

# THE ROLE OF METADATA IN THE SECOND MACHINE AGE

DC-2016 / KØBENHAVN / 13 OCTOBER

Bradley P. Allen

Chief Architect, Elsevier

@bradleypallen

# DUBLIN CORE IN ITS THIRD DECADE



A basic description mechanism for digital information that:

- can be used in all domains
- can be used for any type of resource
- is simple, yet powerful
- can be extended and can work with specific solutions.

Dekkers, M. (2009). History, objectives and approaches of the Dublin Core Metadata Initiative. Retrieved September 29, 2016, from [http://dublincore.org/resources/training/frd\\_20091217/Tutorial\\_FRD\\_dekkers-1.pdf](http://dublincore.org/resources/training/frd_20091217/Tutorial_FRD_dekkers-1.pdf)

## DCMI IN ITS THIRD DECADE



Since its early days in the mid-1990s, DCMI's founding principle has been the discovery and management of resources through metadata across the boundaries of information silos on the Web and within intranets.

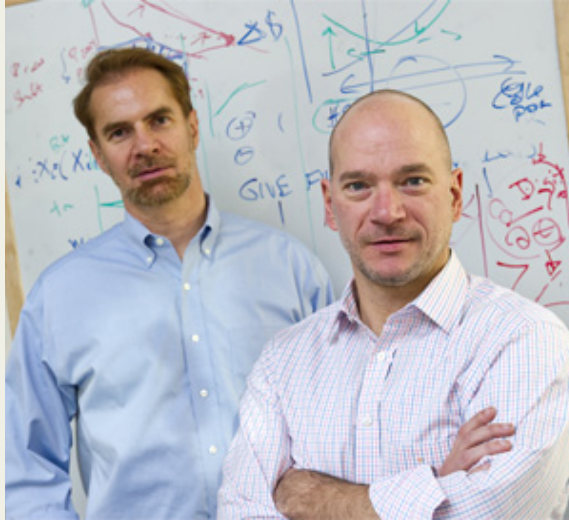
Mission and Principles. (n.d.). Retrieved September 29, 2016, from <http://dublincore.org/about-us/>

# THE IMPACT OF DC AND DCMI

- Arguably the most impactful of many efforts to bring to bear the library science approach to information organization on the Web at large
- Simple, easy to understand, easy to apply, well documented
- While the user experience of discovery has come to be dominated by search engines such as Google, metadata standards are pervasive in the infrastructure of content curation and management, and underpins search infrastructure
- All of this has been a huge contribution to the Web
- It established the approach to bringing the Semantic Web and its associated vocabularies and ontologies online as the Linked Open Data Cloud



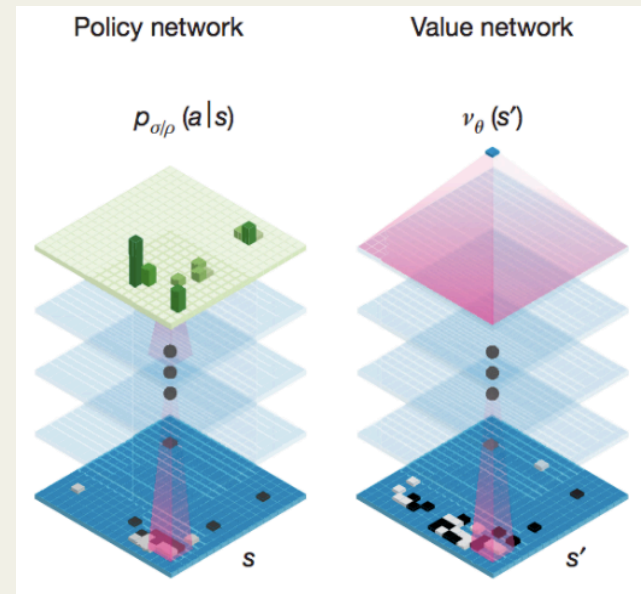
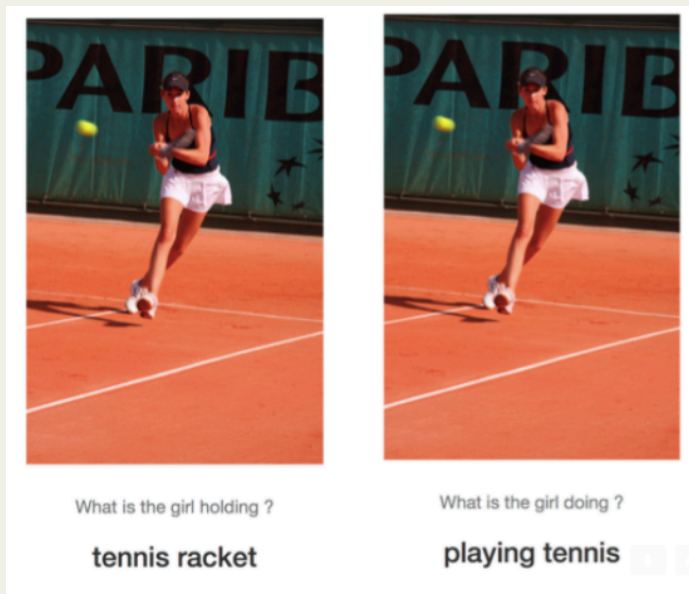
# THE NEXT 20 YEARS: THE SECOND MACHINE AGE



The second machine age will be characterized by countless instances of machine intelligence and billions of interconnected brains working together to better understand and improve our world.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. New York: W.W. Norton & Company.

# MACHINE INTELLIGENCE IS ON THE RISE








# WHAT CAN MACHINE INTELLIGENCE DO TODAY?



If there's a task that a normal person can do with less than one second of thinking, there's a very good chance we can automate it with deep learning.

Andrew Ng, Chief Scientist, Baidu (lecture at Bay Area Deep Learning School, Stanford, CA, September 24, 2016)

# MACHINE INTELLIGENCE IN USE

	Application areas	Specific examples	Recent M&A and hiring
	<ul style="list-style-type: none"> <li>• Speech understanding</li> <li>• Web search</li> <li>• Image search</li> <li>• Machine translation</li> <li>• Personalization and contextual search</li> <li>• Logistics</li> </ul>	<ul style="list-style-type: none"> <li>• “Hummingbird” release contextual search</li> <li>• Google Now intelligent assistant</li> <li>• Knowledge Graph/Vault</li> <li>• Spam filtering in Gmail</li> <li>• Self-driving cars</li> </ul>	<ul style="list-style-type: none"> <li>• Deepmind</li> <li>• Dark Blue Labs</li> <li>• Vision Factory</li> <li>• Timeful</li> <li>• DNNResearch</li> <li>• Hinton (U. Toronto)</li> </ul>
	<ul style="list-style-type: none"> <li>• Speech understanding</li> <li>• Cloud services</li> <li>• Personalization and contextual search</li> </ul>	<ul style="list-style-type: none"> <li>• Bing contextual search</li> <li>• Cortana intelligent assistant</li> <li>• Azure Machine Learning Services</li> <li>• Satori Knowledge Base</li> <li>• Microsoft Cognitive Services</li> </ul>	<ul style="list-style-type: none"> <li>• Merging Bing, Cortana and MSR into ~5,000-person AI division</li> </ul>
	<ul style="list-style-type: none"> <li>• Speech understanding</li> <li>• Personalization and contextual search</li> </ul>	<ul style="list-style-type: none"> <li>• Siri intelligent assistant</li> <li>• iOS9 Proactive Suggestions</li> </ul>	<ul style="list-style-type: none"> <li>• Turi</li> <li>• Establishing AI division in Seattle</li> </ul>
	<ul style="list-style-type: none"> <li>• Question answering</li> <li>• Cloud services</li> <li>• Healthcare decision support</li> </ul>	<ul style="list-style-type: none"> <li>• Watson Discovery for Life Sciences</li> <li>• Watson Discovery Services</li> <li>• Watson Health Cloud</li> <li>• Watson for Health</li> </ul>	<ul style="list-style-type: none"> <li>• AlchemyAPI</li> <li>• MergeHealth</li> <li>• Truven</li> </ul>
	<ul style="list-style-type: none"> <li>• Face detection &amp; recognition</li> <li>• Personalization and contextual search</li> <li>• Question answering</li> </ul>	<ul style="list-style-type: none"> <li>• Facebook Open Graph</li> <li>• Facebook Graph Search</li> <li>• Face recognition in Facebook, Instagram</li> </ul>	<ul style="list-style-type: none"> <li>• Wit.ai</li> <li>• LeCun (NYU)</li> <li>• Bottou (Microsoft Research)</li> </ul>

## WHAT DOES THIS MEAN FOR THE FUTURE OF METADATA?

- Our focus as a community has been on helping people find and use information on the Web
- Metadata standards and machines have been a means to that end
- The emergence of machine intelligence and machine reading in the second machine age will make it even easier to automate the production of metadata to help people find, filter and organize information
- Things information workers do in less than a second that we can train machines to do: read a page for concepts and facts, recognize the type of image on a page, ...

