

# Programme

- *Data Science & Artificial Intelligence*

## **Data Science and AI at the Scholarly Communications Frontier**

*Understanding content, workflow and impact, to inform publishing and editorial decision-making*

### **John Sack**

- *Founding Director, HighWire Press*



*New applications of AI/machine learning and predictive analytics are beginning to influence scholarly communication. Data mining across platforms can deliver insight into the lifecycle of individual papers from submission to citations and downloads. This information can help researchers, editors and publishers make decisions and adjust strategy. In this presentation, we will discuss examples that can help editors and publishers use analytics and AI to tune products and uncover emerging opportunities.*



**Researcher  
to Reader**

26 & 27  
February  
2018





Data Science and  
Artificial Intelligence at the  
Scholarly Communications Frontier

**John Sack**  
**Founding Director**  
**HighWire Press**

27 February 2018



# Topics: AI/DS. Authors/Editors

- AI Applications Overview
- AI and Data Science in Publishing

	Readers / Authors	Editors / Publishers
Artificial Intelligence	1	2
Data Science	3	4

- Some Cautions

# Topics: AI/DS. Authors/Editors

- **AI Applications Overview**
- **AI and Data Science in Publishing**

	Readers / Authors	Editors / Publishers
Artificial Intelligence	1	2
Data Science	3	4

- **Some Cautions**

# AI writes Financial Reports

WEB REPORT ENTERTAINMENT

34

## AP's 'robot journalists' are writing their own stories now

By [Ross Miller](#) | Jan 29, 2015, 11:55am EST



SHARE



TWEET



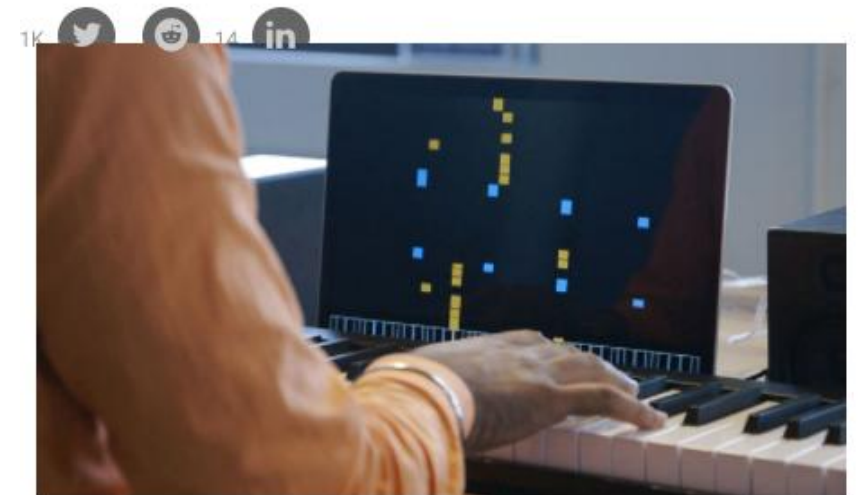
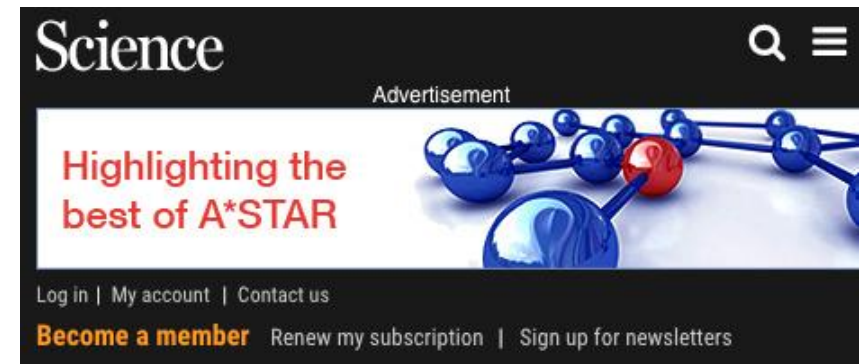
LINKEDIN



Typing robot photo by Shutterstock

Minutes after Apple released its [record-breaking quarterly earnings](#) this week, the Associated Press published (by way of [CNBC](#), [Yahoo](#), and others) "Apple tops Street 1Q forecasts." It's a story without a byline, or rather, without a *human* byline — a financial story

# AI 'Creativity': Making Music



A musician improvises alongside A.I. Duet, software developed in part by Google's Magenta

google

## How Google is making music with artificial intelligence

By [Matthew Hutson](#) | Aug. 8, 2017 , 3:40 PM

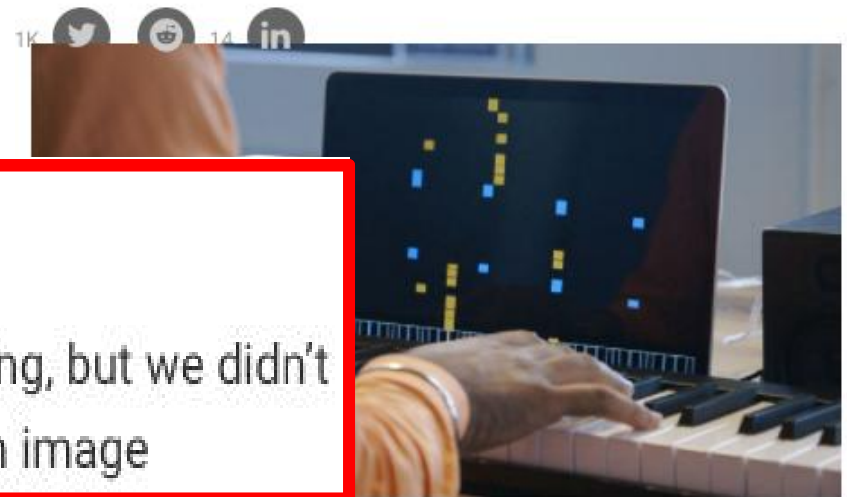
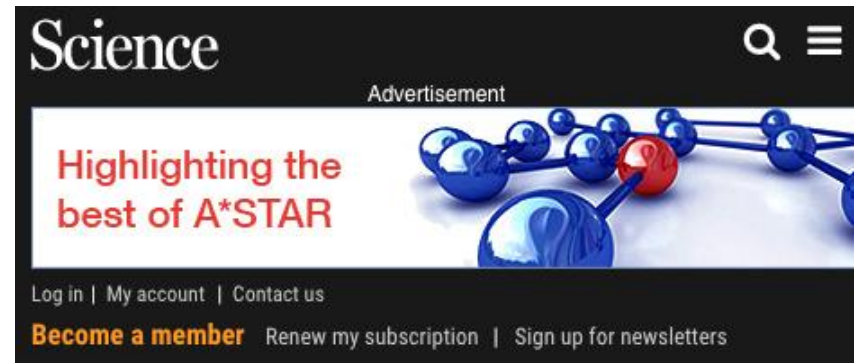
Can computers be creative? That's a question **bordering on the philosophical**, but artificial intelligence (AI) can certainly make music and artwork that people find pleasing. Last year, Google launched **Magenta**, a research project aimed at pushing the limits of what AI can do in the arts. *Science* spoke with Douglas



# AI 'Creativity': Making Music ...but AI Creativity has its limits...

**Q: What else is Magenta doing?**

**A:** We did a summer internship around joke telling, but we didn't generate any funny jokes. We're also working on image



A musician improvises alongside A.I. Duet, software developed in part by Google's Magenta

google

## How Google is making music with artificial intelligence

By **Matthew Hutson** | Aug. 8, 2017, 3:40 PM

Can computers be creative? That's a question **bordering on the philosophical**, but artificial intelligence (AI) can certainly make music and artwork that people find pleasing. Last year, Google launched **Magenta**, a research project aimed at pushing the limits of what AI can do in the arts. *Science* spoke with Douglas

# AI in the Lab

## Research information

The essential  
li



News

Analysis & Opinion

Features

Interviews

Events



For news updates direct  
from the editorial team  
Click here to follow us

Research Information

### NEWS

Tags: **INDUSTRY TRENDS**

## AI 'has potential to revolutionise life sciences' – Pistoia Alliance

9 January 2018

Tweet

9

Share

Some 44 per cent of life science professionals are using or experimenting with AI and deep learning, while 94 per cent expect an increase in use of machine learning within two years.

