Accelerating Innovation at the Confluence of Human-Curated Content and Artificial Intelligence

CAS®
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

www.cas.org

Mark R. Grabau
Chief Analytics Officer, CAS

Accelerating innovation improves human outcomes

- Cure diseases
- Save resources
- Improve quality of life

Copyright 2019 American Chemical Society. All rights reserved.

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

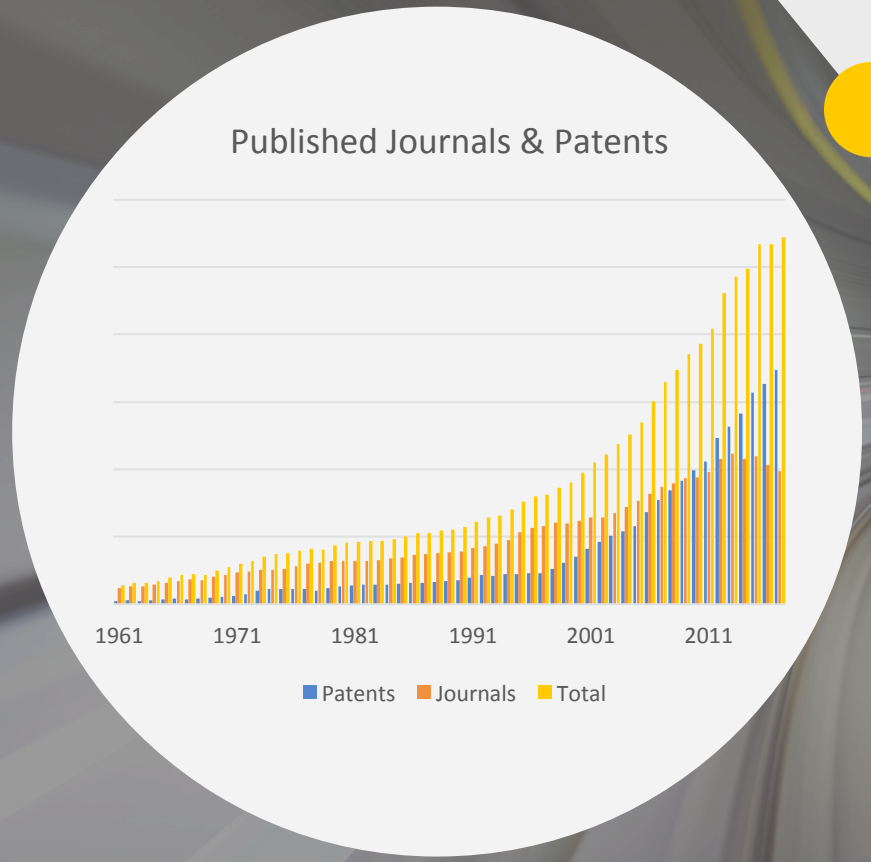
Data is both a blessing and a curse to innovation

The Data Paradox

Copyright 2019 American Chemical Society. All rights reserved.

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

Increasing volume, complexity and connections in the information landscape require new solutions

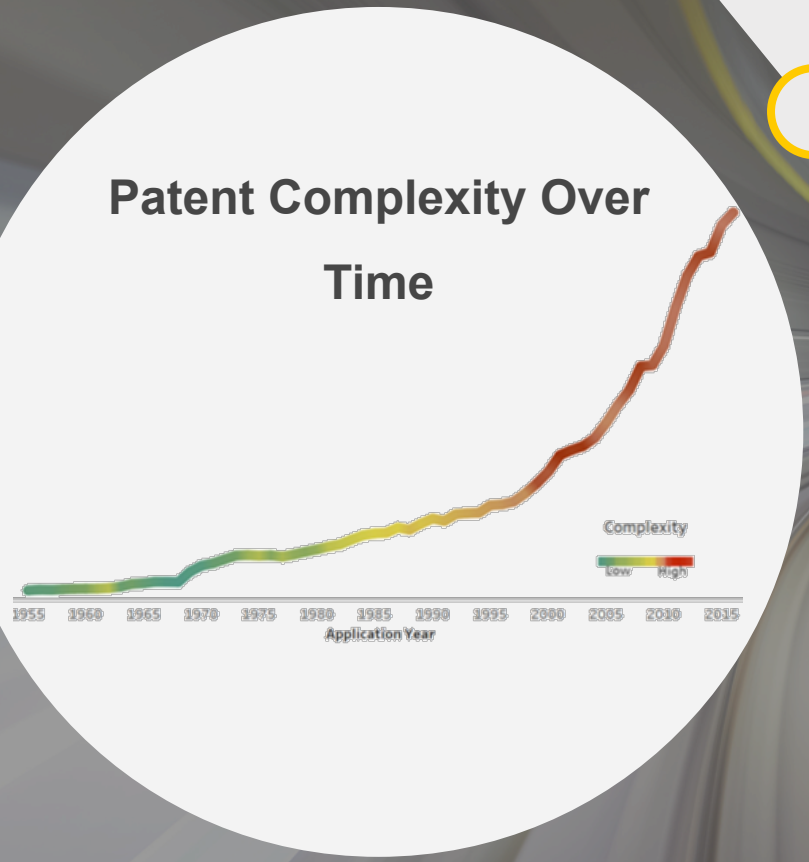


- Volume
- Complexity
- Connectedness

Copyright 2019 American Chemical Society. All rights reserved.

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

Increasing volume, complexity and connections in the information landscape require new solutions

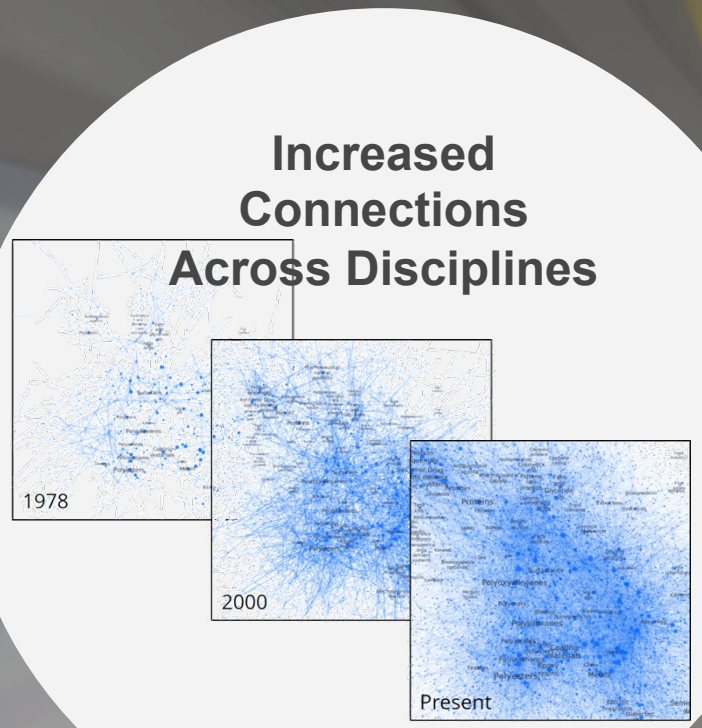


- Volume
- Complexity
- Connectedness

Copyright 2019 American Chemical Society. All rights reserved.

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

Increasing volume, complexity and connections in the information landscape require new solutions



- Volume
- Complexity
- Connectedness

Copyright 2019 American Chemical Society. All rights reserved.

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

Artificial Intelligence has great potential to help us leverage all of this information to accelerate innovation

- Identifying new opportunities
- Increasing efficiency
- Finding unseen connections

Copyright 2019 American Chemical Society. All rights reserved.