# ELSEVIER'S BUSINESS: PROVIDING ANSWERS FOR RESEARCHERS, DOCTORS AND NURSES



My work is moving towards a new field; what should I know?

- Journal articles, reference works, profiles of researchers, funders & institutions
- Recommendations of people to connect with, reading lists, topic pages



How should I treat my patient given her condition & history?

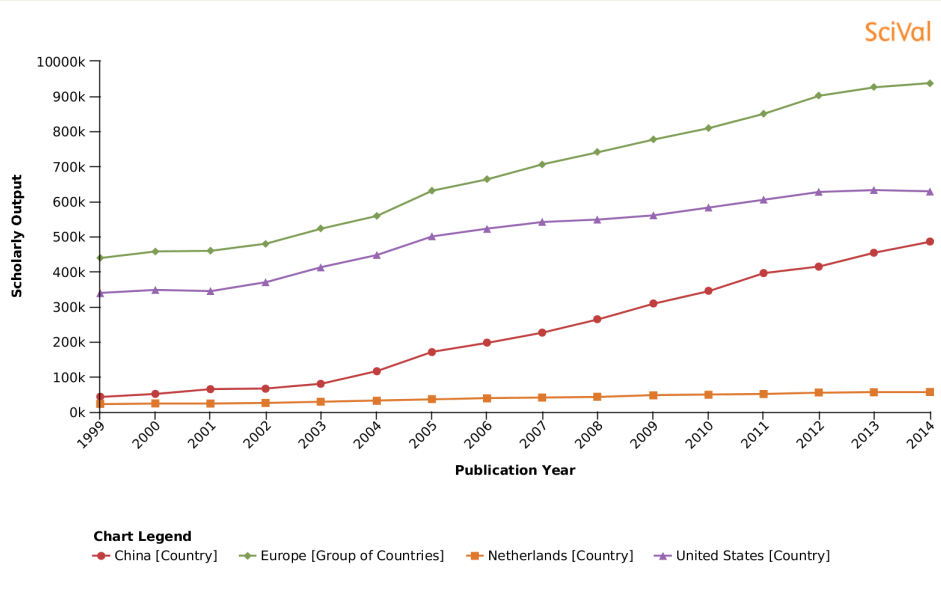
- Journal articles, reference works, medical guidelines, electronic health records
- Treatment plan with alternatives personalized for the patient



How can I master the subject matter of the course I am taking?

- Course syllabus, reference works, course objectives, student history
- Quiz plan based on the student's history and course objectives

# THE GROWTH OF SCIENCE COMPLICATES OUR EFFORTS



**karencoye**  
@karencoye



Following

Science information overload has been the impetus for changes in knowledge organization for nearly 2 centuries. Here it is again.

**Roxanne Khamsi** @rkhamsi

Mission (totally) impossible: "To keep up with the cancer literature one would have to read 17 articles per waking hour... 365 days a year" [twitter.com/DrKhouryCDC/st...](https://twitter.com/DrKhouryCDC/status/784111111111111111)

RETWEETS

2

LIKES

4



10:40 AM - 8 Oct 2016



2



4



# ANSWERS ARE ABOUT THINGS, NOT JUST WORKS



Why shouldn't a search on an author return information about the author, including the author's works? Where was the author born, when did she live, what is she known for? ... All of this is possible, but only if we can make some fundamental changes in our approach to bibliographic description. ... The challenge for us lies in transforming what we can of our data into interrelated "things" without overindulging that metaphor.

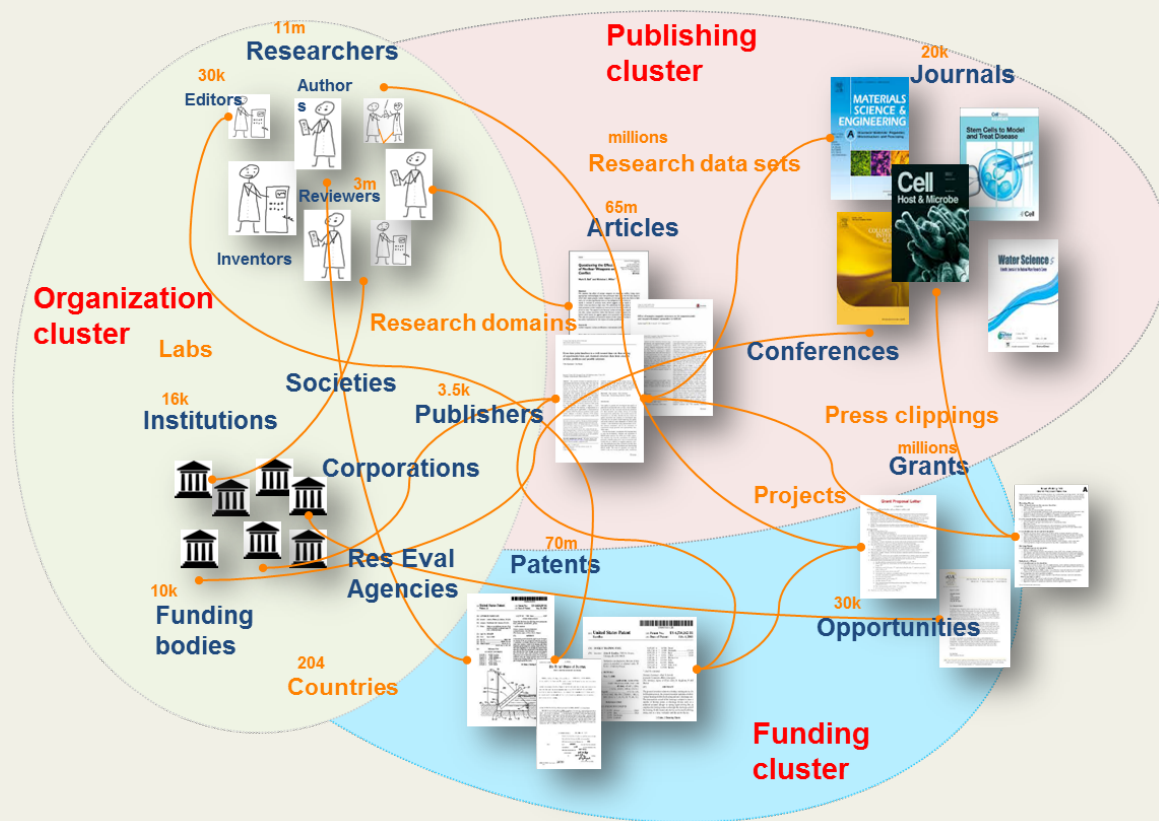
Coyle, K. (2016). *FRBR, before and after: a look at our bibliographical models*. Chicago: ALA Editions.