These are findings from a survey carried out by the Pistoia Alliance, a global, not for profit alliance that works to lower barriers to innovation in life sciences R&D.

The organisation surveyed 374 life science professionals on AI, machine learning



HighWire



# AI-based Portfolio Management

- Could an AI "edit" a journal as it can manage a stock portfolio?
- Discuss.

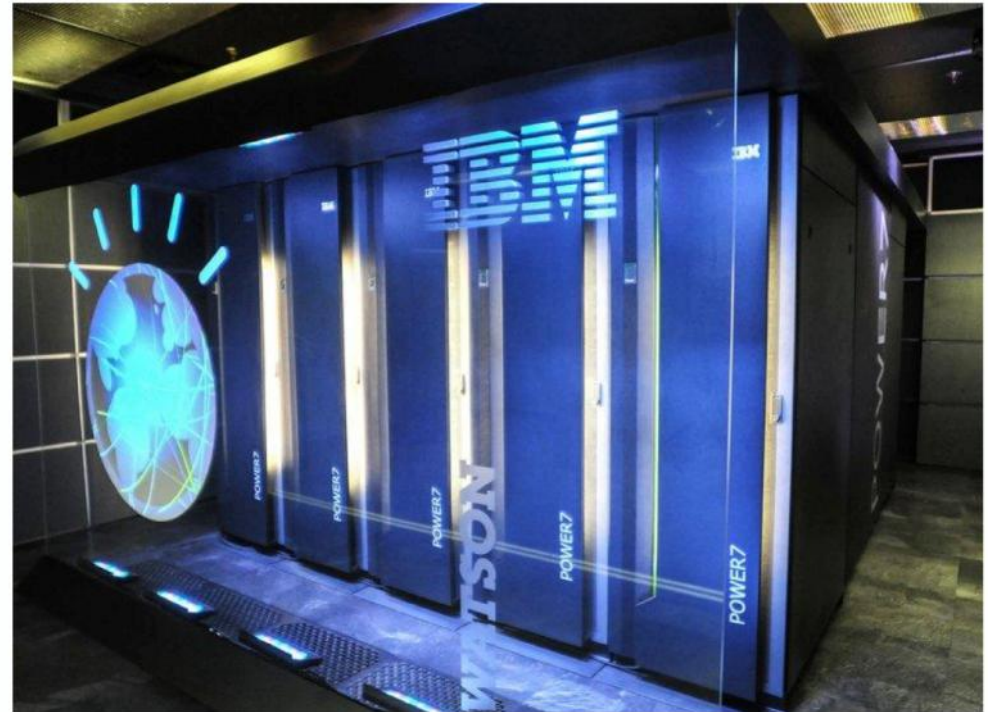
## A.I. Has Arrived in Investing. Humans Are Still Dominating.

By Conrad De Aenlle, [www.nytimes.com](http://www.nytimes.com)

January 14th, 2018

[View Original](#)

ai



I.B.M.'s Watson supercomputer in 2011. An investment fund that emphasizes artificial intelligence now uses Watson to power many of its algorithms. Credit Associated Press  
*Photo by: Associated Press*

Machines are starting to take the place of the people who flip burgers, drive across town and, lately, manage stock portfolios.

Artificial intelligence is taking on a bigger role in making investment decisions.



# Topics: AI/DS. Authors/Editors

- AI Applications Overview
- **AI and Data Science in Publishing**



- Some Cautions

# AI for Researchers: Readers/Authors

- “Keeping Up”
  - Which new articles match my fields?
  - Which articles are about my specific interests?
  - Which articles contain something related to my specific interests?
- “Writing Up”
  - Draft my newest article for me
  - Where shall I submit my newest article first?
- “Dreaming Up”
  - How might Gene A and Protein C be related?

# Keeping Up With My Field

## Arxiv Sanity Preserver

Built in spare time by @karpathy to accelerate research.  
Serving last 41922 papers from cs.[CV|CL|LG|AI|NE]/stat.ML

User:  Pass:  [Login or Create](#)

Fork me on GitHub



[most recent](#) [top recent](#) [top hype](#) [friends](#) [discussions](#) [recommended](#) [library](#)

[Last day](#) [Last week](#) [Last month](#)

Top papers mentioned on Twitter over last day:

### The Shape of Art History in the Eyes of the Machine

Ahmed Elgammal, Marian Mazzone, Bingchen Liu, Diana Kim, Mohamed Elhoseiny

2/12/2018 (v1: 1/23/2018) [cs.AI](#) | [cs.CV](#)

1801.07729v2 [pdf](#)

[show similar](#) | [discuss](#)



How does the machine classify styles in art? And how does it relate to art historians' methods for analyzing style? Several studies have shown the ability of the machine to learn and predict style categories, such as Renaissance, Baroque, Impressionism, etc., from images of paintings. This implies that the machine can learn an internal representation encoding discriminative features through its visual analysis. However, such a representation is not necessarily interpretable. We conducted a comprehensive study of several of the state-of-the-art convolutional neural networks applied to the task of style classification on 77K images of paintings, and analyzed the learned representation through correlation analysis with concepts derived from art history.

titanium dioxide x



## Finding a Needle in a Haystack (UNSILO Example)

### Publisher

<input type="checkbox"/> Elsevier	9,341
<input type="checkbox"/> RSC Publishing	7,284
<input type="checkbox"/> Springer	6,147
<input type="checkbox"/> ACS Publications	4,727
<input type="checkbox"/> Wiley	1,380

✓ See all (26)

### Journal

<input type="checkbox"/> ACS Applied Materials & Interfaces	2,409
<input type="checkbox"/> RSC Advances	2,135
<input type="checkbox"/> Applied Surface Science	1,926
<input type="checkbox"/> Electrochimica Acta	1,065
<input type="checkbox"/> Journal of Alloys and Compounds	987

✓ See the top 100

33,839 articles

8,004 nanomaterials

202 patents

Sort by

Relevance v

☐ Curated summaries for nanomaterials

☐ Select all to export

#### ☐ Progress on mesoporous titanium dioxide: Synthesis, modification and applications

Samira Bagheri | Zul Adlan Mohd Hir | Amin Termeh Yousefi ... in **Microporous and Mesoporous Materials** (2015)

Mesoporous materials have a remarkable interest due to their exceptional characteristics and favorable applications in various fields of technology. Mesoporous Titanium Dioxide is one of the most broadly [more](#)

**This article discusses:** Titania with Mesoporous Titania, Surfactant, Surface, Titania Nanoparticles, Pore Citations according to ReadCube: 29

#### ☐ Biomembrane-Compatible Sol-Gel-Derived Photocatalytic Titanium Dioxide

Kaitlin E. Johnson | Sukriti Gakhar | Yue Deng ... in **ACS Applied Materials & Interfaces** (2017)

**This article discusses:** Titania with Titania Gel, BR, Ethanol, Photocatalytic, Methylene Blue





# AI Writes Research Articles

## Research information

The essential  
li



News

Analysis & Opinion

Features

Interviews

Events



For news updates direct  
from the editorial team  
Click here to follow us

Research Information

### NEWS

Tags: [INDUSTRY TRENDS](#)

## Artificially creating articles

30 November 2017

Tweet

22

Share

Some 99 drafts of scientific papers have been generated so far by a manuscript writer launched three weeks ago, according to the electronic lab notebook company sciNote.

The Manuscript Writer is described as the first one of its kind, allowing researchers to generate a draft of a scientific manuscript using data stored by the user on its software and references that are accessible in open access journals.

The device was developed by a team of scientists and experts on machine learning and software development of sciNote (creators of sciNote, a free open source electronic lab notebook) and it has the potential to notably simplify the process of preparing scientific manuscripts by using the technological advances in artificial intelligence.

Manuscript Writer is aimed at reducing the time needed to prepare initial content. It gathers the data scientists organise and saved in sciNote during their research, and presents it in the form of a manuscript draft. This allows the scientists to save time on gathering their relevant data so they don't need to start writing from scratch. Once they receive the draft, they can start editing and improving the text.



# AI Writes Research Articles



News

Analysis & Opinion

Features

Interviews

Events

### SCINOTE CAN WRITE A DRAFT OF YOUR SCIENTIFIC MANUSCRIPT USING ARTIFICIAL INTELLIGENCE

Do you ever get that feeling that you would like to have a magic spell to organize all your data?

And once it is organized, wouldn't it be magnificent if there would be a software that could put together all relevant data from your projects, add some new references and present you with a manuscript draft you can build upon?

