Towards reproducible Artificial Intelligence

Roles and Responsibilities of Researchers and Publishers.

Anita de Waard, **Sweitze Roffel**, Catriona Fennel, Sergios Petridis, Thom Pijnenburg, Efthymios Tsakonas, Rinke Hoekstra, **George Tsatsaronis**

Venue; New York , Feb 7 , AAAI 2020 Workshop Reproducablilty in AI (RAI 2020)

Future Direction and Reproducablilty Challenge

Introductions

- Ir. Sweitze Roffel
  - Originally educated as a chemical engineer
  - Now a publisher at Elsevier – a scientific information and analytics provider
  - Since 2006 responsible for Elsevier's AI & CS journals
  - First based in New York , now in Amsterdam

- Dr. George Tsatsaronis
  - Academic career in data mining, machine learning, NLP between 2004-2016
  - Now a VP Data Science at Elsevier
  - Based in Amsterdam

What is Reproducibility?

Many different definitions and semantics*

One example from Goodman et al. (2016) and Gundersen and Kjensmo (2018):

- **R1: Experiment Reproducible (or 'Repeatable')**

  Draw the same conclusions from independent replication of a study or exact implementation of the AI method.

- **R2: Data Reproducible (or 'Replicable')**

  Obtain the same results from independent study with procedures matched to the original study, i.e. alternative implementation of the AI method.

- **R3: Method Reproducible (or 'Reproducible')**

  "*As a minimal requirement, you should at least be able to reproduce the results yourself.*"

ELSEVIER

* "A program can fail as a scientific contribution in many different ways for many different reasons. Borrowing the terms coined by Goble (2016), for a program to contribute to science, it should be re-runnable (R1), repeatable (R2), reproducible (R3), reusable (R4), and replicable (R5)."

Goble, C. (2016). "What is reproducibility? The R*brouhaha," in First International Workshop on Reproducible Open Science (Hannover). Available online at: http://repscience2016.research-infrastructures.eu/img/CaroleGoble-ReproScience2016v2.pdf (September 9, 2016).

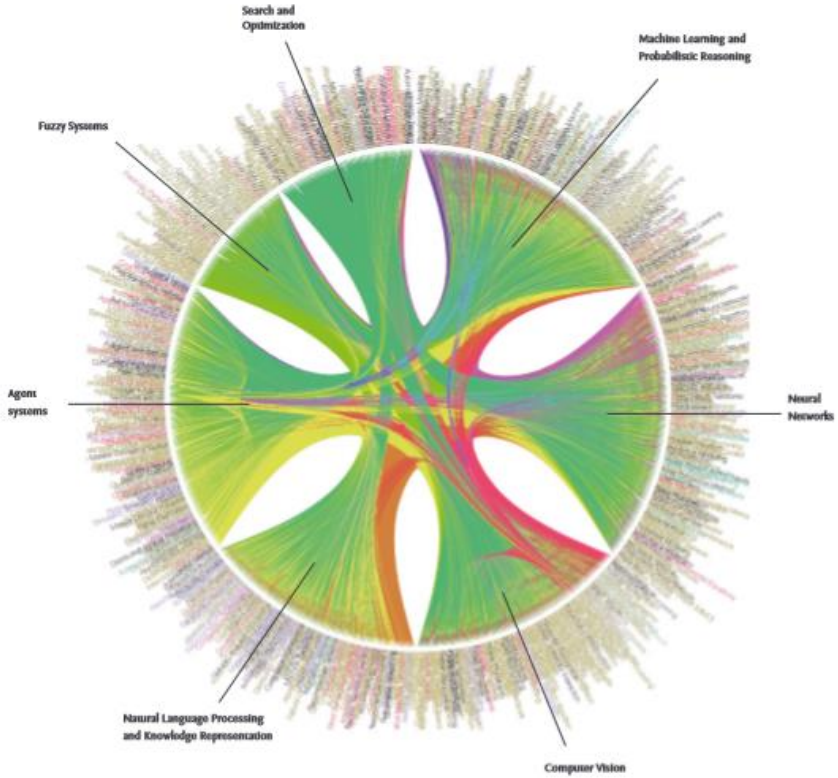
What is AI ?