# KNOWLEDGE GRAPHS AND MACHINE READING TURN CONTENT INTO ANSWERS

- Knowledge graphs are "graph structured *knowledge bases* (KBs) which store factual information in form of relationships between entities" (Nickel, M., Murphy, K., Tresp, V. and Gabrilovich, E. (2015). A review of relational machine learning for knowledge graphs. arXiv:1503.00759v3)
- Knowledge graphs are metadata evolved beyond the focus on the work, linking people, concepts, things and events
- Knowledge graphs organize data extracted from content through machine reading so that queries can provide answers

# ELSEVIER: KNOWLEDGE GRAPHS FOR RESEARCH



# ELSEVIER: KNOWLEDGE GRAPHS FOR LIFE SCIENCES

Biological Pathways extracted via semantic text mining

Bioactivities through text analysis

Chemical Structures And Properties



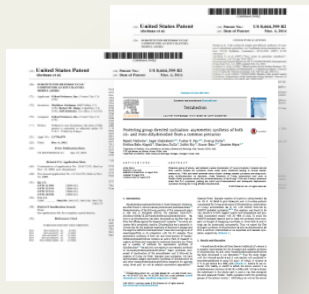
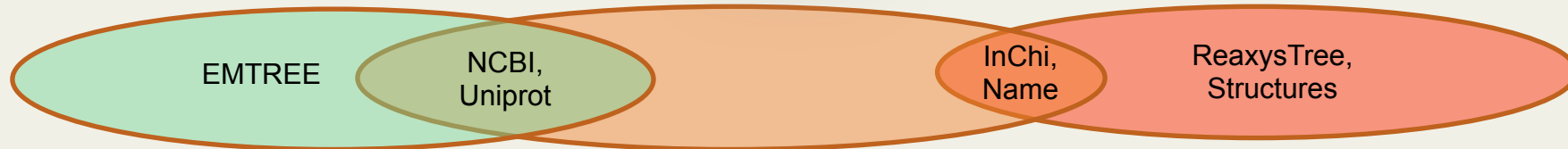
A upregulates B

B upregulates C

C increases disease D

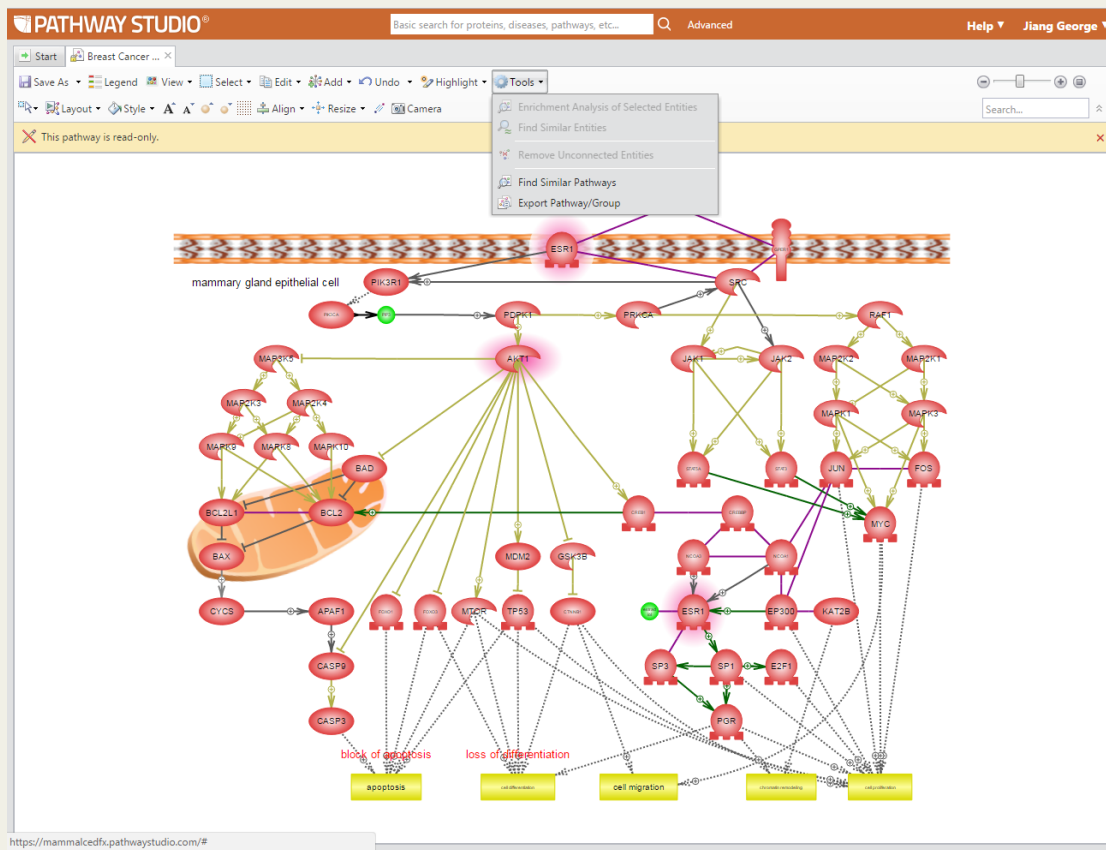
$A \Rightarrow B \Rightarrow C \Rightarrow D$

$IC_{50}$  6.3nM, kinase binding assay  
10mM concentration

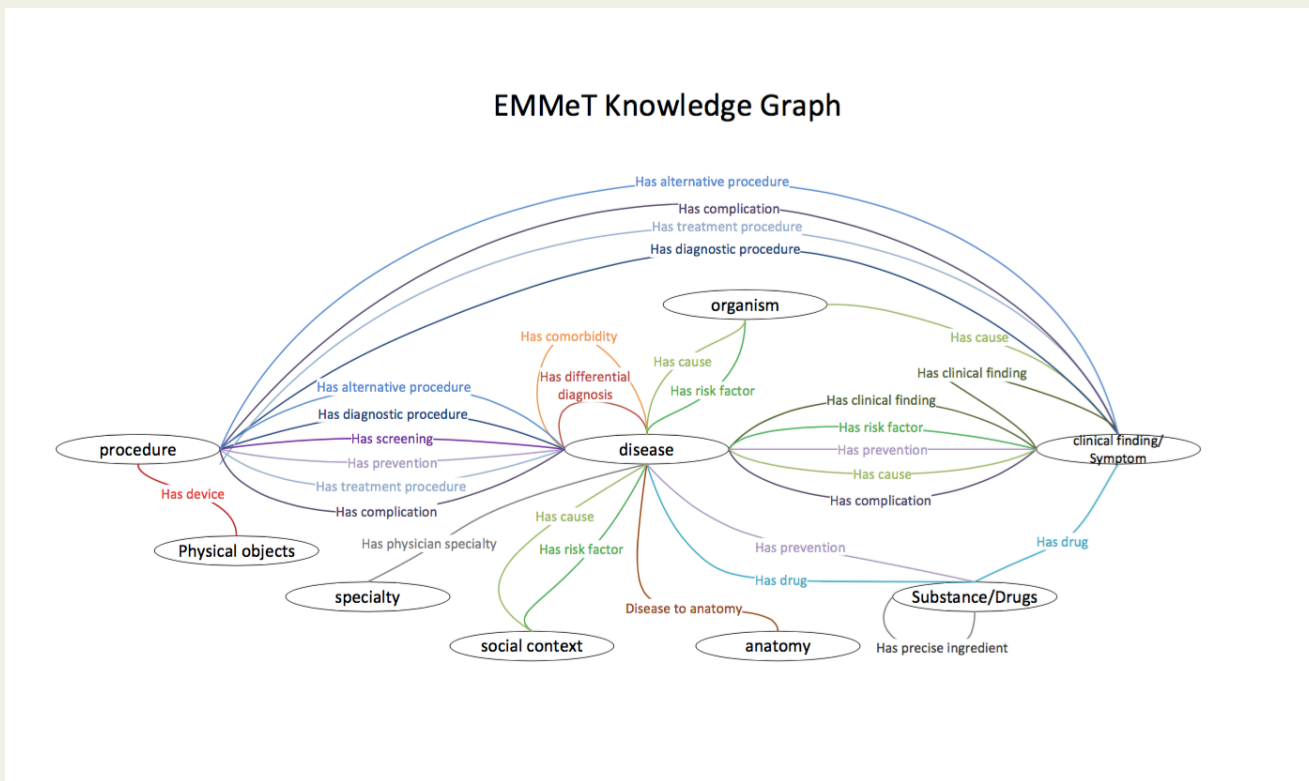
A table with multiple columns. The first column contains chemical structures. The second column contains numerical values. The table is organized into rows, each corresponding to a different chemical structure.