For news updates direct  
from the editorial team  
[Click here to follow us](#)

Research Information

RY TRENDS

### ally creating articles

17

[Tweet](#)

22

[in Share](#)

fts of scientific papers have been generated so far by a manuscript ed three weeks ago, according to the electronic lab notebook Note.

ript Writer is described as the first one of its kind, allowing to generate a draft of a scientific manuscript using data stored by the software and references that are accessible in open access journals.

was developed by a team of scientists and experts on machine software development of sciNote (creators of sciNote, a free open onic lab notebook) and it has the potential to notably simplify the preparing scientific manuscripts by using the technological advances elligence.

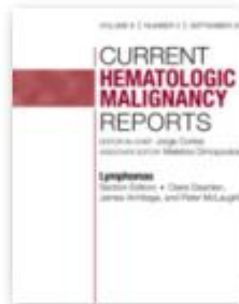
Writer is aimed at reducing the time needed to prepare initial content. s data scientists organise and saved in sciNote during their research, and presents it in the form of a manuscript draft. This allows the scientists to save time on gathering their relevant data so they don't need to start writing from scratch. Once they receive the draft, they can start editing and improving the text.



# Which Journal Should I Submit My Paper To?

## Journal Cascading

Journal Cascading is a ranking of the best journal matches for a submitted manuscript. This ranking is computed by extracting 61 unique features from the manuscript and comparing them to over 15 thousand previously-published manuscripts across thousands of journals.



Based on analysis, **Current Hematologic Malignancy Reports** is the best journal match for this manuscript. This journal publishes articles consistent with the projected article-level impact and topical fingerprint of this manuscript.

Additional suggestions are listed below:

**Blood Cancer Journal**  
**Molecular Cancer**  
**Current Oncology Reports**  
**Nature Communications**  
**Investigational New Drugs**  
**Leukemia**  
**Current Treatment Options in Oncology**  
**Molecular Psychiatry**  
**Stem Cell Research & Therapy**





“Dreaming Up”:  
What  
assertions  
does an article  
make?

Constructing  
“Truth Tables”

october • 27 • 2007

## Automatic Generation of One Million Structured Digital Abstracts

4:15pm  
talk  
**10**

**Balaji S.  
Srinivasan**

*Rion Snow  
Kim Branson  
Vijay S. Pande  
Andrew Ng  
Russ Altman  
Serafim  
Batzoglou*

### **Purpose:**

The Structured Digital Abstract Consortium is a group of researchers and publishers engaged in producing machine-readable summaries of the biomedical literature. These summaries, or Structured Digital Abstracts (SDAs), are to a machine what a text abstract is to a human: a concise description of the factual content of the article.

### **Materials and Methods:**

Biomedical classification systems like the Gene Ontology (GO) have proven invaluable for converting disordered collections of free text into machine-readable knowledge representations. However, the scalability of these ontologies is currently limited because they are populated manually from the literature at great expense. Here, we present an algorithm which removes this limitation by automatically extracting ontological relationships from a massive corpus of more than 1 million full text biomedical articles. We define the set of ontological relationships asserted in each article to be its “Structured Digital Abstract” or SDA.

In more detail, our algorithm uses a small training set of biological objects with known relationships such as ‘is\_a’ or ‘regulates\_a’ to find the lexico-syntactic patterns which specify these relationships in plain text. These learned patterns are then used to find many more examples of objects which satisfy these relationships. Importantly, our method (1) requires no manual specification of regular expressions, (2) is able to move beyond simple co-occurrences to learn directional relationships (e.g. “X localized\_to Y”), and (3) returns a probabilistic estimate for the truth of each ontological relationship.

### **Results:**

We demonstrate that our algorithm is capable of extracting dozens of different kinds of ontological relationships from free text, ranging from gene localizations to chemical modifications to anatomical structures. More-

### **Conclusion:**

By reducing each paper to the ontological relationships asserted therein, we have shown that it is possible to automatically convert unstructured free text into tables of machine-readable facts, or Structured Digital Abstracts. Moreover, by creating SDAs for more than 1 million articles representing roughly 7% of PubMed, we

“Dreaming  
Up”:  
What  
assertions  
does an  
article make?

SourceData  
at  
EMBO

SourceData is a novel platform for researchers and publishers to make their papers discoverable based on their data content.

New!  
Test SmartFigures

Register to test  
the curation tool



WINNER  
Sponsored by MPS



Try a DataSearch with SourceData

**Perturbation ?** **Measured entity ?**

Does  influence  ?

Try these searches: does insulin influence glucose? or does glucose influence insulin?

Our partners



WILEY



nature publishing group npg



# Topics: AI/DS. Authors/Editors

- AI Applications Overview
- **AI and Data Science in Publishing**



- Some Cautions

# HighWire Editors' Workshop on AI: What Editors Asked Us For

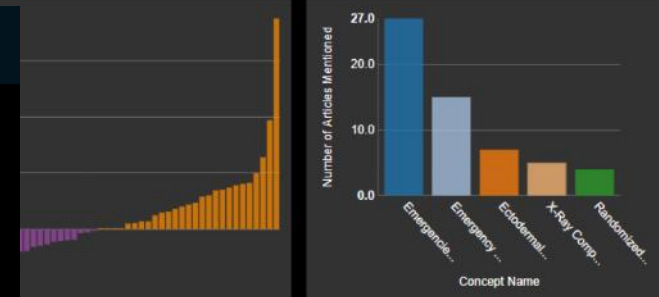
- “Make long meetings shorter”



# HighWire Editors' Workshop on AI: What Editors Asked Us For

- **Journal Performance**
  - Visualize the performance of a journal, its context and competitors
- **Community Engagement**
  - Tools to attract and retain the best authors, and the best papers, in advance of a paper's submission
- **Emergence of the New: fields, topics, authors, labs**
  - Tools to see over the horizon
- **Workflow Productivity/Efficiency**
  - Deal with repetitive, expertise-burning tasks

# Meta's Publisher Dashboard



Paper Title	Date Published	Article Type
Insertion and Use in the Emergency	10-15-2017	Journal Article
Advanced Cardiac Life Support for Out-	10-15-2017	Journal Article
with Severe Burns: A Meta-Analysis of	10-10-2017	Journal Article
Review	10-08-2017	Journal Article
after Emergency Department	10-08-2017	Journal Article
ture Research?	10-08-2017	Journal Article
assess impact of implicit bias: ematic review	10-04-2017	Journal Article

Journal Title ▲	# Papers (3 mo)	Paper Volume Change	Change Avg EF %
Academic Emergency Medicine	57	▲	27.33 ●
Acta Anaesthesiologica Scandinavica	51	▲	31.70 ●
Acta Biochimica et Biophysica Sinica	19	▼	60.83 ●
Acta Crystallographica Section A: Foundations and Advances	15	▲	43.03 ●
Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials	25	▼	33.92 ●

# Meta's “Bibliometric Intelligence” to evaluate submissions

- Is my journal an appropriate venue for this manuscript?
- What is the potential impact of this paper?
- To whom should I send this paper for review?

## Predictive Analytics

The major research areas identified in this manuscript are:



Chronic Lymphocytic  
Leukaemia Refractory



Alemtuzumab

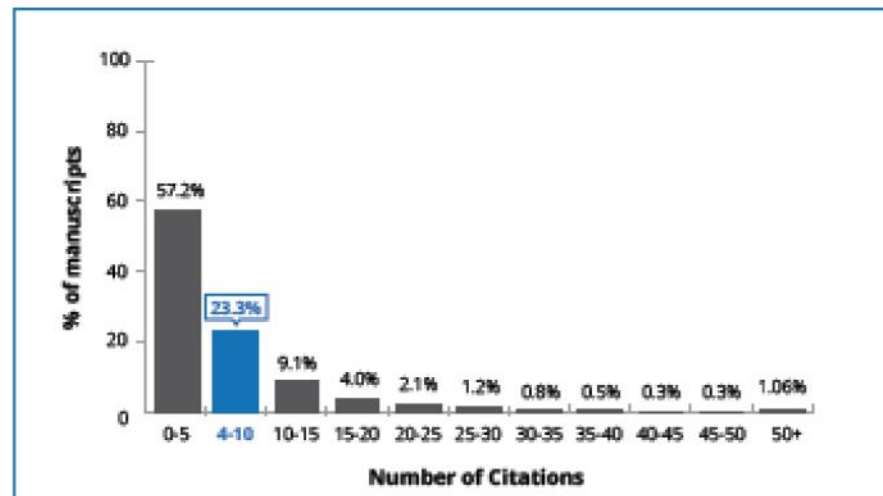


Therapy-related  
Myelodysplastic Syndrome



Fludarabine

## Predicted 3-Year Citation Count



Cheng-Hwai Tzeng

Clinicopathologic features and outcome of acute erythroid leukemia based on 2008 revised World Health Organization classification  
*Leukemia & Lymphoma*