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

It’s why everyone is investing in digitalization and AI

89% of businesses plan to invest in AI by 2020

Deloitte’s Digital Disruption Index

But 85% of AI projects are not meeting expectations

Gartner

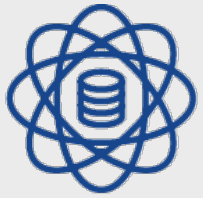
Copyright 2019 American Chemical Society. All rights reserved.

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

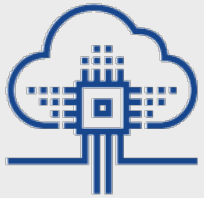
CAS is a specialist in scientific information solutions

We provide products and services that power discovery to solve our world's biggest challenges by helping organizations predict, plan and protect their innovations

With over 110 years of experience, no one knows more about scientific information and related technology than CAS



UNPARALLELED
SCIENTIFIC CONTENT



SPECIALIZED
TECHNOLOGY

UNMATCHED
HUMAN EXPERTISE

Copyright 2019 American Chemical Society. All rights reserved.

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

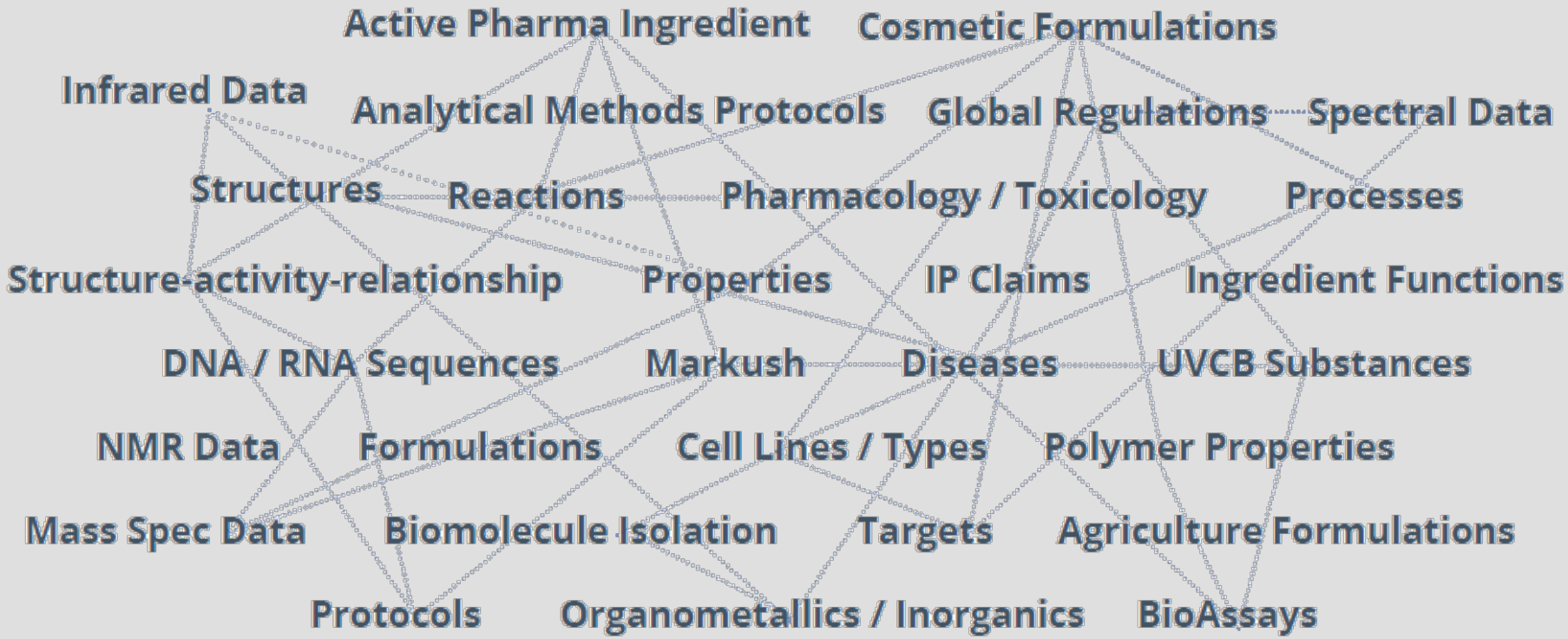
Our global scientific coverage is unmatched

50 THOUSAND scientific journals and documents

63 PATENT OFFICES worldwide

153 MILLION substances

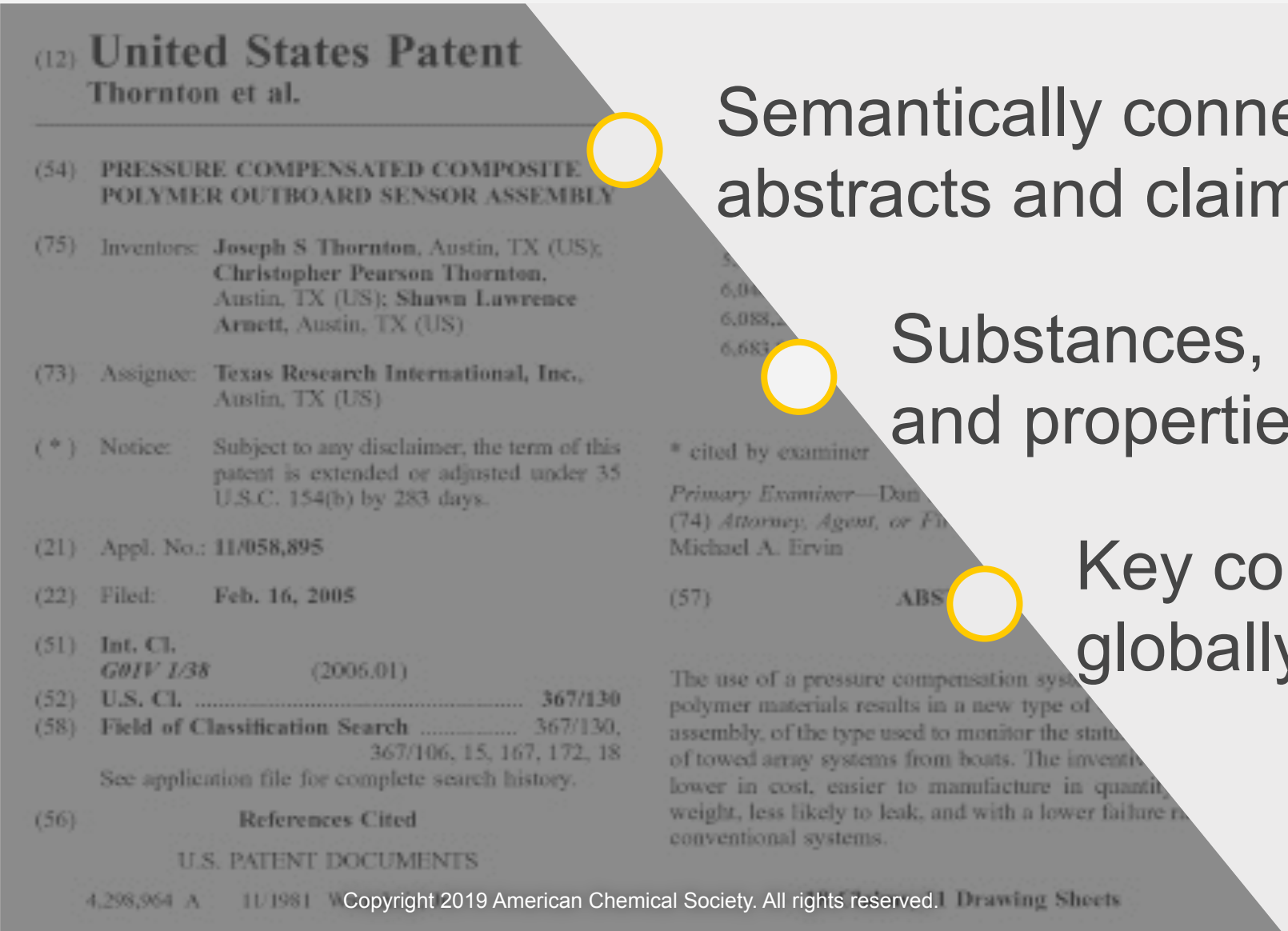
50 LANGUAGES translated



CAS A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

Copyright 2019 American Chemical Society. All rights reserved.