Normalizing vocabularies required: proteins, diseases, drugs, chemicals

# ELSEVIER: KNOWLEDGE GRAPHS FOR LIFE SCIENCES



# ELSEVIER: KNOWLEDGE GRAPHS FOR HEALTHCARE



# ELSEVIER: USING JSON-LD FOR ANNOTATIONS

```

{
  "@context": [ "http://vtw.elsevier.com/metadata/context.jsonld",
    { "eaf": "http://vtw.elsevier.com/data/ns/Formats/Annotation",
      "oca": "http://ontochem.com/ns/ann"
    }
  ],
  "@type": "eaf:AnnotationDocument",
  "eaf:annotatedSourceInfo": {
    "dct:identifier": "E-602148748",
    "dct:title": "A rare mutation in MYH7 gene occurs with overlapping phenotype",
    "dct:type": "http://data.elsevier.com/vocabulary/ElsevierContentTypes/1.3"
  },
  "eaf:annotationProvenance": {
    "dct:format": "application/xhtml+xml",
    "cef:enrichmentTool": "OCMiner",
    "cef:enrichmentToolVersion": "RxBT edition 2.2.Milestone6",
    "cef:dictionary": "ReaxysTree",
    "cef:dictionaryVersion": "13.0",
    "prov:generatedAtTime": "2016-02-16T16:11:07Z"
  },
  "eaf:annotations": [
    {
      "eaf:localID": "1",
      "eaf:sourceText": "beta-myosin",
      "eaf:conceptID": "",
      "eaf:annotationType": "concept",
      "eaf:annotationScope": "text range",
      "eaf:range": "2726-2737",
      "eaf:confidence": "77",
      "eaf:relevance": "72",
      "oca:d": "chemCompound",
      "oca:CNType": "Compound",
      "oca:CNHasStructure": "NO",
      "oca:CNSubType": "Trivial"
    }, {
      "eaf:localID": "2",
      "eaf:sourceText": "liquid",
      "eaf:conceptID": "817100004230",
      "eaf:annotationType": "concept",
      "eaf:annotationScope": "text range",
      "eaf:range": "7448-7454",
      "eaf:confidence": "65",
      "eaf:relevance": "22",
      "oca:d": "reaxystree",
      "oca:p": "liquid"
    }
  ]
}

```

# ELSEVIER: MACHINE LEARNING FOR LINK DISCOVERY

diseases 2791370 glaucoma  
 diseases 2791370 glaucoma  
 diseases 2791370 glaucoma  
 diseases 2791370 glaucoma  
 diseases 2791370 glaucoma

have been documented to cause  
 is assessed through  
 progresses more rapidly than  
 recommend  
 supports the assumption that  
 is the death of

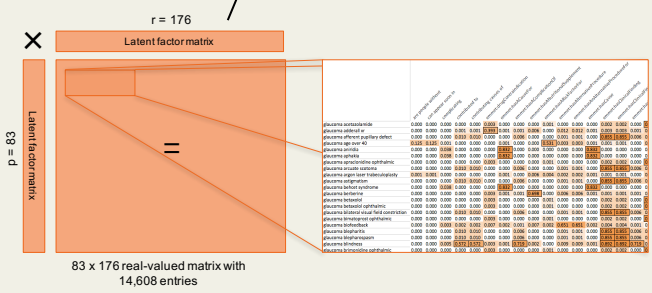
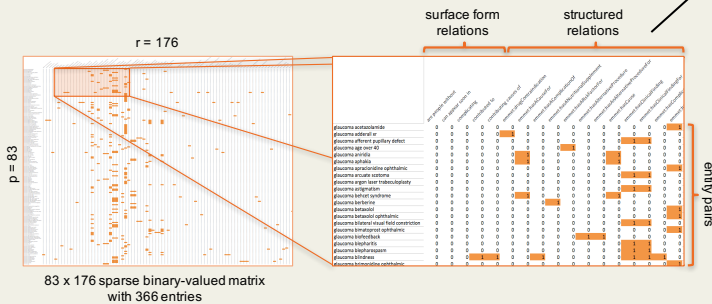
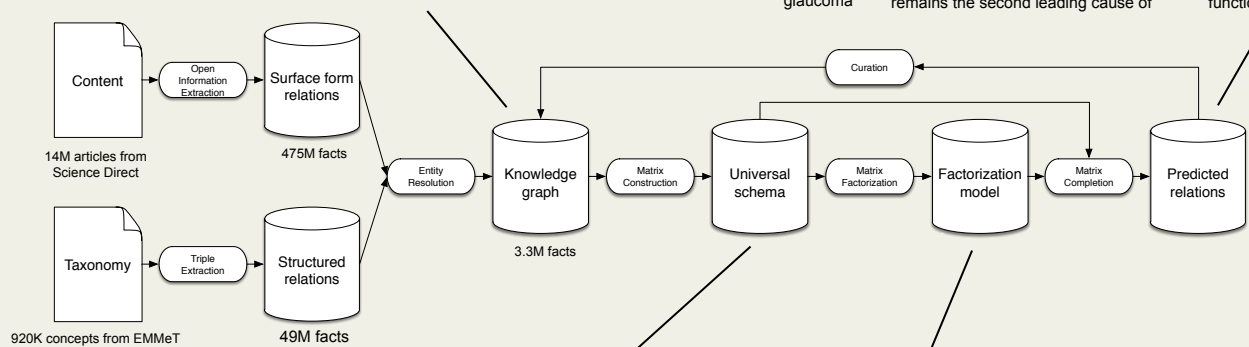
contact dermatitis  
 evaluation  
 primary open-angle glaucoma  
 treatment  
 oxidative stress  
 retinal ganglion cells

3815093  
 5415395  
 8247149  
 5216597  
 8184588  
 8002088

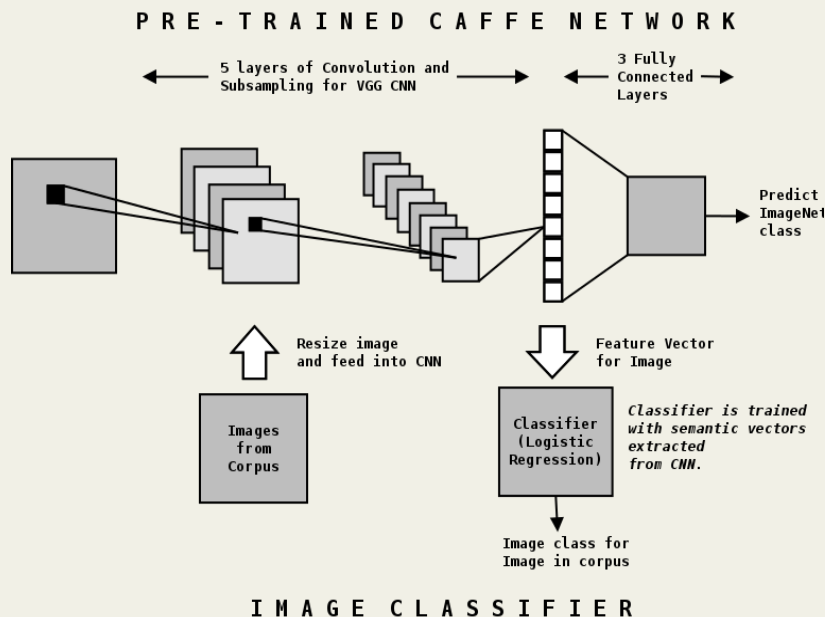
diseases  
 qualifier  
 diseases  
 procedures  
 diseases  
 anatomy

developed many years after  
 chronic inflammation of uveal tract  
 can appear soon in  
 can appear soon in  
 the risk of  
 contributing causes of  
 contributed to  
 is considered the second leading cause of  
 remains the second leading cause of

chronic inflammation of uveal tract  
 family history of glaucoma  
 age over 40  
 functional visual field loss  
 functional visual field loss  
 functional visual field loss  
 functional visual field loss



# ELSEVIER: DEEP LEARNING FOR IMAGE ANNOTATION



**ClinicalKey** English CME Login Register

All Types **hypertension** Books Journals More

Filter By: Source Type Study Type Specialties Date

34991 results **[+] Rate Results**

Relevance Subscribed Content

**CLINICAL OVERVIEW**  
**Hypertension**  
 Synopses Terminology Diagnosis Treatment Complications and Prognosis Screening and Prevention

**FIRST CONSULT**  
**Hypertension**  
 Yonghong Huan, MD, Assistant Professor of Medicine, Renal, Endocrine and Hypertension Division, University of Pennsylvania Health System, Philadelphia, Pennsylvania; Ruben J. Nazaro, MD, MA, Contributing Editor, First Consult. Published January 4, 2014. Last updated September 18, 2013.

**FIRST CONSULT**  
**Hypertension in children**  
 Charles Kwon, MD, Director, Center for Pediatric Nephrology, Cleveland Clinic, Cleveland, Ohio; Ruben Nazaro, MD, MA, Published February 19, 2014. Last updated September 17, 2013.

Searches related to hypertension  
 pulmonary hypertension pulmonary hypertensive arterial disease essential hypertension portal hypertension intracranial hypertension

**BOOK CHAPTER**  
**Hypertension**  
 Rosen's Emergency Medicine. Lavy, Philip D. Published January 1, 2014. Pages 1113-1123. © 2014.