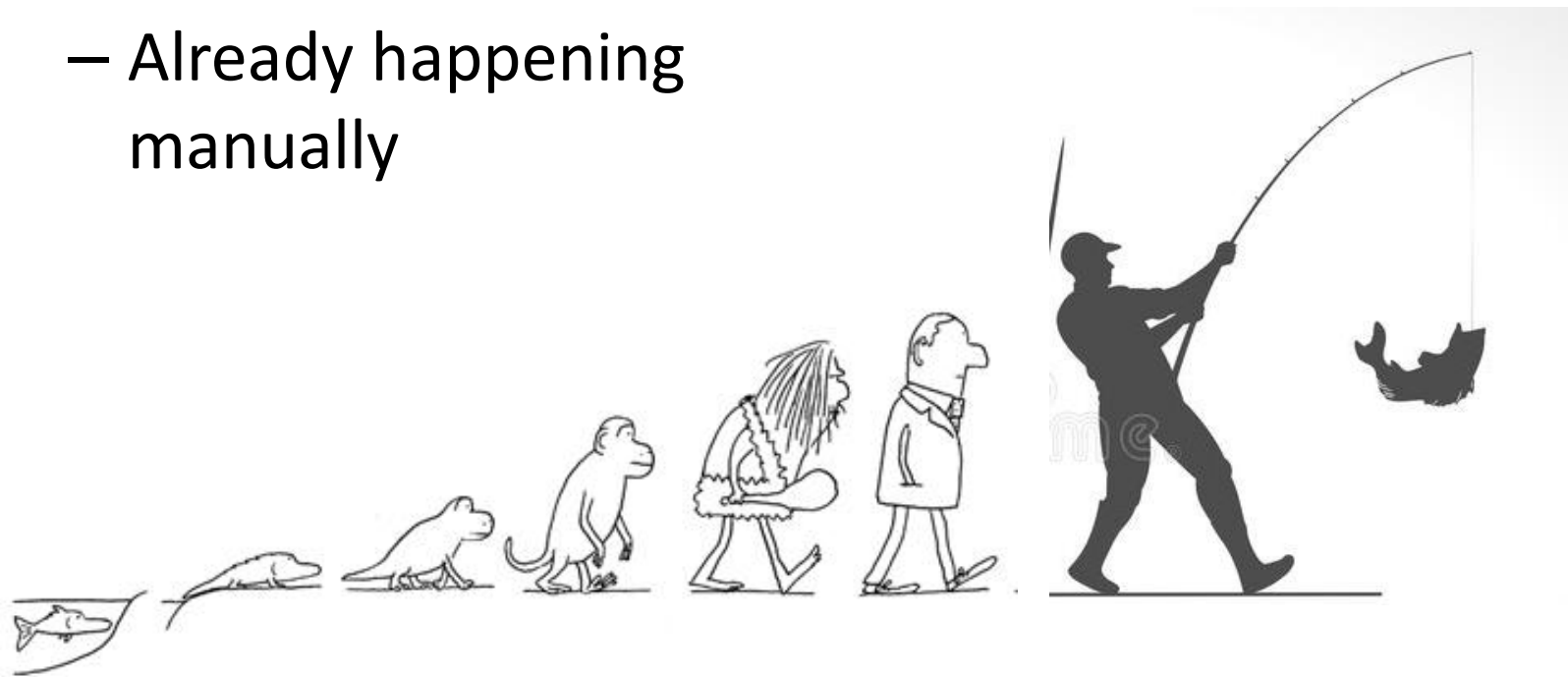
Gail J. Robez

Targeted Deletion of Autophagy Genes Atg5 Or Atg7 In The Chondrocytes Promotes Caspase-Dependent Cell Death And Leads To Mild Growth Retardation  
*Journal of Bone and Mineral Research*