Robust, high-quality data require in-depth curation by scientists with expertise in the field and the language



(12) **United States Patent**
Thornton et al.

(54) **PRESSURE COMPENSATED COMPOSITE POLYMER OUTBOARD SENSOR ASSEMBLY**

(75) Inventors: **Joseph S Thornton**, Austin, TX (US); **Christopher Pearson Thornton**, Austin, TX (US); **Shawn Lawrence Arnett**, Austin, TX (US)

(73) Assignee: **Texas Research International, Inc.**, Austin, TX (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 283 days.

(21) Appl. No.: **11/058,895**

(22) Filed: **Feb. 16, 2005**

(51) **Int. Cl.**
G01V 1/38 (2006.01)

(52) **U.S. Cl.** **367/130**

(58) **Field of Classification Search** 367/130, 367/106, 15, 167, 172, 18

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,298,964 A 11/1981 W

* cited by examiner

Primary Examiner—Dan

(74) *Attorney, Agent, or Fi*
Michael A. Ervin

(57) ABS

The use of a pressure compensation sys
polymer materials results in a new type of
assembly, of the type used to monitor the statu
of towed array systems from boats. The inventi
lower in cost, easier to manufacture in quanti
weight, less likely to leak, and with a lower failure r
conventional systems.

1 Drawing Sheets

Semantically connected titles, abstracts and claims

Substances, reactions, sequences and properties connected across pubs

Key concepts and inventions globally translated and indexed

Copyright 2019 American Chemical Society. All rights reserved.

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

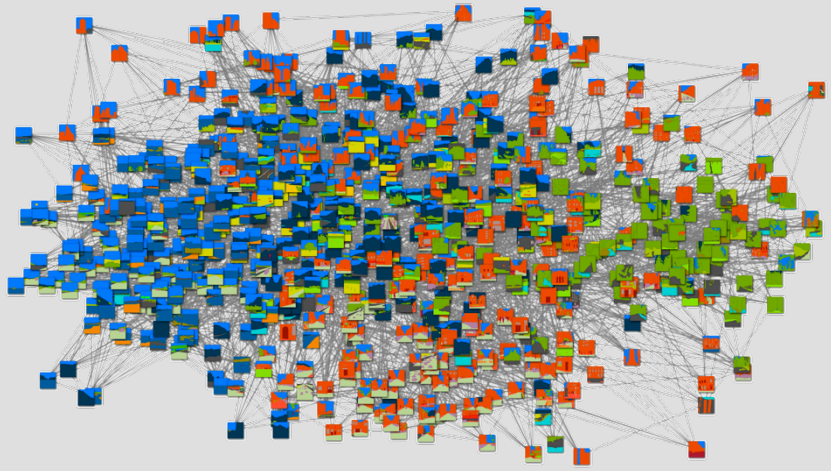






CAS conquers the chaos

- Global Translation
- Document Indexing
- Lexicon Development
- Semantic Indexing
- Reaction Indexing
- Markush Indexing



- Data Modeling
- Search Architecture
- App Development

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

Copyright 2019 American Chemical Society. All rights reserved.

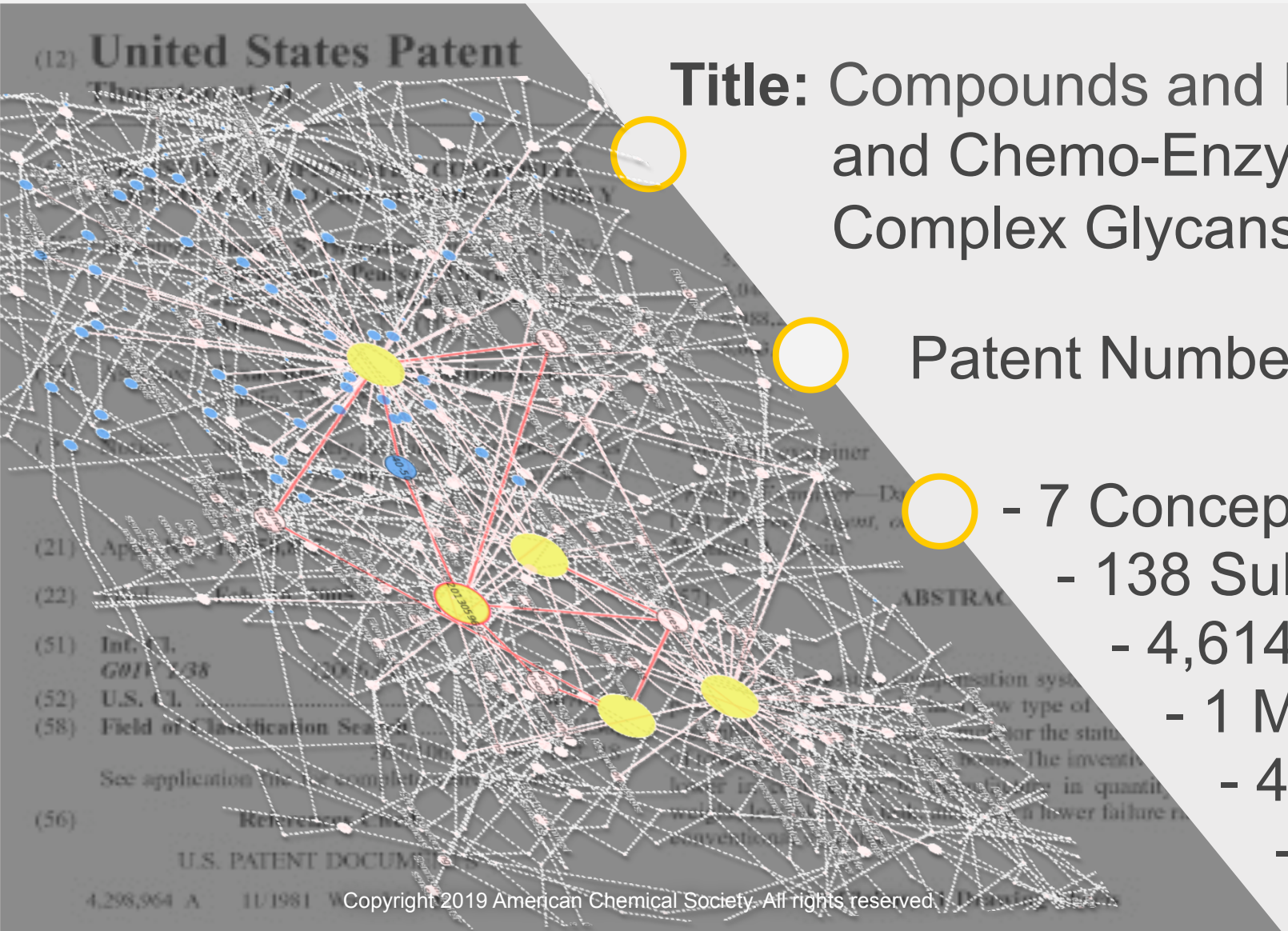
Algorithms miss details within scientific content that only intellectual curation delivers

	Finds all chemistry	Comprehensive retrieval	Scientific connections	50+ languages translated	Well-structured and connected	
Human Indexing	✓ (tables, text, images & Markush)	✓	✓	✓ (native speakers with scientific expertise)	✓	**HIGH CONFIDENCE**
Automated Indexing	✗ (Simple chemistry, not novel or complex)	✗ (Inconsistent)	✗ (Lacks context & connections)	✗ (limited in language and scope)	✗ (unstructured)	**LOW CONFIDENCE**

Copyright 2019 American Chemical Society. All rights reserved.