**Full Text Article**

**Selected Content (6)** **CC BY**

Comparison of Hypertension Guidelines  
 Mayo Clinic Proceedings - February 2015

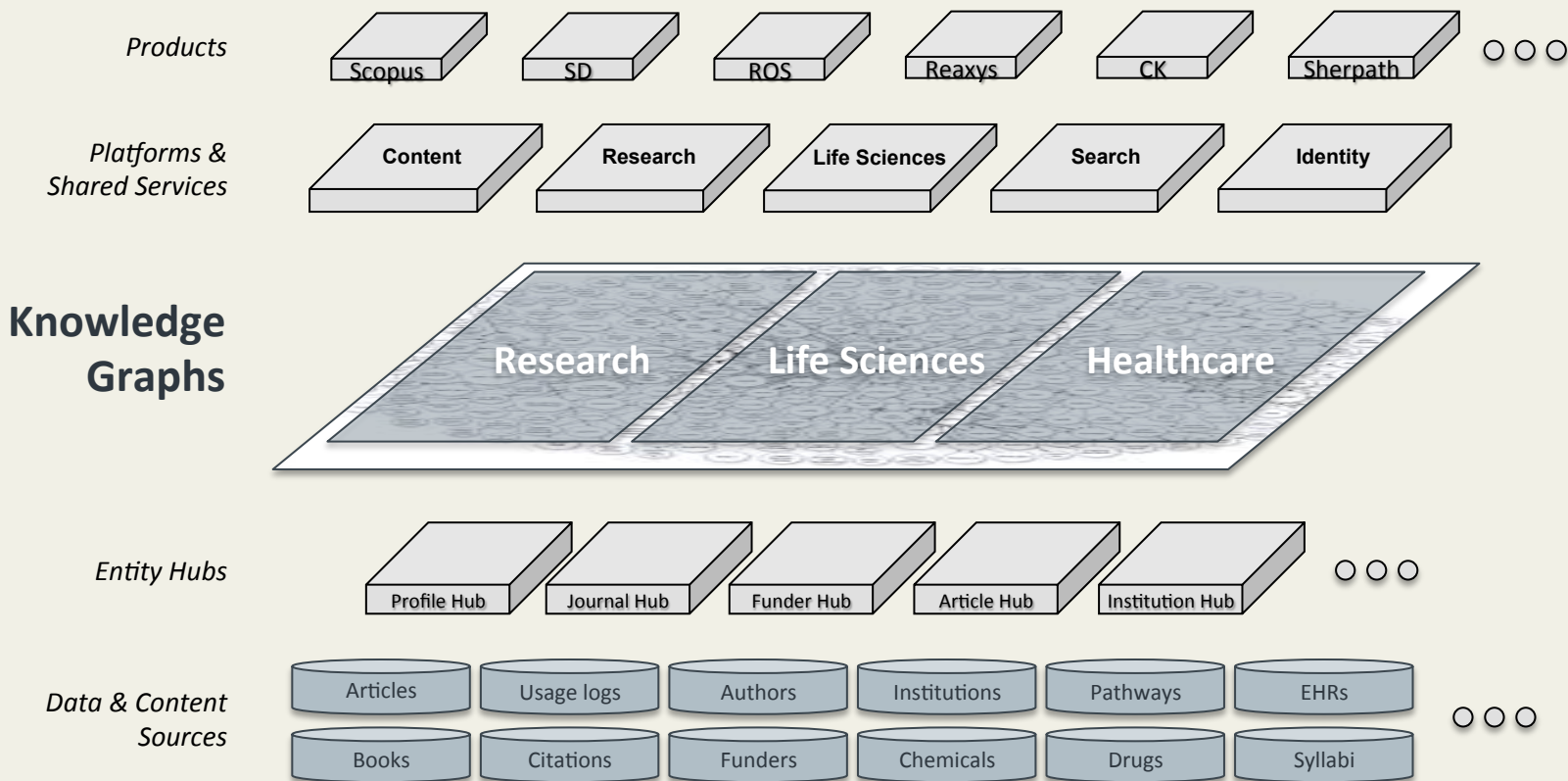
Standardized effects of a 10 mm Hg reduction in systolic blood pressure stratified by blood pressure  
 The Lancet - March 2016

Evidence-based dosing of antihypertensive medications  
 Medical Clinics of North America - July 2015

Was this helpful? Yes or No



# ELSEVIER'S KNOWLEDGE PLATFORM



## MOVING FORWARD: QUESTION ANSWERING AS AN AI-COMPLETE PROBLEM



Question answering (QA) is a complex natural language processing task which requires an understanding of the meaning of a text and the ability to reason over relevant facts. Most, if not all, tasks in natural language processing can be cast as a question answering problem ...

Kumar, A., Ondruska, P., Iyyer, M., Bradbury, J., Gulrajani, I., Zhong, V., Paulus, R. & Socher, R. (2016). Ask Me Anything: Dynamic Memory Networks for Natural Language Processing. Proceedings of the 33rd International Conference on Machine Learning (ICML 2016).

# THE BATTLE FOR THE KNOWLEDGE GRAPH



I really believe that the key battleground in any industry is that of its knowledge graph. Google has it for media/advertising, Netflix has it for filmed entertainment, Uber has it for inner city transportation, Facebook has it across social media as well as messaging and the multiples speak for themselves.

Tony Askew, Founder/Partner at REV (personal communication, September 29, 2016)

# METADATA STANDARDS ARE AND WILL CONTINUE TO BE FOUNDATIONAL FOR KNOWLEDGE GRAPHS

- Semantic Web and Linked Data standards are the consensus approach for representing and sharing knowledge graphs over the Web
- There is a single thread from the establishment of the Dublin Core through Open Linked Data to the emergence of knowledge graphs
  - Linked data principles
  - Standard vocabularies, taxonomies and ontologies

# BUT WHAT DO METADATA STANDARDS DO FOR MACHINE READING?

- Machines' proficiency in constructing knowledge graphs from text, audio, images and video will depend on our ability to train them effectively to read information from the Web
- How machines read the Web today
  - Crawling and indexing Web resources, possibly semantically tagged (e.g. using schema.org)
  - Find-and-follow crawling of open linked data resources for ontology and data sharing and reuse
  - Programmatic access to APIs mediated through HTTP/S and other Internet protocols

## THE SEMANTIC WEB WAS INTENDED FOR MACHINE READING



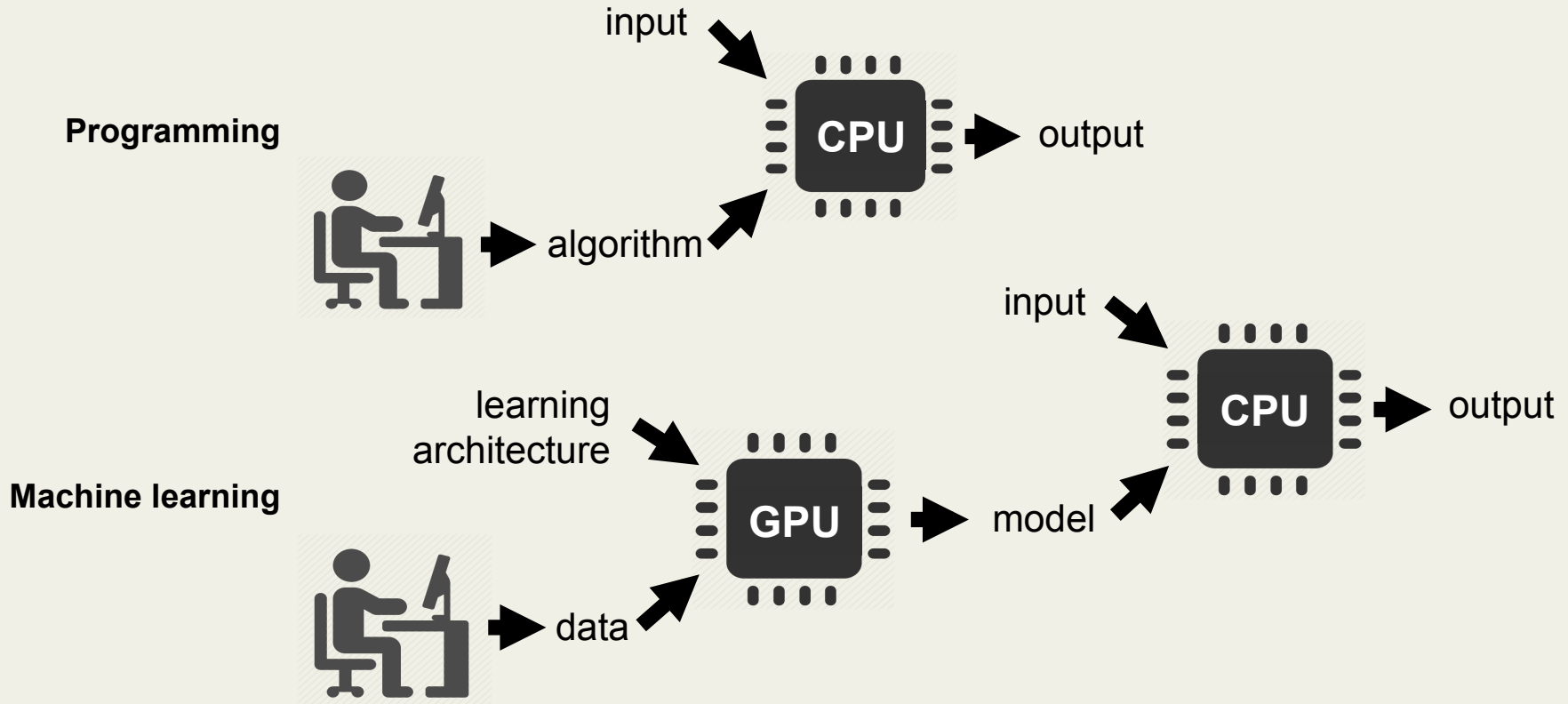
... that's the real idea behind the Semantic Web: letting software use the vast collective genius embedded in its published pages.