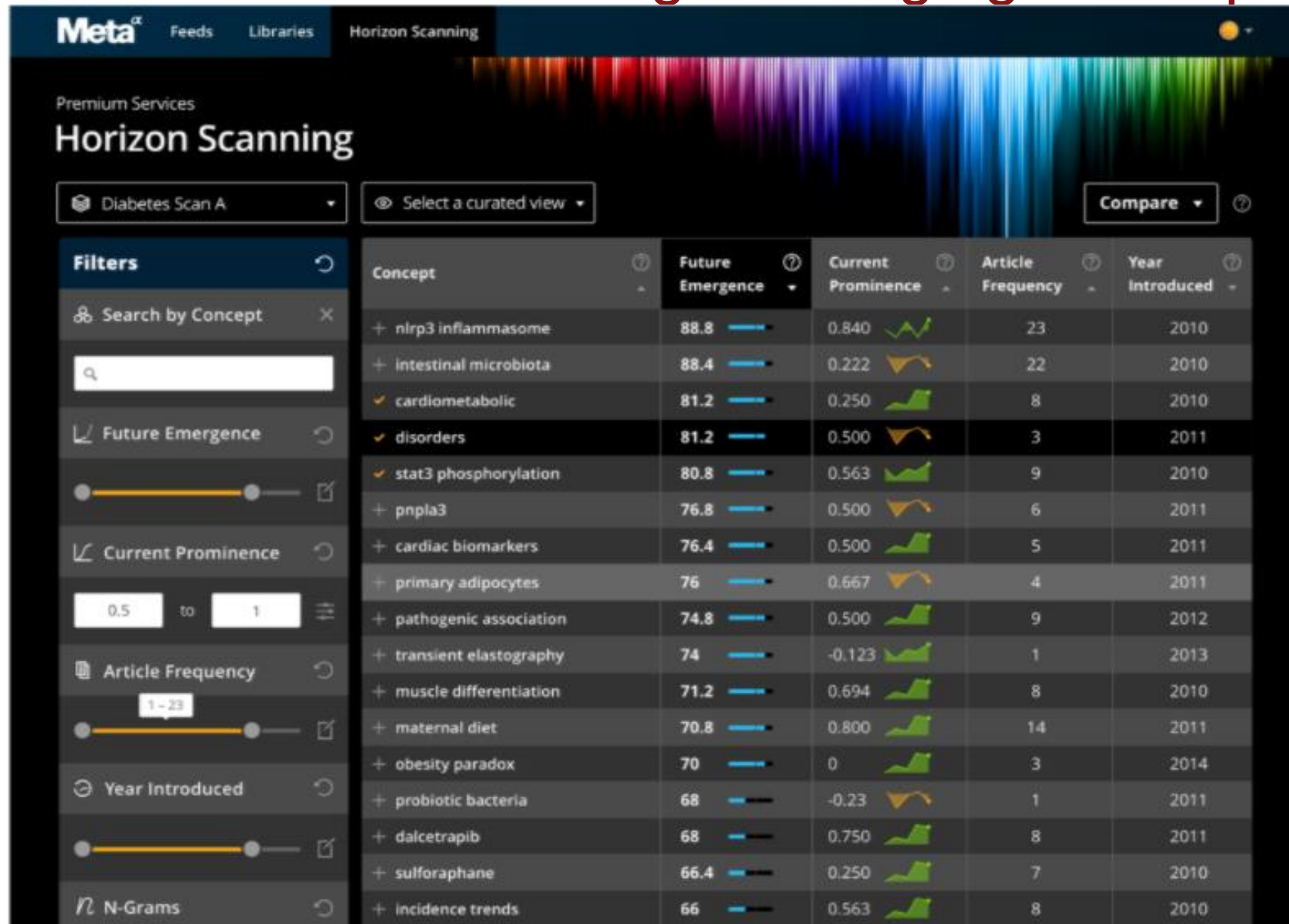


# Fishing in the Preprint Stream for Prospective Papers

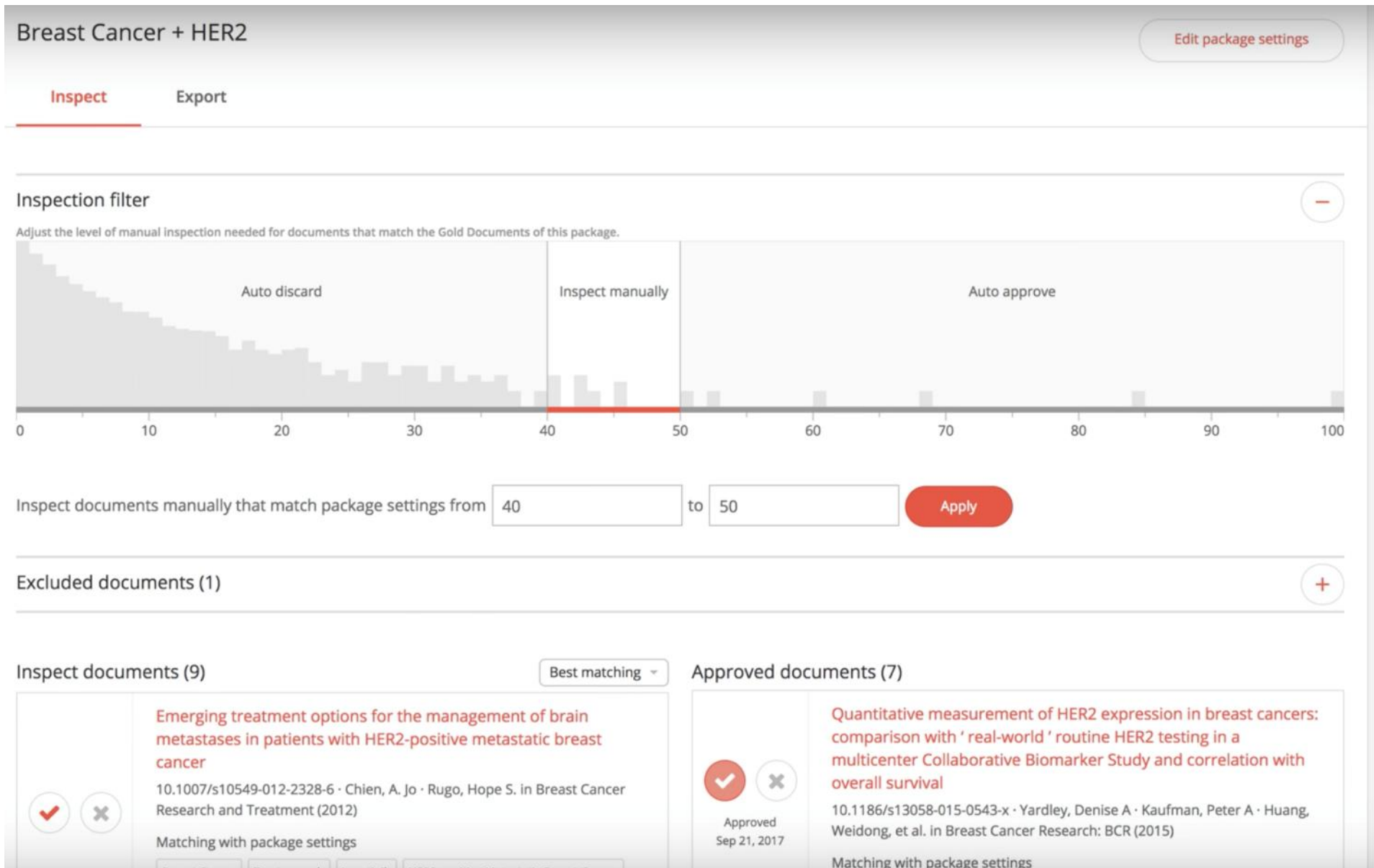
- Journals can leverage preprints by finding the best and most appropriate articles to solicit
- Already happening manually



# Meta “Horizon Scanning”: Emerging Concepts



# UNSILO Article Categorization Dashboard



# Evaluation of Figure Integrity

New Results

## Bioscience-scale automated detection of figure element reuse

Daniel E Acuna, Paul S Brookes, Konrad P Kording

doi: <https://doi.org/10.1101/269415>

This article is a preprint and has not been peer-reviewed [what does this mean?].

Abstract

Info/History

Metrics

 Preview PDF

### Abstract

Scientists reuse figure elements sometimes appropriately, e.g. when comparing methods, and sometimes inappropriately, e.g. when presenting an old experiment as a new control. To understand such reuse, automatically detecting it would be important. Here we present an analysis of figure element reuse on a large dataset comprising 760 thousand open access articles and 2 million figures. Our algorithm detects figure region reuse, while being robust to rotation, cropping, resizing, and contrast changes, and estimates which of the reuses have biological meaning. Then a

“We suggest that a great deal of scientific fraud will be, sooner or later, detectable by automatic means.”

# Topics: AI/DS. Authors/Editors

- AI Applications Overview
- **AI and Data Science in Publishing**



- Some Cautions



Data-Science  
Guided Decision on  
Which journals to  
Submit a paper to:

Is this journal  
a club  
I aspire  
to join?

Journal of Biological Chemistry	
h5-index:107 h5-median:134	
#3 Biochemistry	
Title / Author	Cited by
Exosome-associated Tau Is Secreted in Tauopathy Models and Is Selectively Phosphorylated in Cerebrospinal Fluid in Early Alzheimer Disease S Saman, WH Kim, M Raya, Y Visnick, S Miro, S Saman, B Jackson, ... Journal of Biological Chemistry 287 (6), 3842-3849	323
Peroxiredoxin Functions as a Peroxidase and a Regulator and Sensor of Local Peroxides SG Rhee, HA Woo, IS Kil, SH Bae Journal of Biological Chemistry 287 (7), 4403-4410	270
$\alpha$ -Synuclein in Central Nervous System and from Erythrocytes, Mammalian Cells, and Escherichia coli Exists Predominantly as Disordered Monomer B Fauvet, MK Mbefo, MB Fares, C Desobry, S Michael, MT Ardah, E Tsika, ... Journal of Biological Chemistry 287 (19), 15345-15364	263
Microtubule-associated Protein 1 Light Chain 3 (LC3) Interacts with Bnip3 Protein to Selectively Remove Endoplasmic Reticulum and Mitochondria via Autophagy RA Hanna, MN Quinsay, AM Orogo, K Giang, S Rikka, AB Gustafsson Journal of Biological Chemistry 287 (23), 19094-19104	236
Peroxynitrite, a Stealthy Biological Oxidant R Radi Journal of Biological Chemistry 288 (37), 26464-26472	232
Mitochondrial Complex II Can Generate Reactive Oxygen Species at High Rates in Both the Forward and Reverse Reactions CL Quinlan, AL Orr, IV Perevoshchikova, JR Treberg, BA Ackrell, ... Journal of Biological Chemistry 287 (32), 27255-27264	229

Data-Science  
Guided  
Reading

[Browse](#)
[Publish](#)
[About](#)

advanced search

OPEN ACCESS
 PEER-REVIEWED

RESEARCH ARTICLE

### Detecting Individual Sites Subject to Episodic Diversifying Selection

Ben Murrell, Joel O. Wertheim, Sasha Moola, Thomas Weighill, Konrad Scheffler, Sergei L. Kosakovsky Pond

368 Save

388 Citation

15,767 View

8 Share

# Topics: AI/DS. Authors/Editors

- AI Applications Overview
- **AI and Data Science in Publishing**

	Readers / Authors	Editors / Publishers
Artificial Intelligence	1	2
<b>Data Science</b>	3	<b>4</b>

- Some Cautions

# Impact Vizor:

## Data-driven Insight into Editors' Big Questions

- Rejected Articles
  - Who publishes my rejected articles?
  - Am I rejecting high-impact articles?
- Published Articles
  - Which articles are driving my impact up, or down?
  - Which types of articles are resonating?
  - Are this year's articles higher impact than last year's?
- Cohort Articles
  - What are the trending topics in my fields?
  - And who is publishing them?