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

CAS captures critical details that algorithms alone miss



Title: Compounds and Methods for Chemical and Chemo-Enzymatic Synthesis of Complex Glycans

Patent Number: WO 2012135049

- 7 Concepts
  - 138 Substances
    - 4,614 Reactions
      - 1 Markush Structure
        - 4 Patent Family Members
          - 3 Cited Documents

Copyright 2019 American Chemical Society. All rights reserved.

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

More data beats clever algorithms, but BETTER DATA beats more data

Peter Norvig, PhD
Director of research at Alphabet, notable author & AI expert

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

Copyright 2019 American Chemical Society. All rights reserved.

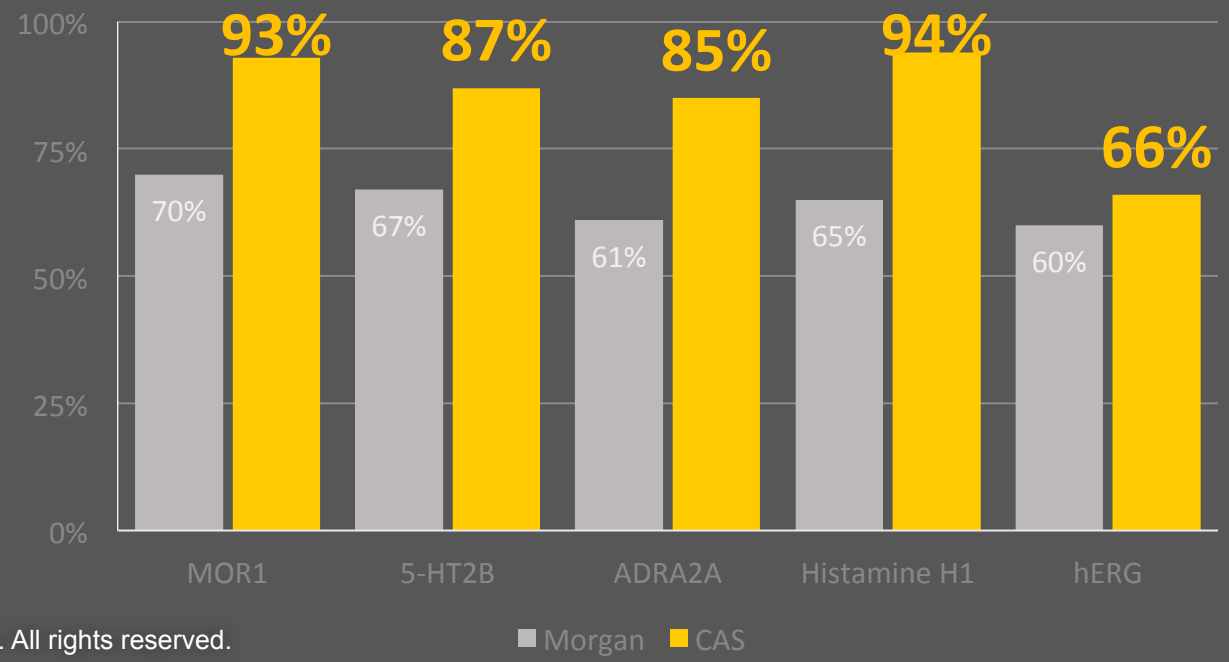
What is the measurable impact of clean, human-curated data on predictive outcomes?

CHALLENGE: A recently published paper classified almost 10,000 chemical entities on predicted biological activity to five different targets using Morgan fingerprints using a support vector machine model

OUR QUESTION: Does substituting Morgan fingerprints with CAS proprietary fingerprints have a measureable impact on prediction accuracy?

RESULTS:

The classification accuracy increased by over 30% when using higher-quality CAS data



CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

Copyright 2019 American Chemical Society. All rights reserved.

CAS uses knowledge graphs to leverage this data for unique insights

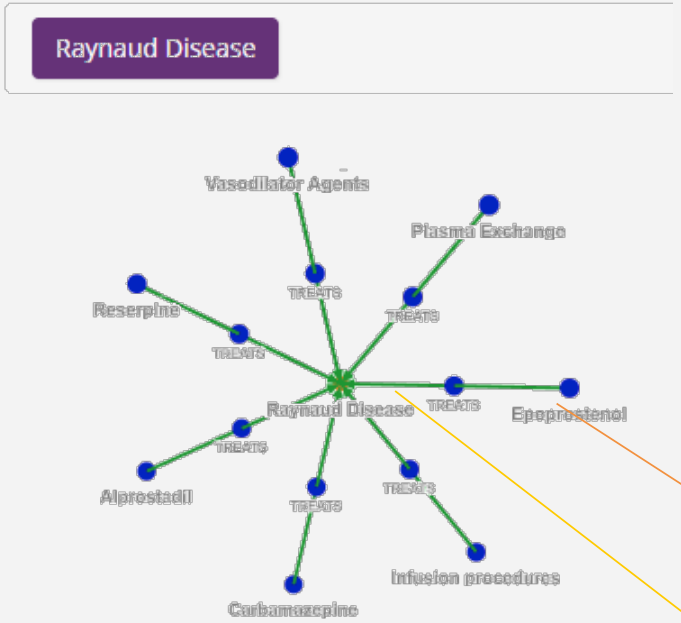
Outcomes:	**Literature Discovery**	**Prior Art Search**	**Decision Support**
			
Approach:	**Text-based knowledge graphs**	**AI + knowledge graphs**	**IAP + real-time analytics**

Copyright 2019 American Chemical Society. All rights reserved.

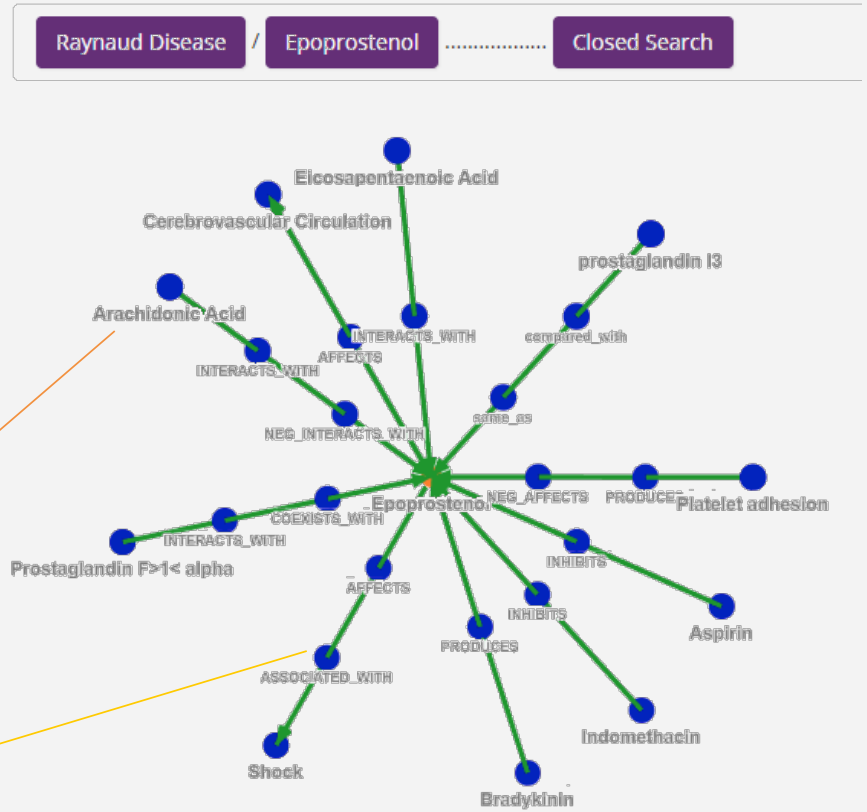
CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

Literature based discovery via knowledge graphs allows for open exploration

Begin with a known curated concept



Explore other related concepts extracted from content



Concept and Substance Details

Source Reference Exploration

Literature Discovery

Copyright 2019 American Chemical Society. All rights reserved.