Swartz, A. (2013). Aaron Swartz's A programmable Web: An unfinished work. San Rafael, Calif.: Morgan & Claypool Publishers.

## **BUT THE SEMANTIC WEB IS BUILT FOR PEOPLE, NOT MACHINES**

- The Semantic Web is largely a logicist take on the way knowledge is to be represented
- The latest advances in machine intelligence are based on a connectionist approach to knowledge representation
- There is a gap between how knowledge is represented in the Semantic Web and what deep learning is exploiting to such good effect
- The Semantic Web is silent about how machines can become better readers, and hence better partners in the second machine age
- How will we evolve metadata standards to better accommodate machines?

# MACHINE READING IS ENABLED BY MACHINE LEARNING





# MACHINES SEE THINGS DIFFERENTLY THAN PEOPLE

```

81a9c 32 3d 4b 70 b7 5b ef 53 e1 38 ea 40 2a 5e d2 79 df d2 0e 21 cf 88 ba
81acb 3b 7a cf 3e db 7d 31 8d 99 88 04 e1 8d 1c 2d 6d 38 22 37 70 4a 8d bf
81afa 8f cd 2e 1d 8a 9f bc 3f 50 ef 47 e5 4e 84 2d c6 09 79 52 4a 77 22 07
81b29 49 b5 34 b2 2b 53 e0 97 06 e4 ee 22 3d fd b1 e9 f8 72 b0 62 ee ee bc
81b58 13 8e 6f 5f 73 21 0d 7f ba d8 17 14 6d 25 5e 7a 91 72 6c 59 d9 ba 69
81b87 e3 23 3a ac ea a6 a0 55 d2 7c 4d 0a 3c cb 71 63 58 e2 26 49 3f 94 63
81bb6 27 ca 9a 74 21 64 a7 68 09 9d c9 fa 1f 8e 38 5d 77 05 90 63 ce f3 f5
81be5 54 7f 48 38 e6 30 5a d7 39 ad 6f 52 79 5d 04 d3 be 3c 27 16 f5 a5 52
81c14 27 b0 05 b2 3e f8 f4 a8 08 c0 cb 82 31 d1 e4 ee bf a7 65 c8 e3 63 0c
81c43 a8 cb 74 4d 78 31 85 c9 c1 8d 34 7a 93 a2 af 4f 28 d1 3f 87 1a 52 c6
81c72 b0 f8 47 1d d7 a5 e8 b1 b9 b0 ed be 13 81 96 a8 fa 65 9b ae 75 cf b4
81ca1 20 c9 8b d3 9b c6 6b 5e 63 c8 f7 65 22 8f 42 5a 44 84 90 21 49 dc 1e
81cd0 1a 9b 5d ed a3 69 a9 65 b7 c2 54 15 a2 24 09 de 67 d7 db 91 38 bf 9e
81cff cb ef 43 5e 2d 59 d6 da 76 48 2a 52 47 1d 80 27 0d 7e b0 3f d3 da d7
81d2e 09 fd fa 6c 4d 78 44 27 85 e9 00 c7 e4 71 c7 f8 2f 16 4c dd 4b 22 ba
81d5d cb 4c a8 3e 52 be 55 ce de bb e3 d4 f0 80 43 6e 27 f4 0b 87 d5 32 24
81d8c 51 9f b9 02 7d b1 d3 45 83 17 95 bd 70 8f cb 91 d3 9a 3d 57 a0 f2 a6
81ddb 63 8e d5 1f 1c 99 1b 01 5d 96 81 2c 98 63 cc 0b 09 ea 46 6e ae 46 7a
81dea af 8c 35 19 4e a8 25 8c f6 0a 53 e0 6d 3d 49 b4 37 5f 67 a8 02 b6 dc
81e19 99 80 fd a5 e8 de 8a 84 24 14 7e d3 d1 25 2c a4 13 c1 29 d3 09 3e d3
81e48 56 cc ea aa 57 9e 0d 8a 67 11 ad 71 04 05 7a 8f 4f fb b1 df 66 e3 9c

```

(a) hex dump of picture of a lion

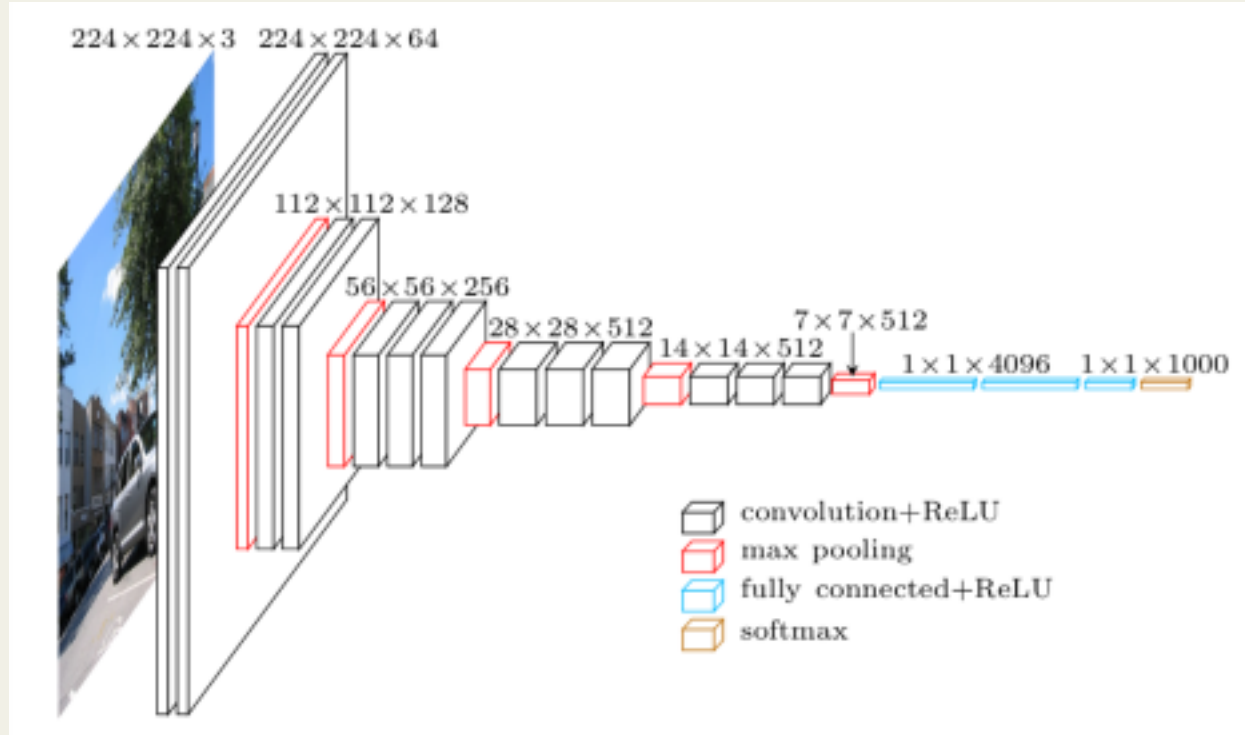


(b) same lion in human-readable format

Figure 1: The hex dump represented on the left has more information contents than the image on the right. Only one of them can be processed by the human brain in time to save their lives. Computational convenience matters. Not just entropy.