Many different definitions and semantics

We studied this in 2018

We had to use advanced analytics to wrangle the data - we had to use ai to study ai

2018 rapport freely available*



ELSEVIER

Artificial Intelligence: How knowledge is created, transferred, and used
*2018 Elsevier study comparing AI trends across Europe , USA and China, available at;
https://www.elsevier.com/research-intelligence/resource-library/ai-report

Why does this matter ?

TRUST

- science vs opinion
- 3 different formats to bring new knowledge into the world;
  1. Share R&D in Scientific literature - books, journals, proceedings , open source
  2. Protect R&D - patents & trademarks
  3. Keep R&D secret - share artifacts & closed source
- The knowledge we share “reasons” differently
  - Empirical (shared understanding of how to discover truths about the natural world -eg Chemistry)
  - Formal (shared understanding of rules – eg Mathematics)
  - Narrative (shared understanding of scholarship – eg Histography)
  - Engineering (shared understanding around utility of any humanly engineered artefacts – eg Aerospace)
- Whence we all complain* differently, for many different reasons

ELSEVIER

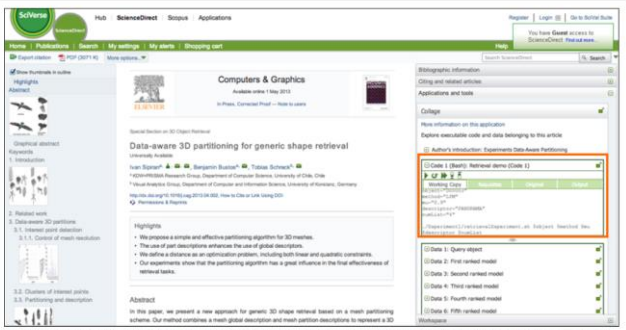
- *see tons of papers or quotes like Iqbal (2016), Stodden , etc where yet another study shows their field fails in reproducing X,Y or Z…*
- *https://www.nap.edu/catalog/25303/reproducibility-and-replicability-in-science*

Visions for transformation

- Executable paper challenge
- The data pyramid
- Force 11 – data and software citations

Executable Paper Challenge

- CS is mostly a virtual science… Why not start there and "redo " this science digitally?
- Elsevier challenged CS community to "digitalize" CS research (Collect , validate , disseminate , archive)
- 50 K USD prize competition & winners to be piloted in real journals
- Competition Finals held at International Conference on Computational Science at NTU in Singapore (ICCS 2011)
- 71 ideas submitted, 15 workshop participants, 9 systems build, 3 winners selected by expert jury on things like; Project quality, Usefulness, Innovation/vision, Evaluation, Scope & Feasibility



An article on ScienceDirect with the Collage Executable Paper components in the right-hand pane. Click on the image to access the article.

First Prize
The Collage Authoring Environment, whose team members include:
Piotr Nowakowski, Eryk Ciepiela, Daniel Harężlak, Joanna Kocot, Marek Kasztelnik, Tomasz Bartyński, Jan Meizner, and Grzegorz Dyk, ACC CYFRONET AGH, Kraków, Poland, and Maciej Malawski of the Institute of Computer Science AGH, Kraków, Poland, and the Center for Research Computing, University of Notre Dame, USA.