CAS A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

and closed discovery

Begin by choosing several concepts

Source Concepts

Concept Name

Magnesium

Add Another Subject

AND

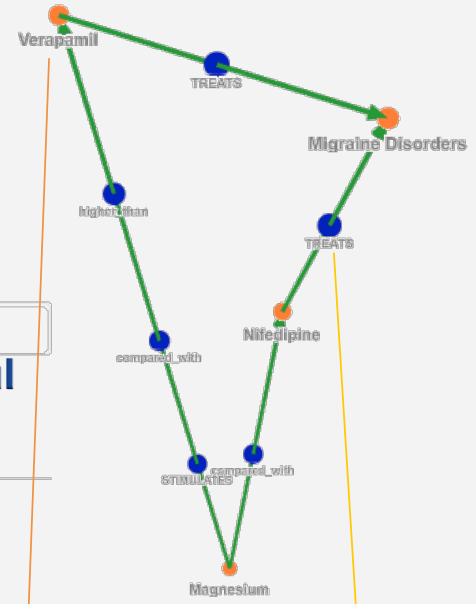
Destination Concepts

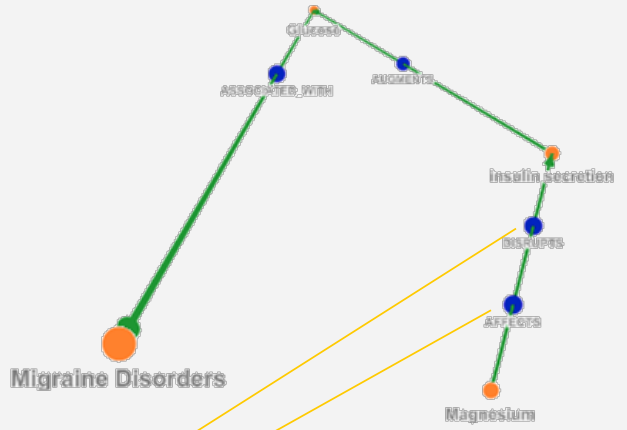
Concept Name

Migraine Disorders

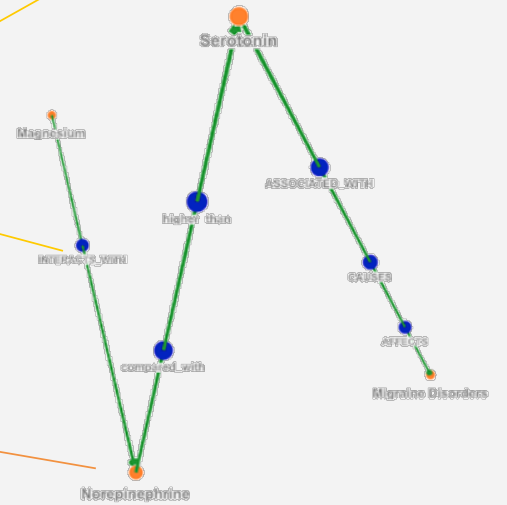
Add Another Object

Find meaningful connections





Source Reference Exploration



Concept and Substance Details

Literature Discovery

Copyright 2019 American Chemical Society. All rights reserved.

CAS A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

CAS uses knowledge graphs to leverage this data for unique insights

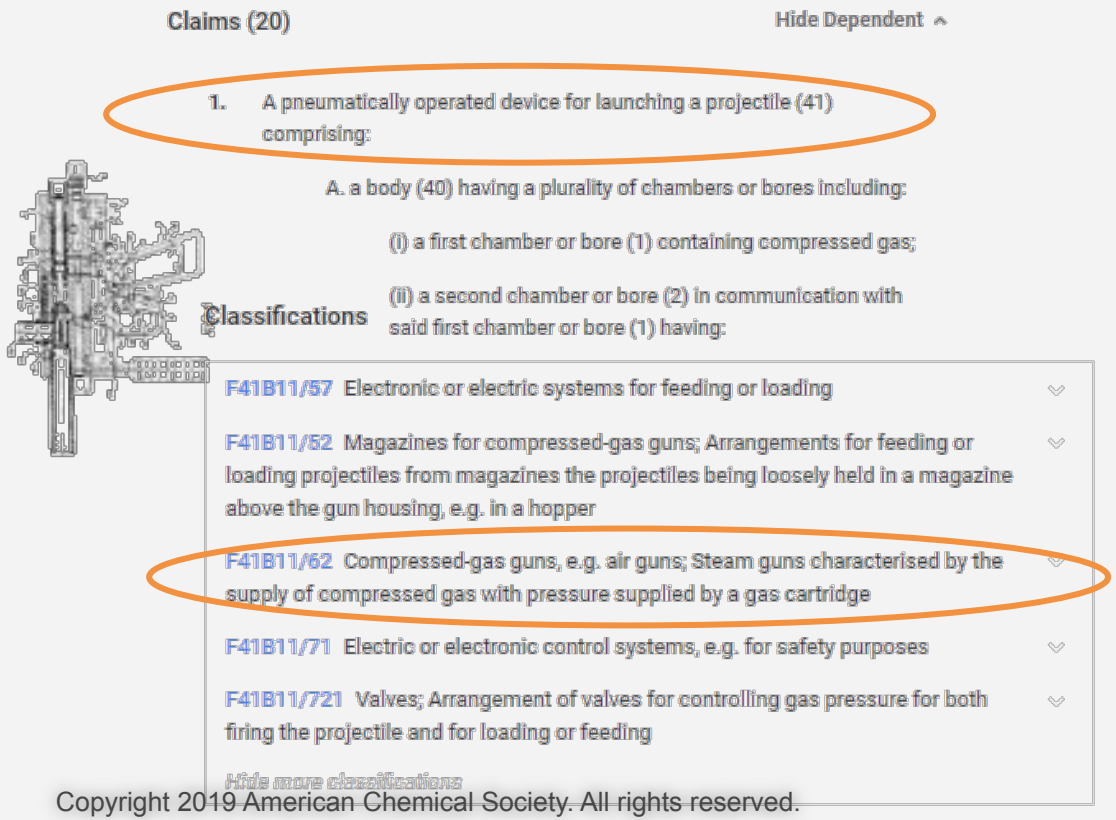
Outcomes:	Literature Discovery	**Prior Art Search**	Decision Support
			
Approach:	Text-based knowledge graphs	**AI + knowledge graphs**	IAP + real-time analytics

Copyright 2019 American Chemical Society. All rights reserved.