# The Different Faces of Impact Vizors' Viewers



< Rejected  
Article  
Tracker

Section >  
Performance  
Analyzer

< Hot Article  
(& Object)  
Tracker

Advance >  
Correlator of  
Citations &  
Usage

< Citation  
Distribution  
Surveyor

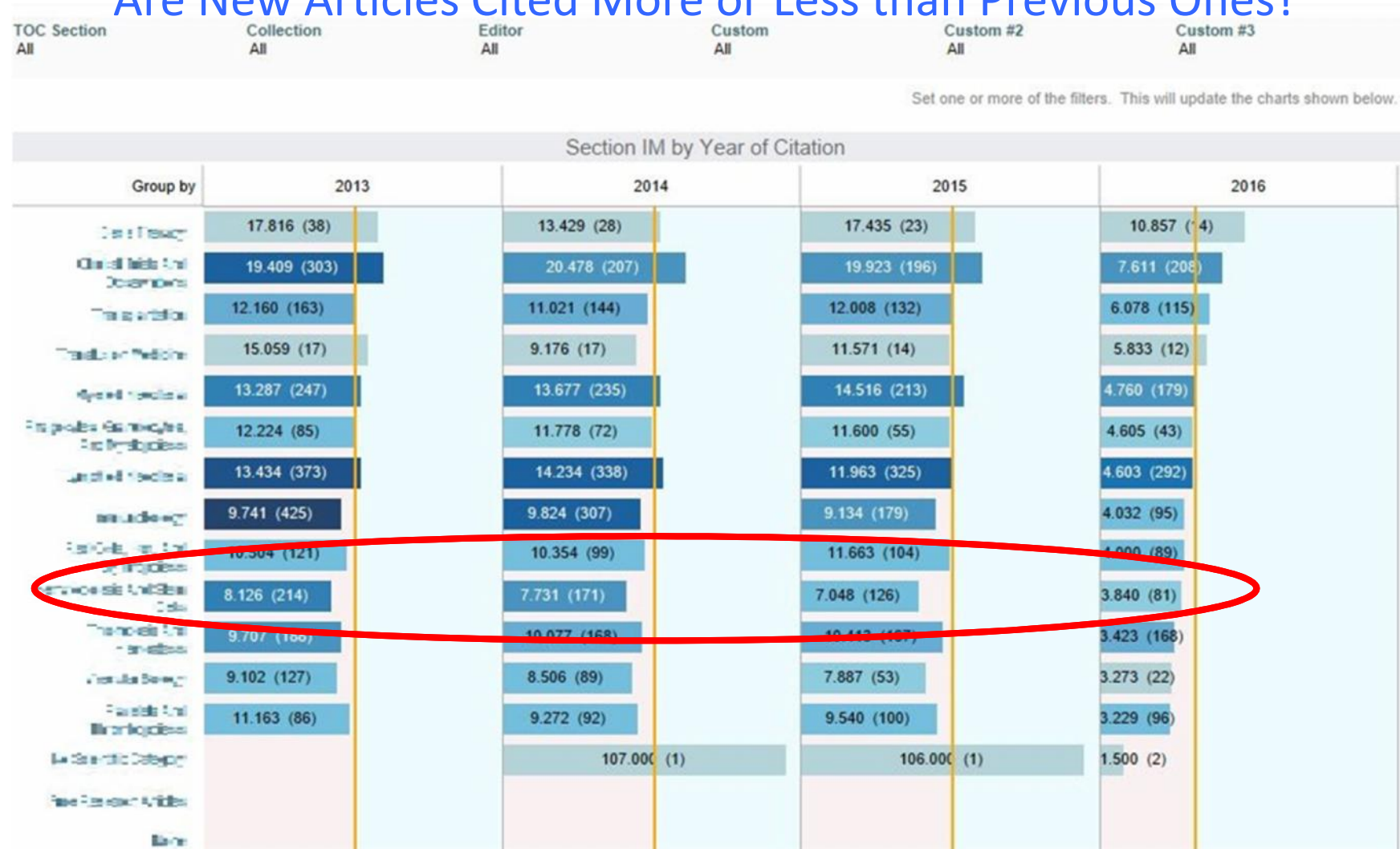
Cohort >  
Comparator



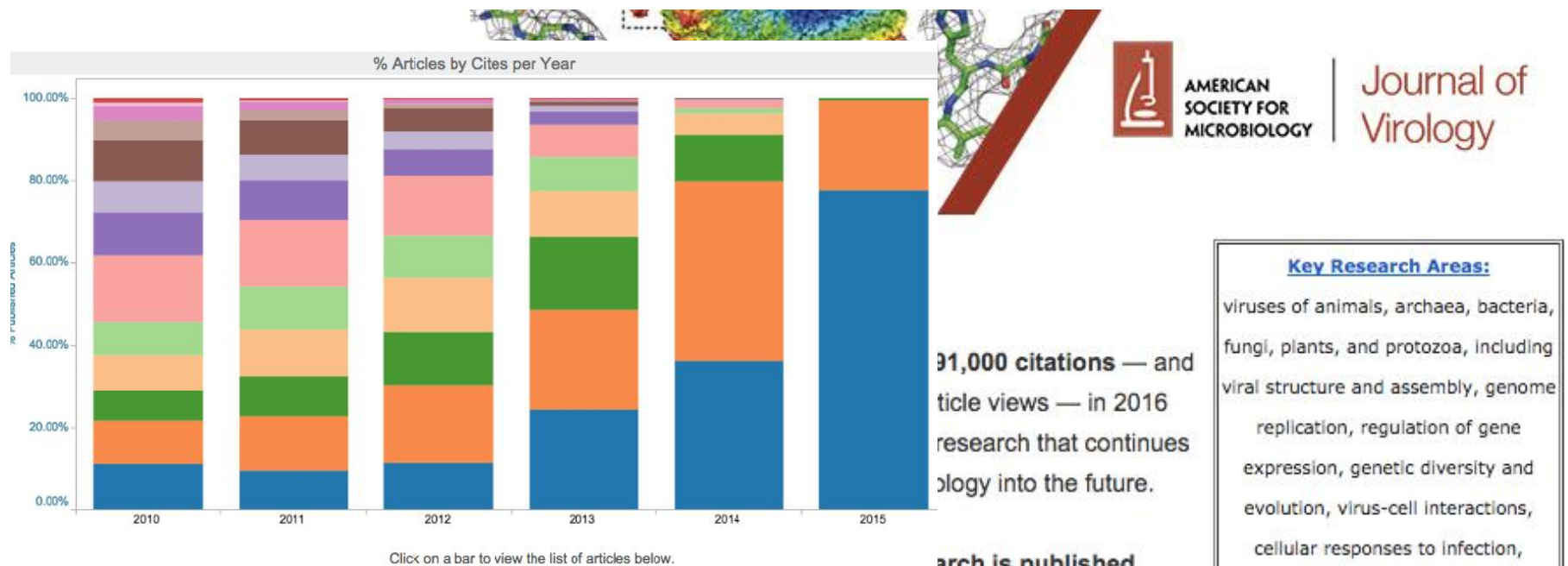
# Which Sections of a Journal are Consistently Below Par?

What Groups of Articles Are Cited Above and Below Average

Are New Articles Cited More or Less than Previous Ones?







Journal of  
Virology

**91,000 citations — and  
article views — in 2016  
research that continues  
virology into the future.**

**Key Research Areas:**

viruses of animals, archaea, bacteria,  
fungi, plants, and protozoa, including  
viral structure and assembly, genome  
replication, regulation of gene  
expression, genetic diversity and  
evolution, virus-cell interactions,  
cellular responses to infection,  
transformation, and oncogenesis,  
gene delivery, viral pathogenesis and  
immunity, vaccines, and antiviral  
agents