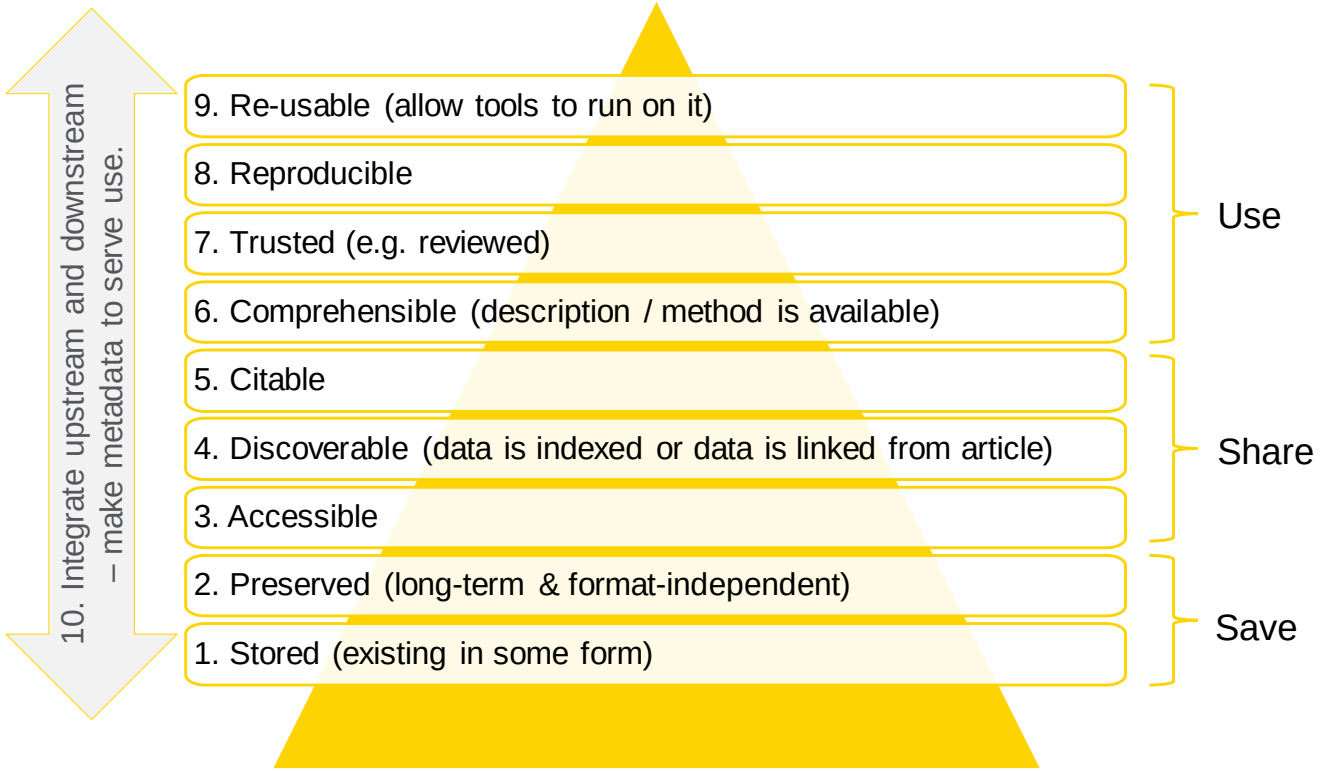
Second Prize
Pieter Van Gorp of TU Eindhoven and Steffen Mazanek, Munich for SHARE: A Web Portal for Creating and Sharing Executable Research Papers (Sharing Hosted Autonomous Research Environments)

Third Prize
Matan Gavish and David Donoho, Stanford University, for A Universal Identifier for Computational Results.

https://www.elsevier.com/about/press-releases/science-and-technology/elsevier-announces-winners-of-the-executable-paper-grand-challenge

Executable Paper Grand Challenge Workshop , Ann Gabriel, Rebecca Capone, https://doi.org/10.1016/j.procs.2011.04.060

Climbing the Data Pyramid



ELSEVIER

Force 11 Data and Software Citation WG

FAIR

Research data should be findable, accessible, inter-operable, and reusable

Force11

Establish a cross-publisher set of author guidelines that mandate data sharing and the use of Data Availability Statements

*https://www.force11.org/datacitationprinciples
** https://www.force11.org/software-citation-principles
*** http://www.copdess.org/enabling-fair-data-project/author-guidelines/

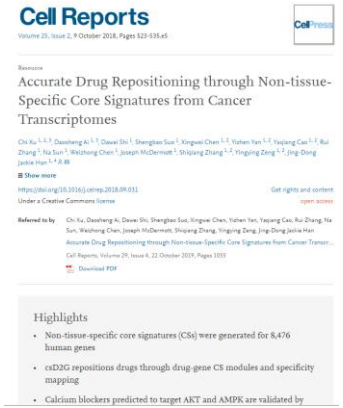
ELSEVIER

Putting vision into practice

- Structuring recommendations & best practices (star methods)
- research elements / software journals / data journals
- computational replication in practice :
  - early pioneers : SIGMOD
  - code ocean
  - Reprozip
  - More

STAR Methods

STAR (Structured, Transparent, Accessible Reporting) Methods promotes transparent reporting of experimental design and methodological details, making the information on how things are done and the resources used clear and easily accessible.