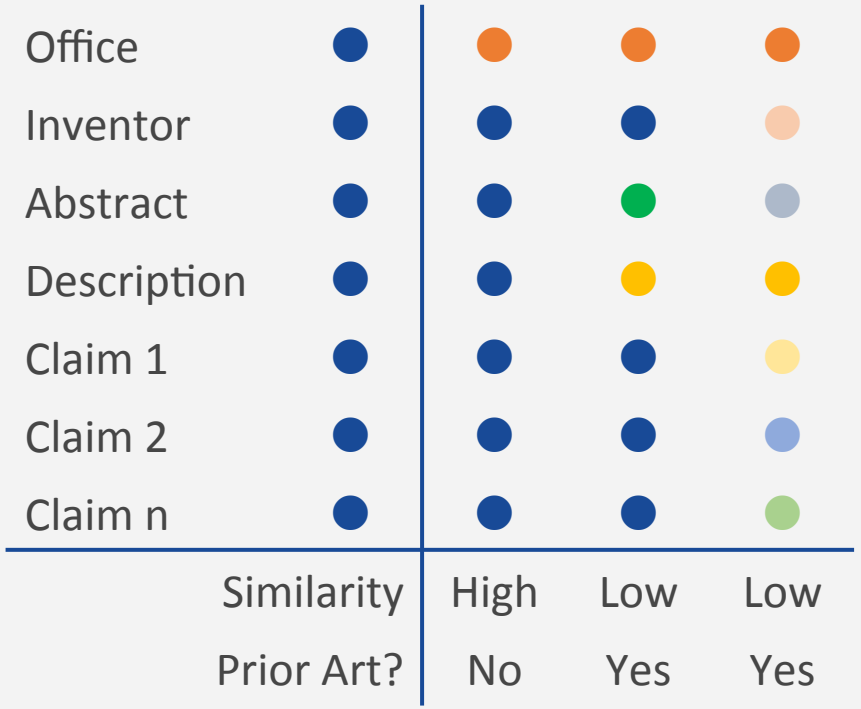
CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

Patents have inherent challenges compared to typical text ingestion and similarity searches

Syntactic Similarity vs Semantic Similarity



Similarity vs Prior-Art

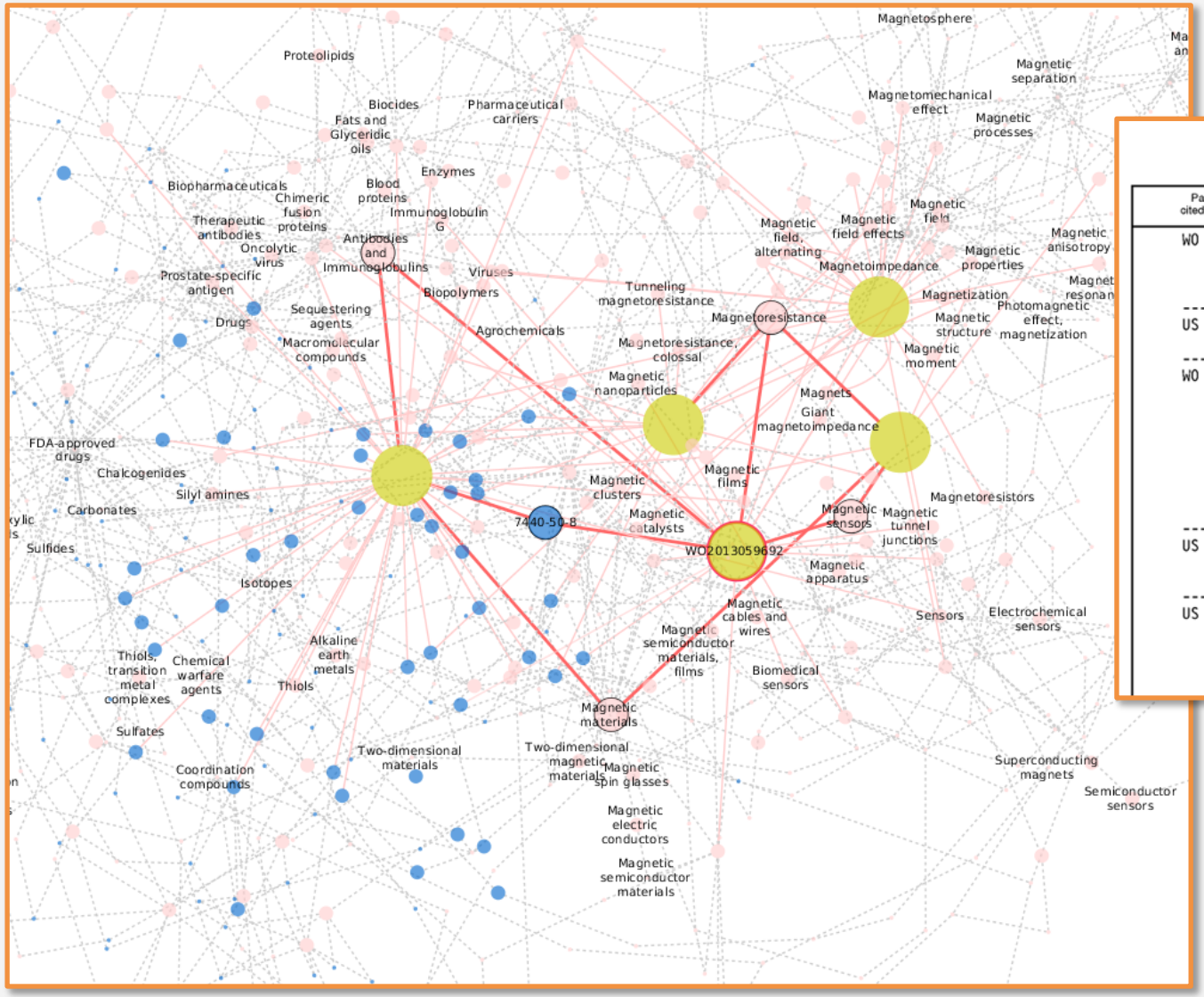


Prior Art Search

Copyright 2019 American Chemical Society. All rights reserved.

CAS A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

Visual representation of search reports support faster patent analysis and examination



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No PCT/US2012/061156

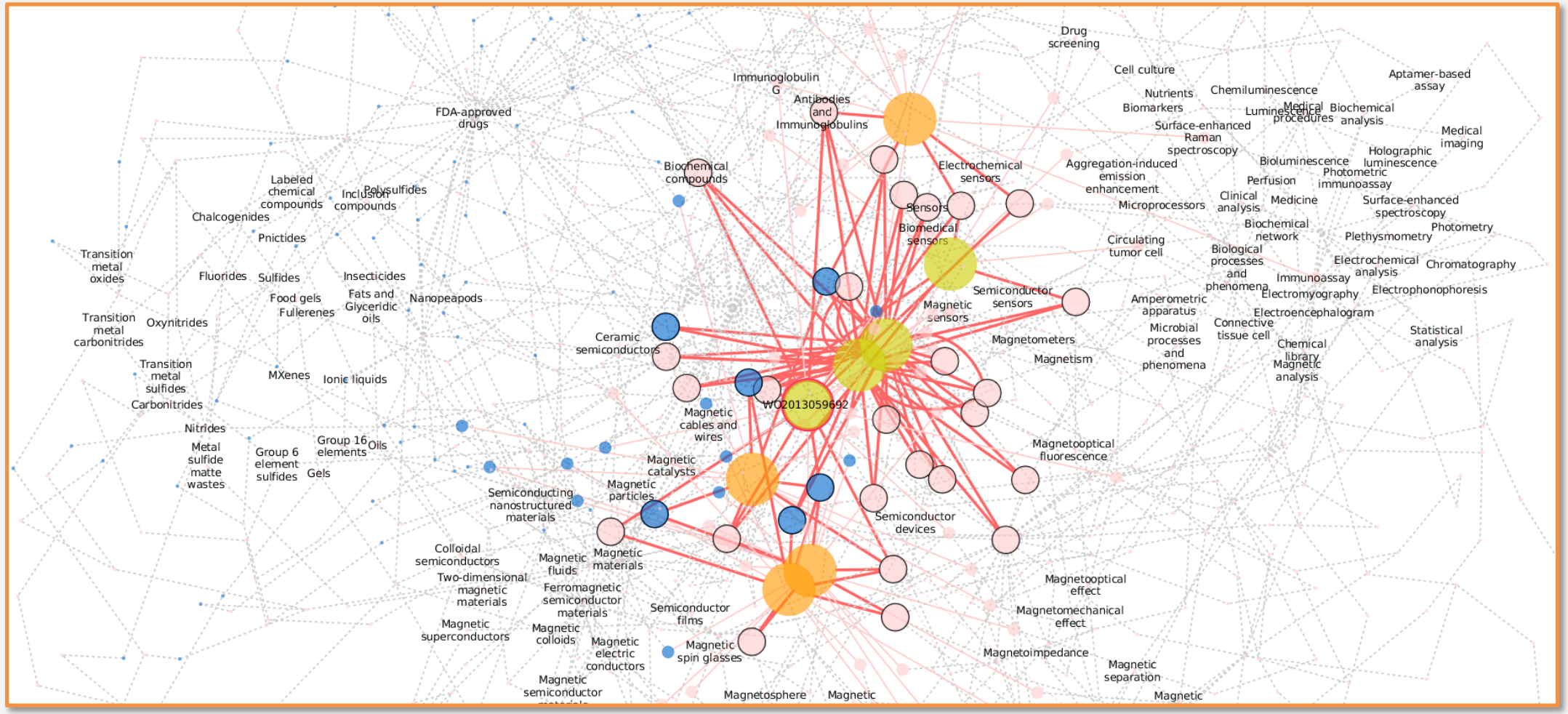
Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 2006080558	A1	03-08-2006	JP	4756868 B2	24-08-2011
			JP	2006208295 A	10-08-2006
			US	2008284419 A1	20-11-2008
			WO	2006080558 A1	03-08-2006
US 2005244873	A1	03-11-2005	JP	2005315678 A	10-11-2005
			US	2005244873 A1	03-11-2005
WO 03054523	A2	03-07-2003	AT	326697 T	15-06-2006
			AU	2002366904 A1	09-07-2003
			CN	1608206 A	20-04-2005
			DE	60211555 T2	22-02-2007
			EP	1456658 A2	15-09-2004
			JP	2005513475 A	12-05-2005
			US	2005087000 A1	28-04-2005
			US	2006128035 A1	15-06-2006
			WO	03054523 A2	03-07-2003
US 2009021250	A1	22-01-2009	CN	101315429 A	03-12-2008
			JP	2009008663 A	15-01-2009
			US	2009021250 A1	22-01-2009
US 2010117641	A1	13-05-2010	US	2010117641 A1	13-05-2010
			US	2010231214 A1	16-09-2010
			US	2010232075 A1	16-09-2010
			US	2011043201 A1	24-02-2011
			US	2011095751 A1	28-04-2011