Marketing to  
authors  
based on the  
assurance  
Of peer attention

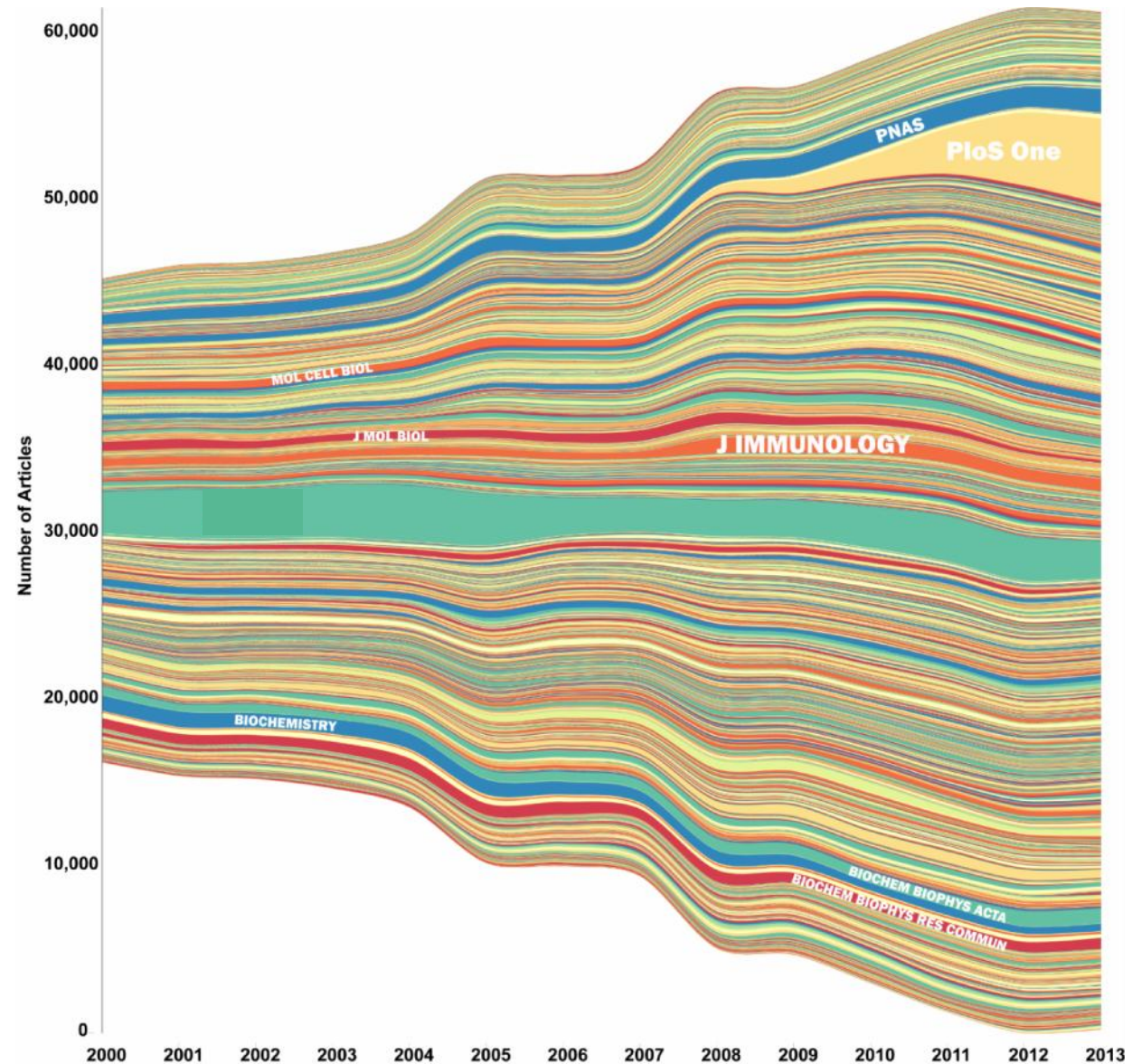
**arch is published  
quickly and accessed easily by worldwide audiences to get  
you cited quickly by your virology peers.**

- **20%** of articles published in 2017 already have been cited this year.
- **75%** of all articles published in 2016 already have been cited at least once.
- Over **90%** of articles published in 2015 have been cited at least once as of this year.
- JVI is the top-cited journal in the Virology Category of the Journal Citation Reports® (JCR).
- **#1** journal ranked by Eigenfactor® in the field of Virology in the JCR.
- Median time to first decision: **23.7 days** (August 2017)

**Read our most cited articles**

# Journal Performance: Competitor Streamgraph

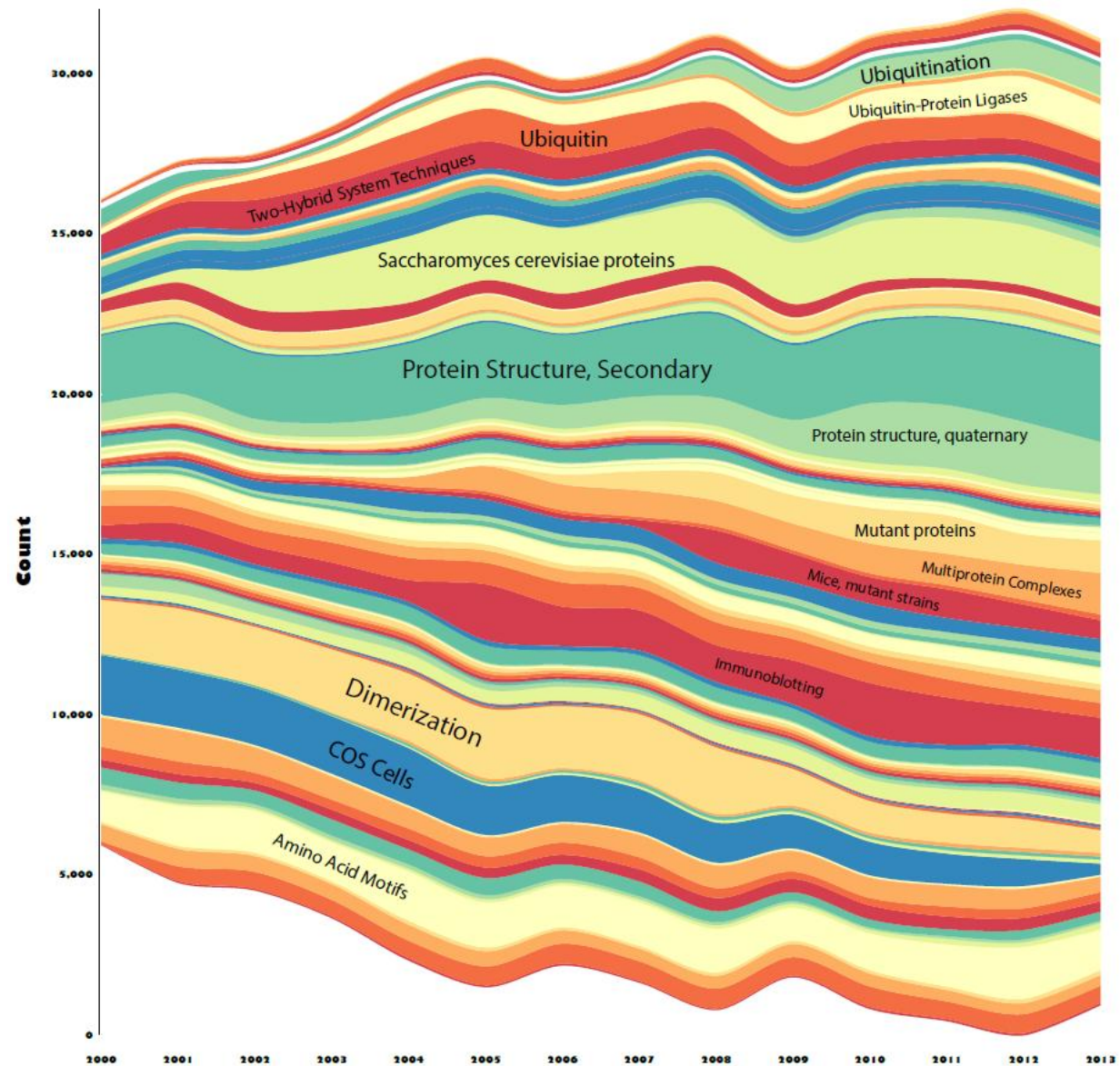
Q: Who is publishing  
In my fields, and  
How much? What is  
My “market share”?





# Journal Performance: Topic Streamgraph

Q: Which fields that I publish in are growing, and which are shrinking?