Cell Reports

Accurate Drug Repositioning through Non-tissue-Specific Core Signatures from Cancer Transcriptomes

Highlights

- Non-tissue-specific core signatures (CSs) were generated for 8,476 human genes
- csD2G repositions drugs through drug-gene CS modules and specificity mapping
- Calcium blockers predicted to target AKT and AMPK are validated by

Manuscript

Automatic extraction

Key Resources Table


STAR★Methods

Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
AKT (pan)	Cell Signaling Technology	Cat# 4691; RRID:AB_915783
phospho-AKT (Ser473)	Cell Signaling Technology	Cat# 4060; RRID:AB_2315049
AMPKα	Cell Signaling Technology	Cat# 2532; RRID:AB_330331
phospho-AMPKα (Thr172)	Cell Signaling Technology	Cat# 2535; RRID:AB_331250
GAPDH	Cell Signaling Technology	Cat# 5174; RRID:AB_10622625
Chemicals, Peptides, and Recombinant Proteins		
Benfluorex	Sigma	B7522-5G
Cyproheptadine	Sigma	279072-5G
Nocodazole	Sigma	M1404-2MG
Perhexiline	Sigma	SML0120-10MG
Ellipticine	Sigma	E2451000
Verteporfin	Sigma	SML0534-5MG
Diperodon	Sigma	D8536-5G
Deposited Data		
Non-tissue specific core signatures of human coding gene mutations	This paper; Mendeley Data	https://data.mendeley.com/datasets/8wzmn46d43; http://www.picb.ac.cn/hanlab/csD2G

ELSEVIER

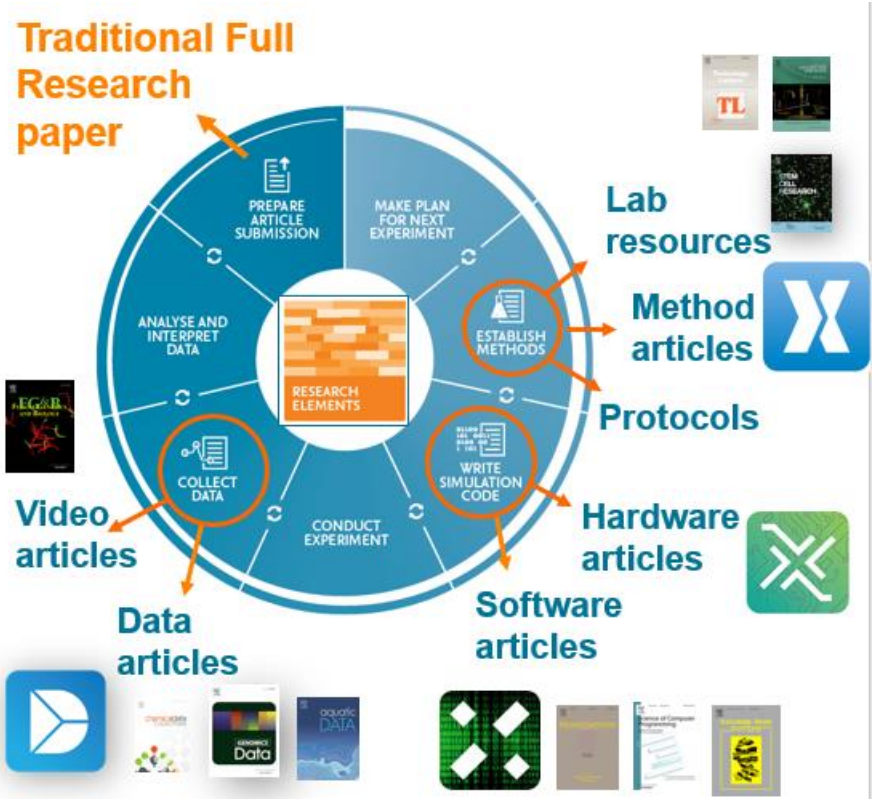
https://www.cs.mcgill.ca/~jpineau/ReproducibilityChecklist.pdf

https://www.elsevier.com/authors/author-resources/key-resources-table

Research Elements

Are …

- Short article types for specific elements of a research cycle
- Complementary to full research papers
- Separately peer-reviewed and indexed
- DOI & citable post-publication updates
- Published both by Multidisciplinary and domain-specific journals
- Primarily Open Access
- *(1) Stored, (2) preserved, (3) Discoverable, (4) Accessible, (5) citable, (6) comprehensible, (7) trusted.*