Prior Art Search

Copyright 2019 American Chemical Society. All rights reserved.

CAS A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

Can extend beyond search reports to include other connected patents and non-patent literature



Prior Art Search

Copyright 2019 American Chemical Society. All rights reserved.

CAS A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

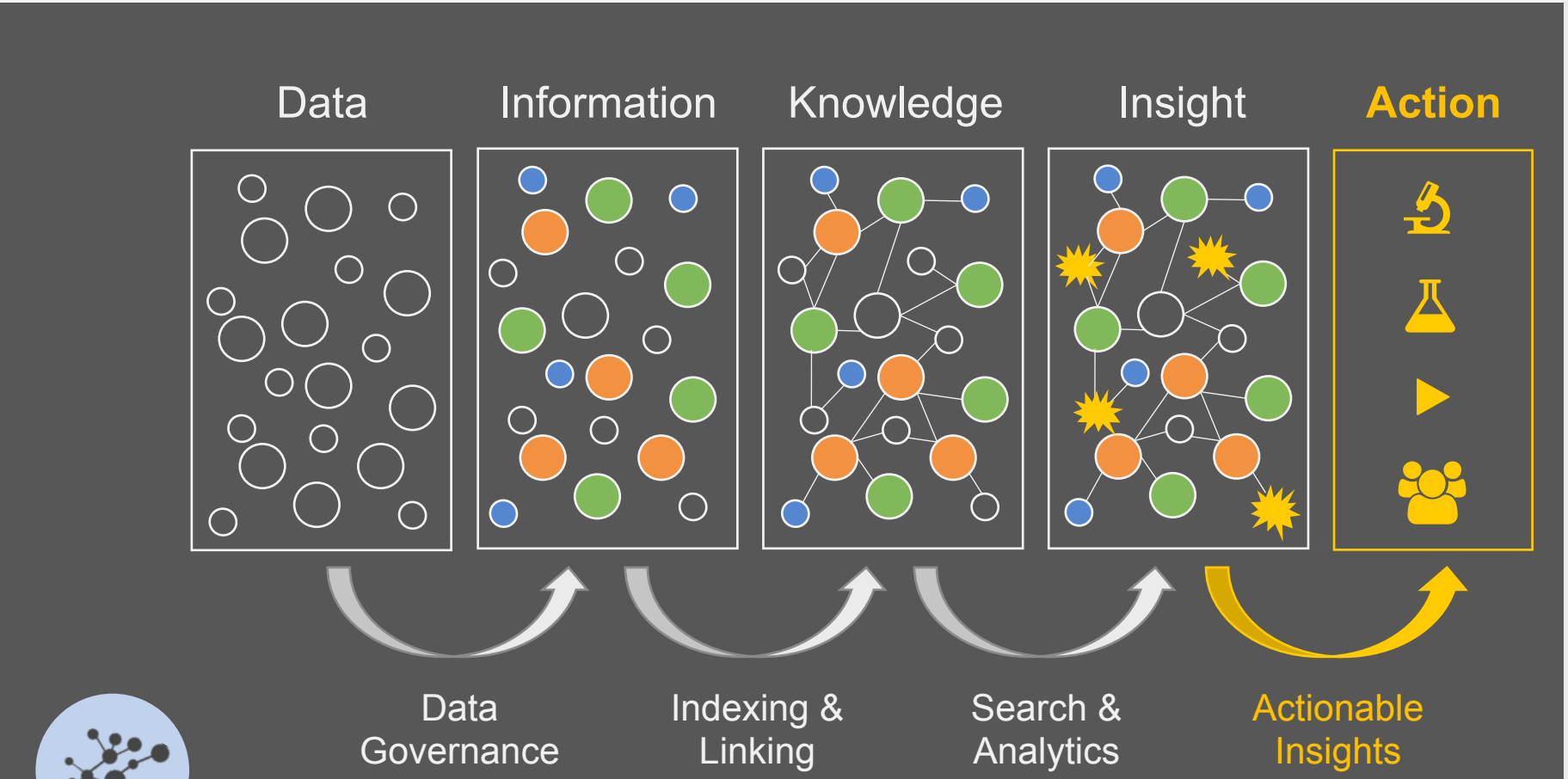
CAS uses knowledge graphs to leverage this data for unique insights

Outcomes:	Literature Discovery	Prior Art Search	Decision Support
			
Approach:	Text-based knowledge graphs	AI + knowledge graphs	IAP + real-time analytics

Copyright 2019 American Chemical Society. All rights reserved.