From: Alain, G. and Bengio, Y. (2016). Understanding intermediate layers using linear classifier probes. arXiv:1610.01644v1.

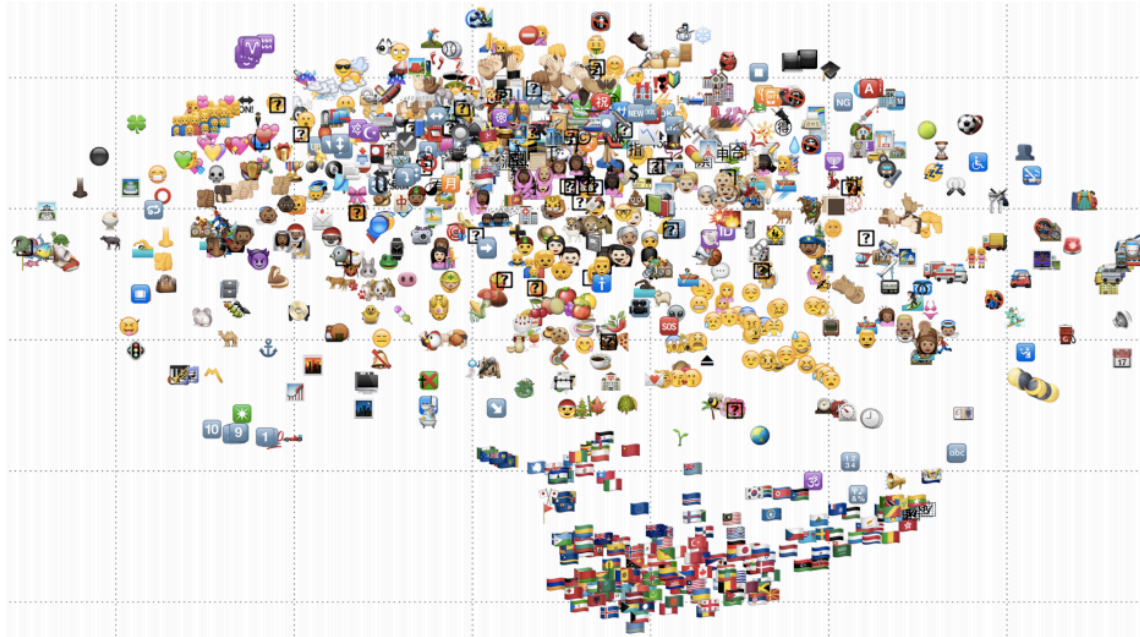
# MACHINES LEARN THINGS DIFFERENTLY THAN PEOPLE



# MACHINE LEARNING DATASETS AND MODELS ARE BECOMING PART OF THE WEB

- Machines need lots and lots of data to learn how to read
- Datasets with ad-hoc formats are being made openly available
  - Open Images “~9 million URLs to images that have been annotated with labels spanning over 6000 categories” (The Open Images Dataset. (n.d.). Retrieved September 29, 2016, from <https://github.com/openimages/dataset>.)
  - YouTube-8M : “8 million YouTube video URLs (representing over 500,000 hours of video), along with video-level labels from a diverse set of 4800 Knowledge Graph entities” (Vijayanarasimhan S. and Natsev, P. (2016). Announcing YouTube-8M: A Large and Diverse Labeled Video Dataset for Video Understanding Research. Retrieved September 29, 2016, <https://research.googleblog.com/2016/09/announcing-youtube-8m-large-and-diverse.html>.)
  - Stanford Natural Language Inference: “570k human-written English sentence pairs manually labeled for balanced classification with the labels *entailment*, *contradiction*, and *neutral*, supporting the task of natural language inference” (The Stanford Natural Language Inference (SNLI) Corpus. (n.d.). Retrieved September 29, 2016, from <http://nlp.stanford.edu/projects/snli/>.)
- Standard architectures for machine (deep) learning are being released as open source
  - Dense neural networks for classification
  - Convolutional neural networks for image, audio and video recognition
  - Recurrent neural networks for sequence processing and generation
- Advances in the field are being published quickly and transferred to industrial application just as quickly

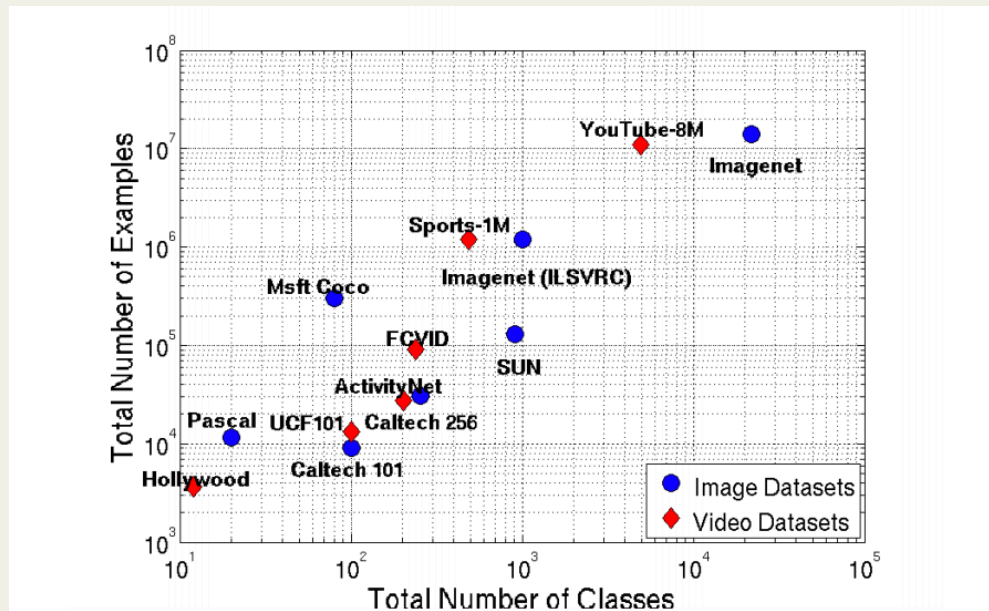
# VOCABULARIES ARE SETS OF VECTOR EMBEDDINGS



**Figure 3:** Emoji vector embeddings, projected down into a 2-dimensional space using the t-SNE technique. Note the clusters of similar emoji like flags (bottom), family emoji (top left), zodiac symbols (top left), animals (left), smileys (middle), etc.

From: Eisner, B., Rocktäschel, T., Augenstein, I., Bošnjak, M. and Riedel, S. (2016). Emoji2vec: learning emoji representations from their description. arXiv:1609.08359v1.

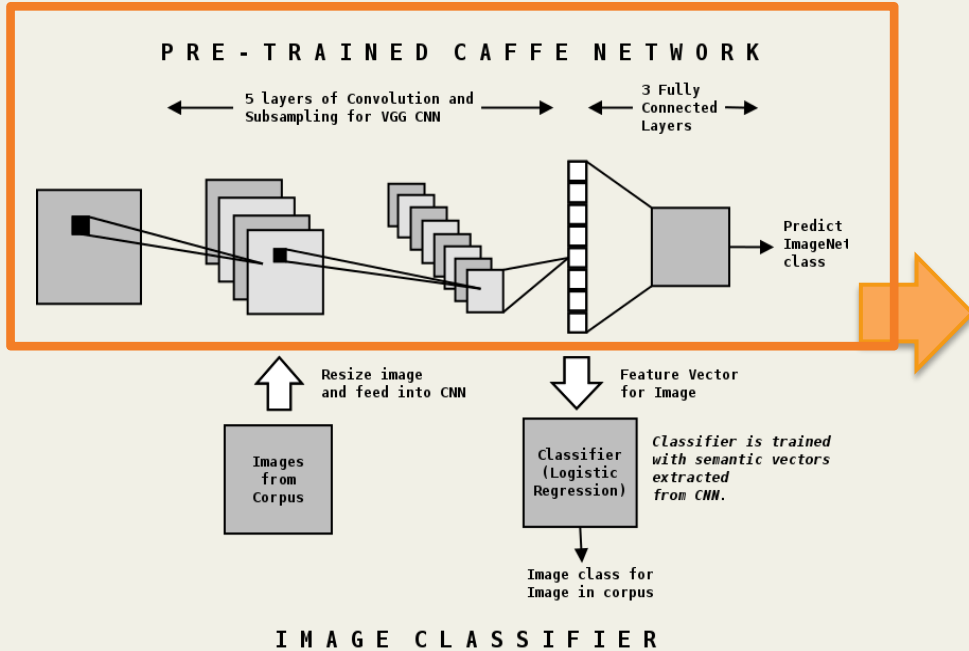
# TRAINING DATASETS ARE GROWING IN VOLUME AND COVERAGE



**Figure 2: The progression of datasets for image and video understanding tasks. Large datasets have played a key role for advances in both areas.**

From: Abu-El-Haija, S., Kothari, N., Lee, J., Natsev, P., Toderici, G., Varadarajan, B. and Vijayanarasimhan, S. YouTube-8M: a large-scale video classification benchmark. arXiv:1609.08675.