ELSEVIER

https://www.elsevier.com/authors/author-resources/research-elements

Software Element

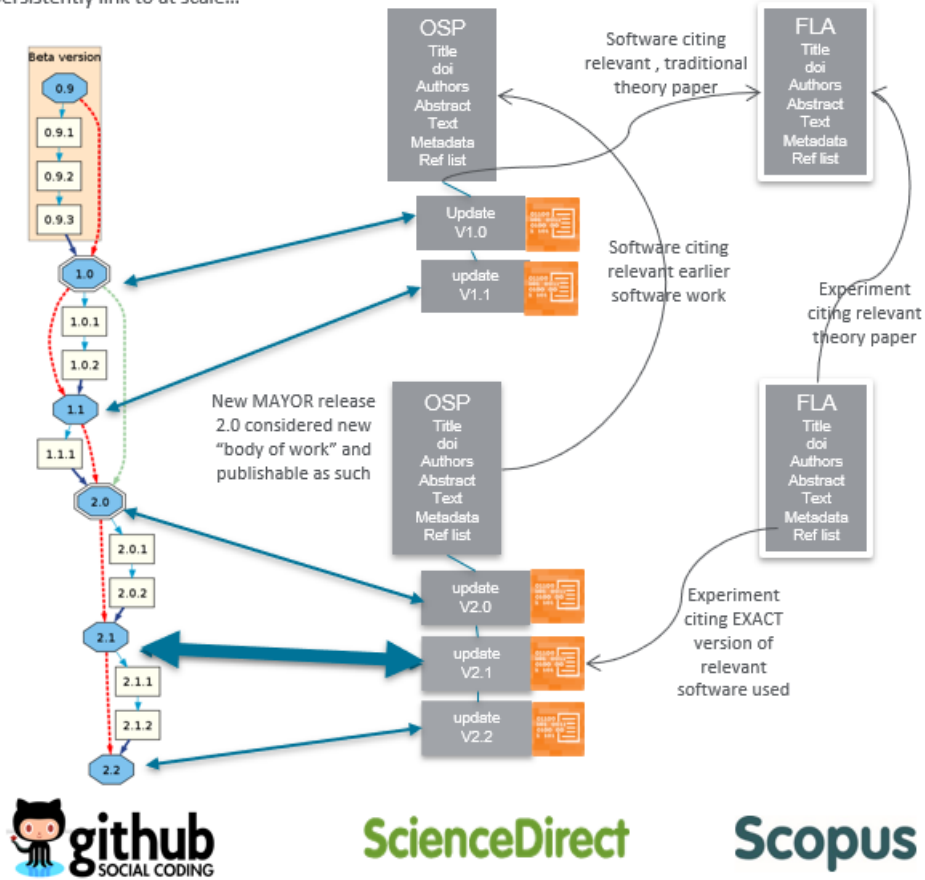
Original Software Publication

- OSP v1 developed in collaboration w Github in 2005
- Many scholars develop software , but scholarly system does not capture this “born digital” research w revision control
  - Developers (authors) can’t claim credit
  - Users (readers) can’t find this valuable content
- *Software is method executed by a machine – and deserved to be treated as a full academic citizen*
- Software papers allow people to formally publish their software (yes ! more papers & citations)

1) Actual Software & code on developers platform use automated versioning systems publishers persistently link to at scale...

2) Though OSP the Software & code can now be peer reviewed, searched, indexed, discovered, archived, cite other work *and* be cited...

3) Creating a richer and connected ecosystem for open science and discovery



ELSEVIER

equal rights 4 soft ware!

BETTER SOFTWARE BETTER RESEARCH

RSE
RESEARCH SOFTWARE ENGINEERS ASSOCIATION

Ian Holmes
@ianholmes

Follow

You can download our code from the URL supplied. Good luck downloading the only postdoc who can get it to run, though #overlyhonestmethods

4:52 PM - 8 Jan 2013

313 RETWEETS **98** FAVORITES

Pioneers

- SIGMOD 2008 was the first database conference that proposed testing the code associated to conference submissions against the data sets used by the authors with the specific aim to test the repeatability of the experiments presented in the submitted papers.
- Report in ACM SIGMOD Record, 37(1):39-45, March 2008.