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

Building connections transforms data to support insights and key decisions



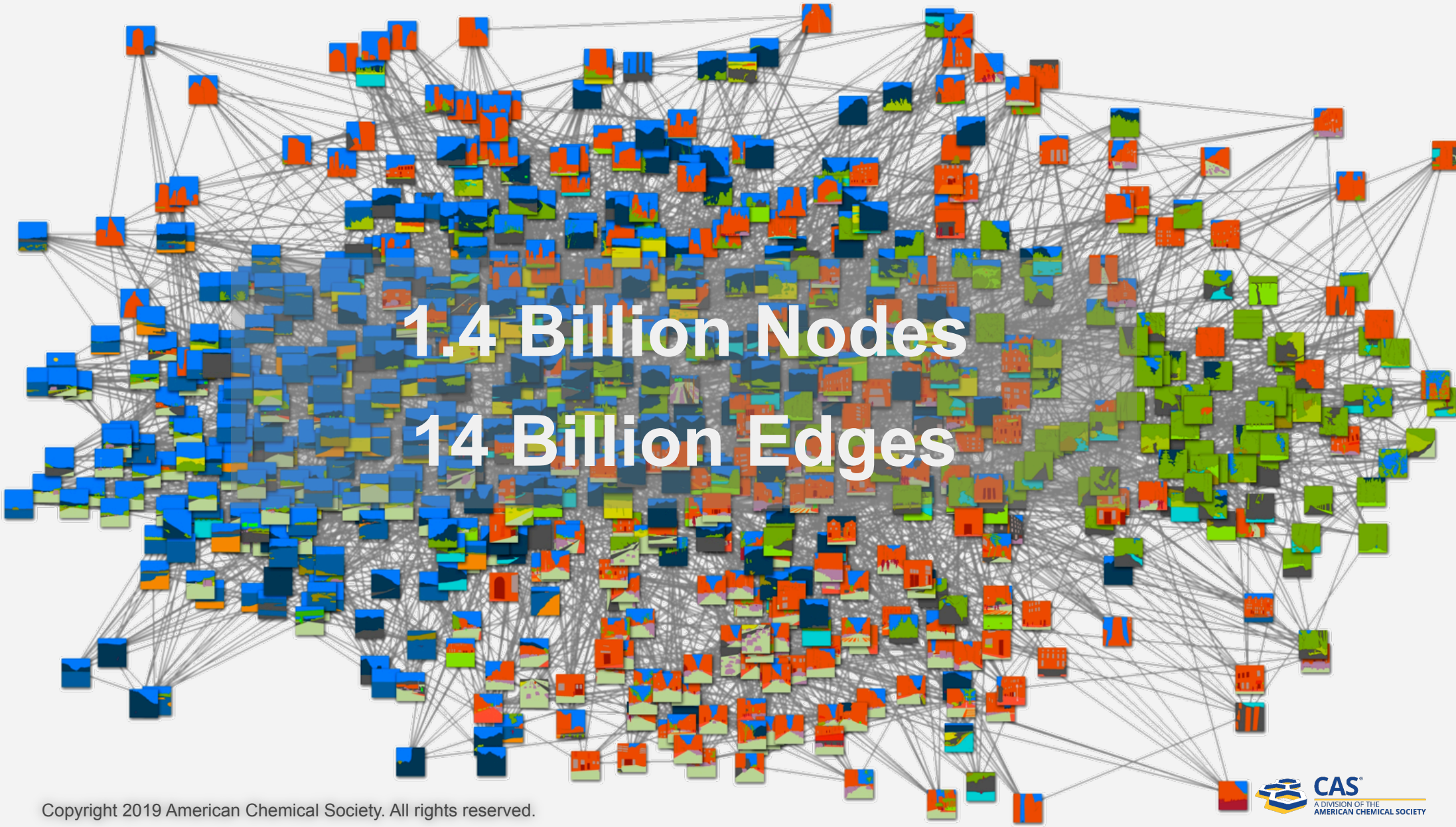
Actionable insights drive faster breakthroughs

Decision Support

Copyright 2019 American Chemical Society. All rights reserved.

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

The CAS knowledge graph is enormous

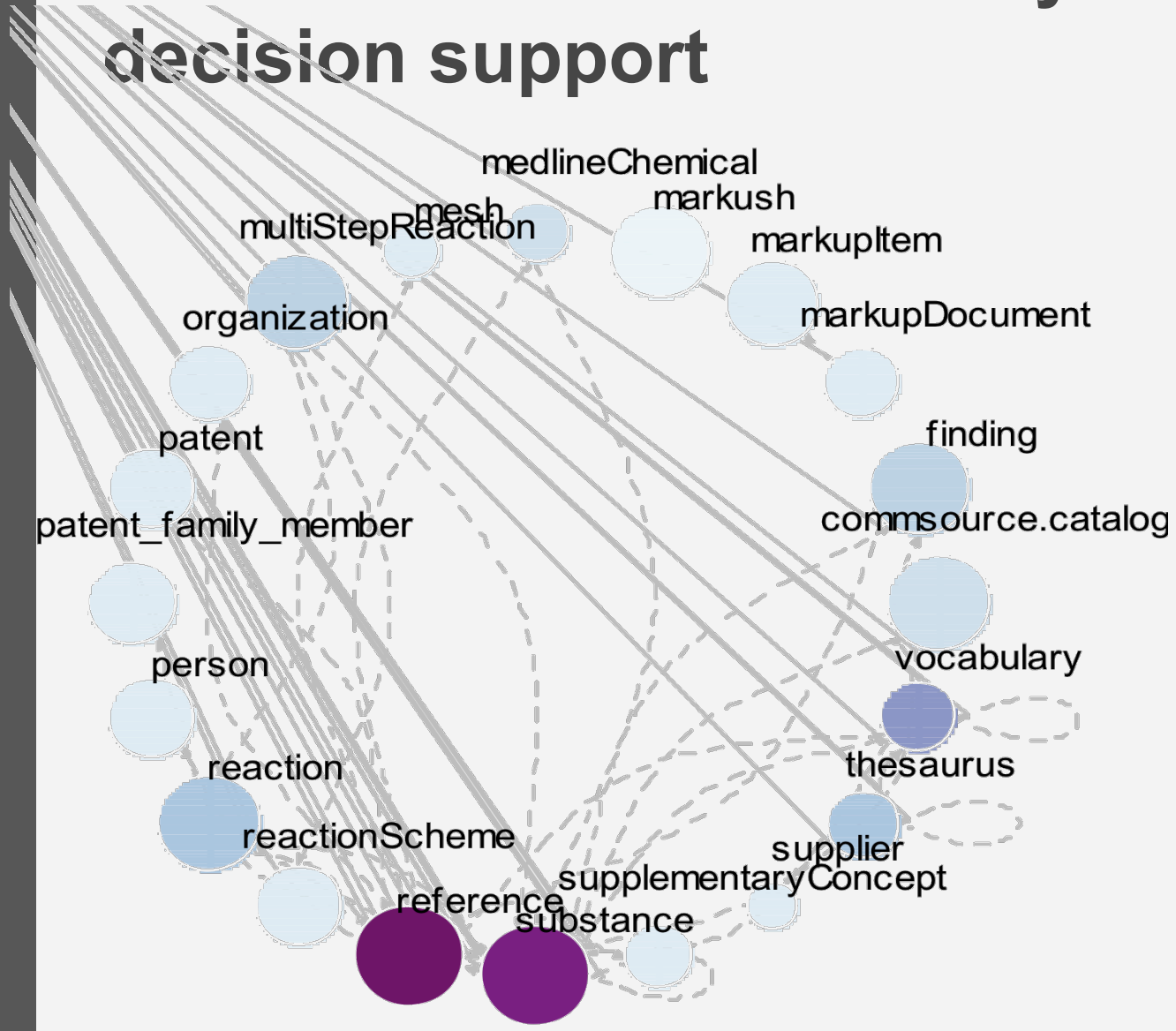


Decision Support

Copyright 2019 American Chemical Society. All rights reserved.

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

IAP enables real-time analytics for more insightful decision support



Nearly 10TB of data and >25 entities for SciFinder-n

Distributed, scalable and cloud-enabled

Integrated text search, structure search and graph traversal in near real time

Decision Support

Copyright 2019 American Chemical Society. All rights reserved.

CAS A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

To Recap

The potential of emerging data technologies to accelerate scientific innovation is
HIGH

The impact of data quality on the success of these technologies is
HUGE

The ROI on human-curated data for scientific content is
MEASURABLE

The potential applications to drive insight are
UNLIMITED

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

Copyright 2019 American Chemical Society. All rights reserved.

CAS®
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

www.cas.org

Mark R. Grabau
mgrabau@cas.org

Copyright 2019 American Chemical Society. All rights reserved.