# MODELS ARE BECOMING REUSABLE DATA RESOURCES THEMSELVES



**ClinicalKey** English CME Login Register

Search: **hypertension**

34991 results

**Hypertension**

**CLINICAL OVERVIEW**

- Synopsis
- Terminology
- Diagnosis
- Treatment
- Complications and Prognosis
- Screening and Prevention

**FIRST CONSULT**

**Hypertension**

Yonghong Huai, MD, Assistant Professor of Medicine, Renet, Electrolyte and Hypertension Division, University of Pennsylvania Health System, Philadelphia, Pennsylvania; Ruben J. Nazaro, MD, MA, Contributing Editor, First Consult, Published January 4, 2014. Last updated September 16, 2013.

**FIRST CONSULT**

**Hypertension in children**

Charles Kwon, MD, Director, Center for Pediatric Nephrology, Cleveland Clinic, Cleveland, Ohio; Ruben Nazaro, MD, MA, Published February 19, 2014. Last updated September 17, 2013.

Searches related to hypertension

- pulmonary hypertension
- pre-eclampsia
- essential hypertension
- portal hypertension
- pulmonary hypertensive arterial disease
- intracranial hypertension

**BOOK CHAPTER**

**Hypertension**

Rosen's Emergency Medicine, Levy, Philip D., Published January 1, 2014. Pages 1113-1123 ad. © 2014.

**Selected Content (6)**

- Comparison of Hypertension Guidelines
- Mayo Clinic Proceedings - February 2015
- Standardised effects of a 10 mm Hg reduction in systolic blood pressure stratified by blood pressure
- The Lancet • March 2016
- Evidence-based dosing of antihypertensive medications
- Selected Cases of Health America • July 2015

Yes this helped? Yes or No

# THE CHALLENGE FOR THE METADATA COMMUNITY: LINKED DATA THAT MACHINES CAN LEARN FROM

- We should be able to keep the core linked data principles that allow us to leverage Web architecture
  - URI as identifiers support (re)use of data in place (e.g. as in Google Open Images)
- We need to understand whether linked data formats need to change to support the needs of machine readers
  - The need for n-ary relations
  - The need for efficient indexing
- We need to investigate how vocabulary management can adapt to support the needs of machine reading
  - Vocabularies define what machines can recognize
  - Vocabularies need to accommodate lexical and vector representations together

# WORK AT ELSEVIER: RESEARCH DATA MANAGEMENT



DataSearch

Filter Results reset 29129 results for *frog phylogeny*

Types

- Image (24252)
- Tabular Data (12621)
- Document (2994)
- Raw Data (595)
- File Set (307)
- Slides (117)
- Video (63)
- Statistical Data (16)

Molecular phylogeny of Malagasy reed frogs, *Heterixalus*, and the relative performance of bioacoustics and color-patterns for resolving their systematics  
*Katharina C. Wallenberg, Frank Glaw, Axel Meyer & Miguel Vences - 2007-06-22*  
 The members of the genus *Heterixalus* constitute one of the endemic frog radiations in Madagascar. Here we present a complete species-level phylogeny based on DNA sequences (4876 base pairs) of three nuclear and four mitochondrial markers to clarify the phylogenetic relationships among...

DOCUMENT IMAGE TABULAR DATA

Data from: Phylogeny of frogs of the *Physalaemus pustulosus* species group, with an examination of data incongruence  
*Cannatella, David C., Hillis, David M., Chippindale, Paul T., Weigt, Lee, Rand, A. Stanley & Ryan, Michael J. - 1998-06-01*

MENDELEY DATA Browse My datasets New dataset Paul Groth

## Reproducible experiments on dynamic resource allocation in cloud data centers

Published: 13 Dec 2015 | Version 6 | DOIs: 10.17632/x26gv65m6d.6 Viewed 957 Downloaded 101  
 Contributor(s): Andreas Wolke

Description of this data

In Wolke et al. we compare the efficiency of different resource allocation strategies experimentally. We focused on dynamic environments where virtual machines need to be allocated and deallocated to servers over time. In this companion paper, we describe the simulation framework and how to run simulations to replicate experiments or run new experiments within the framework.

Experiment data files Download all files (6)

Results.zip	63 KB	↓
github.paper.IS2015-master.zip	8 MB	↓
github.workload-master.zip	222 MB	↓

This data is associated with the following peer reviewed publication:  
 Reproducible experiments on dynamic resource allocation in cloud data centers

Cite this article

Published in: Information Systems



## THE OPPORTUNITY FOR LIBRARIANS AND PUBLISHERS

As machines become increasingly capable of general-purpose language understanding, the burden of effort in building machine intelligences will shift from software engineering to the acquisition, organization and curation of training content and data.

# SAVE THE TIME OF THE MACHINE READER



Perhaps this law is not so self-evident as the others. None the less, it has been responsible for many reforms in library administration and has a great potentiality for effecting many more reforms in the future.

Ranganathan, S.R. (1931). The five laws of library science. Madras: The Madras Library Association.

# SCHOLARLY PUBLISHING IN THE SECOND MACHINE AGE



ELSEVIER

Publishing  
content  
for people



context

content



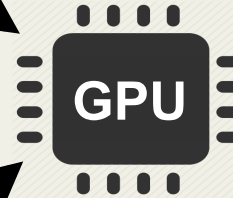
knowledge

Publishing  
data for  
machines



learning architecture

data



model

## IN SUMMARY: THE ROLE OF METADATA IN THE SECOND MACHINE AGE

- We need to save the time of the (human) reader by providing answers, not content
- This requires a shift from document-centric to entity-centric management of knowledge
- Knowledge graphs are "vehicles of communication" between people and machines and the battleground on which dominance in markets will be established
- We can use machine reading to build knowledge graphs
- The Semantic Web falls short in helping machines learn how to read
- Metadata standards to support machine (deep) learning should receive attention from the metadata community
- Our goal: making information resources on the Web discoverable and comprehensible for both people and machines

# THANK YOU

Bradley P. Allen

Chief Architect, Elsevier

@bradleypallen

# IMAGE ATTRIBUTIONS

- p. 2: <https://twitter.com/stuartweibel>
- p. 3: <https://vimeo.com/66143280>
- p. 5: <http://www.prospectmagazine.co.uk/other/erik-brynjolfsson-andrew-mcafee>
- p. 6: <https://arxiv.org/pdf/1411.4555v2.pdf>
- p. 6: [http://media.wix.com/ugd/142eb4\\_7581cfcf090e4e31a52599315f77c648.pdf](http://media.wix.com/ugd/142eb4_7581cfcf090e4e31a52599315f77c648.pdf)
- p. 6: <http://www.nature.com/nature/journal/v529/n7587/full/nature16961.html>
- p. 7: <https://www.wired.com/2013/05/neuro-artificial-intelligence/>
- p. 11: <https://twitter.com/karencoye/status/784810709641158656>
- p. 22: <http://www.pcworld.com/article/3051631/salesforce-buys-ai-specialist-metamind-to-avoid-being-flanked.html>
- p. 23: <https://twitter.com/tonyaskew>
- p. 26: <http://quotesgram.com/aaron-swartz-quotes/>
- p. 29: <https://arxiv.org/pdf/1610.01644.pdf>
- p. 30: <https://blog.heuritech.com/2016/02/29/a-brief-report-of-the-heuritech-deep-learning-meetup-5/>
- p. 32: <https://arxiv.org/pdf/1609.08359v1.pdf>
- p. 33: <https://arxiv.org/pdf/1609.08675v1.pdf>
- p. 38: <https://plus.google.com/photos/112452283652920158570/album/6059593062673732337/6059593180461846098>
- Save corporate logos on p. 8, all other images © 2016 Elsevier, Inc.