MA) to Human (NCIt) Anatomy

- Created: Oct 16, 2023, 13:50
- Modified: Oct 16, 2023, 13:50

MAPPINGS CONFIGURATION



Search

<input type="checkbox"/>	Source Term	Target Term	Mapping Relation	Type	Score	Matched
<input type="checkbox"/>	abdomen blood vessel MA_0000518		NONE	AUTOMATED	0	
<input type="checkbox"/>	abdomen connective tissue MA_0000519		NONE	AUTOMATED	0	
<input type="checkbox"/>	abdomen muscle MA_0000520		NONE	AUTOMATED	0	
<input type="checkbox"/>	abdomen nerve MA_0000521		NONE	AUTOMATED	0	
<input type="checkbox"/>	abdomen organ MA_0000522		NONE	AUTOMATED	0	
<input type="checkbox"/>	abdomen/pelvis/perineum MA_0000021		NONE	AUTOMATED	0	
<input type="checkbox"/>	abdomen/pelvis/perineum blood vessel MA_0000524		NONE	AUTOMATED	0	
<input type="checkbox"/>	abdomen/pelvis/perineum bone MA_0000525		NONE	AUTOMATED	0	

Explore mappings of different quality

Publish your mappings to users



1050 root nodes 29 nodes fetched 1 selected Expand all

Please note that only the first 20 root classes are shown; keep scrolling to load more.

- skin
 - trunk skin
 - abdomen/pelvis/perineum skin
 - abdomen skin
- abdomen
- abdomen/pelvis/perineum
- abdominal aorta
- accessory olfactory bulb
- accessory olfactory bulb glomerular layer

abdomen skin MA:0000523

Primary ID: http://mouse.owl#MA_0000523

Annotations Relationships Graph History

Labels

Label

- abdomen skin

Mappings

HasDbXref

- human_mapping Abdominal_Skin http://human.owl#NCI_C52758
 - Confidence 0.98
 - Source SciBite Workbench

Publish your mappings to our Ontology Management System (CENTree) with provenance