# Analytics Support for Evidence-Based Design

Heat maps →

↓ Confetti (click) maps



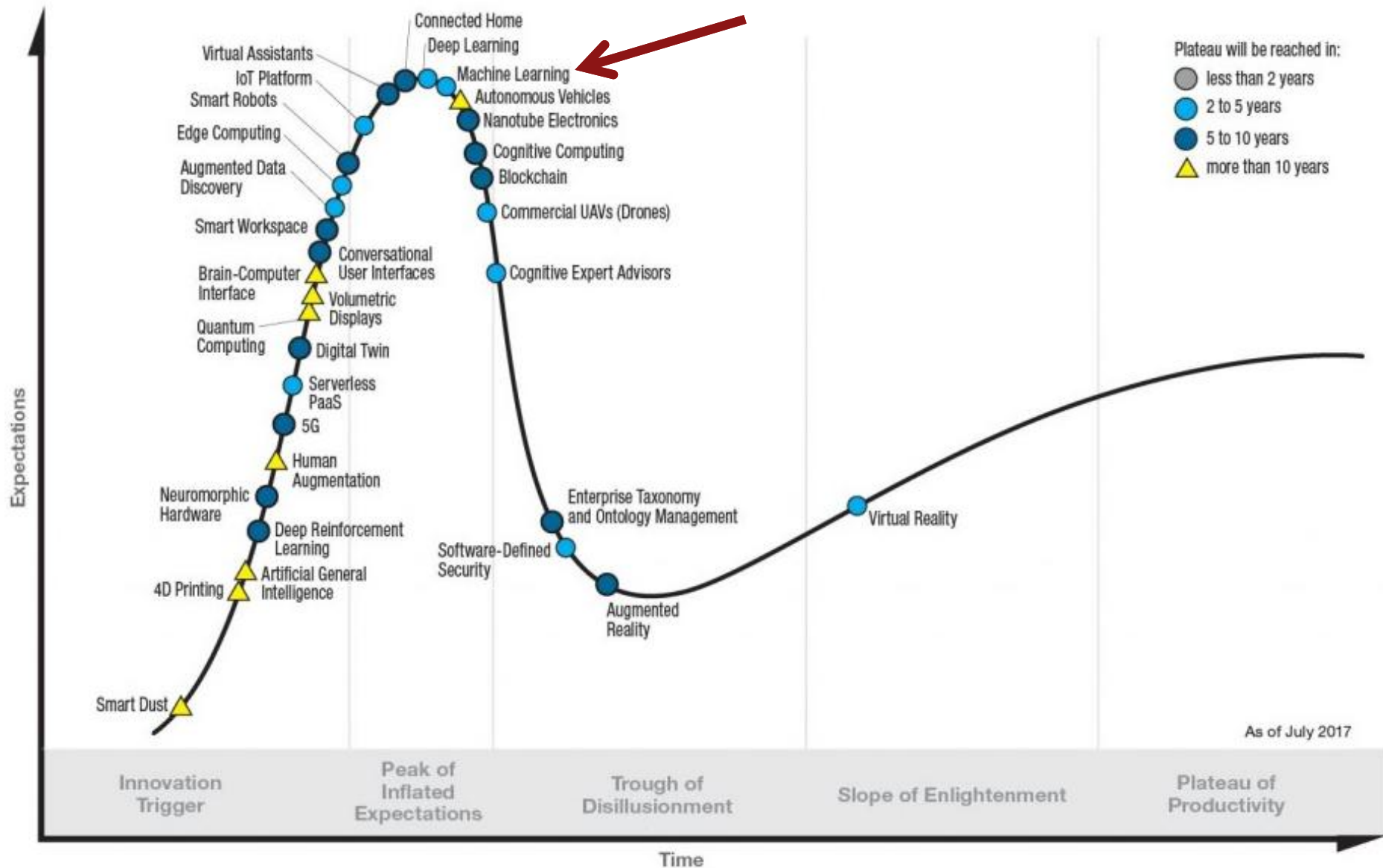
# Topics: AI/DS. Authors/Editors

- AI Applications Overview
- AI and Data Science in Publishing

	Readers / Authors	Editors / Publishers
Artificial Intelligence	1	2
Data Science	3	4

- **Some Cautions**
  - The Hype Cycle, and Validation

# Gartner **Hype Cycle** for Emerging Technologies, 2017



“People tend to

**overstate**

the short-term potential of new technologies and

*understate*

their long-term consequences.”

- Paul Saffo

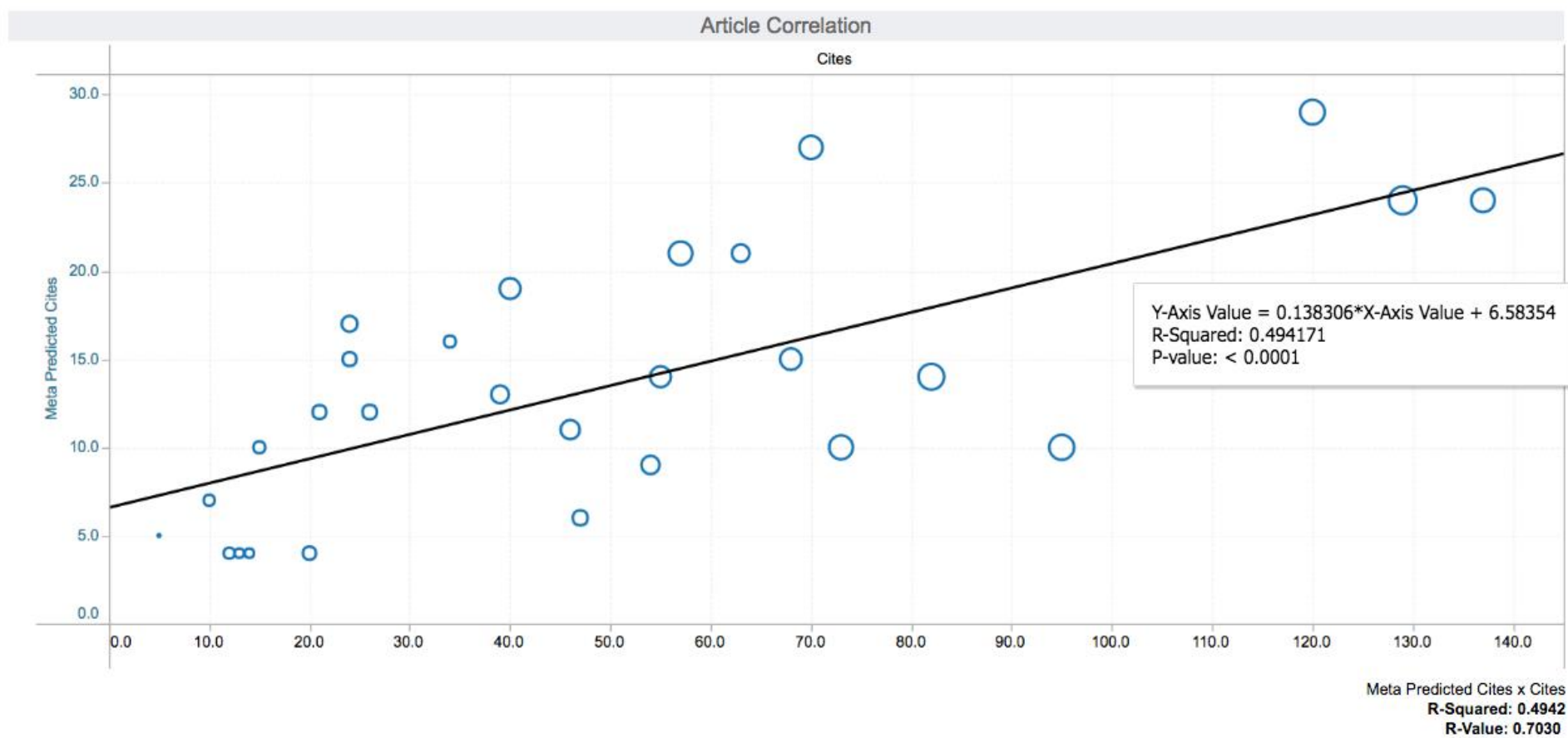
- Ray Amara, “Amara’s Law”





# Validation – Calibration - Confidence

- HighWire's Impact Vizor Validating Meta's Predictions  
– Correlate AI-Predicted Cites to Actual Cites





# Programme

- *Data Science & Artificial Intelligence*

## **Data Science and AI at the Scholarly Communications Frontier**

*Understanding content, workflow and impact, to inform publishing and editorial decision-making*

### **John Sack**

- *Founding Director, HighWire Press*



*New applications of AI/machine learning and predictive analytics are beginning to influence scholarly communication. Data mining across platforms can deliver insight into the lifecycle of individual papers from submission to citations and downloads. This information can help researchers, editors and publishers make decisions and adjust strategy. In this presentation, we will discuss examples that can help editors and publishers use analytics and AI to tune products and uncover emerging opportunities.*



**Researcher  
to Reader**

26 & 27  
February  
2018