- Same DB community also drives ACM badging
- ACM badges distiquish between :
  - Repeatability (Same team, same experimental setup)
  - Replicability (Different team, same experimental setup)
  - Reproducibility (Different team, different experimental setup)

https://event.cwi.nl/SIGMOD-RWE/2008/
http://db-reproducibility.seas.harvard.edu/
https://www.acm.org/publications/policies/artifact-review-badging

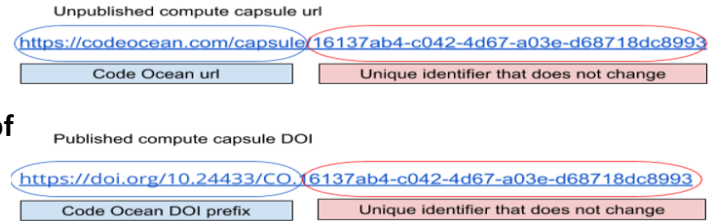
Pilot with Code Ocean

MENDELEY DATA

CODE OCEAN

- cloud-based computational reproducibility platform to run code published in academic journals and conferences.

- Elsevier testing (privately shared) sub-platform for peer-review; journals in pilot:
  - Multidisciplinary **: SoftwareX**
  - Physics **: Computer Physics Communication , Astronomy and Computing**
  - Computer Science **: Neurocomputing**, **Knowledge-based Systems**, **Internet of Things**, **Science of Computer Programming**, **Future Generation Computer Systems**, **Journal of Computational Science**
  - Life sciences (Cell Press) **: Cell Systems, Biophysical Journal**



- Pilot running over 2020, aims to review and publish 40+ compute capsules

1. author uploads code to CO, CO creates link to private compute capsule for reviewer
2. reviewer get link and checks output or runs code on CO, iterates with author in EES/EVISE, corrections implemented in CO before capsule is accepted for publication -
3. Publication assigs DOI and special page both on Code ocean and Mendeley data

ELSEVIER

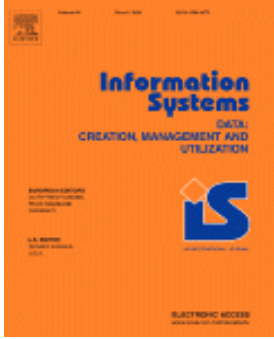
Reproducibility papers

1. Paper is submitted , Information Systems Journal editors check if submission might be reproducible
2. If so ,the journal looks for 2nd lab willing to replicate this research
3. Asks the authors to package all computational resources needed (software, data & environment in a docker container – REPROZIP) providing full OS virtualization and submit this together with the traditional paper
4. 2nd team accesses these resources via the usual peer review systems and process , unpacks this locally to try to replicate this research (same experiment , same methods , different lab)
5. 2nd team documents their experiences in a "reproducibility paper"
6. The original paper , the "How to guide " reproducibility paper and the virtual container containing everything needed to actually replicate this research is permanently linked , published and archived.

MENDELEY DATA

ScienceDirect

ReproZip



This Tuesday Feb 11

New York City

– for ONE DAY only –

Meet & Greet the IS Reproducibility Editor

-Fernando Chirigati –

For a live reproducibility demo at the Elsevier AAAI 2020 Booth

1) Stored, (2) preserved, (3) Discoverable, (4) Accessible, (5) citable, (6) comprehensible, (7) trusted, (8) R1 Reproducible, (9) re-usable

ELSEVIER

A collaborative approach to computational reproducibility, Fernando Chirigati (Reproducibility Editor, Information systems), Rebecca Capone (Publisher), Rémi Rampin, Juliana Freire, Dennis Shasha (Editor-in-Chief, Information systems), https://doi.org/10.1016/j.is.2016.03.002
https://www.elsevier.com/connect/new-article-type-verifies-experimental-reproducibility

Trends* in AI

Research in AI is increasing at a tremendous rate. Over the last 5 years:

- 12,9% annual growth in AI publications.
- 37,4% annual growth in pre-prints in arXiv.

After current AI Summer,

…is the next Winter coming?

- Winter brought on by loss of TRUST in community's claims
- SCIENCE <=> REPRODUCEBILITY <=> TRUST

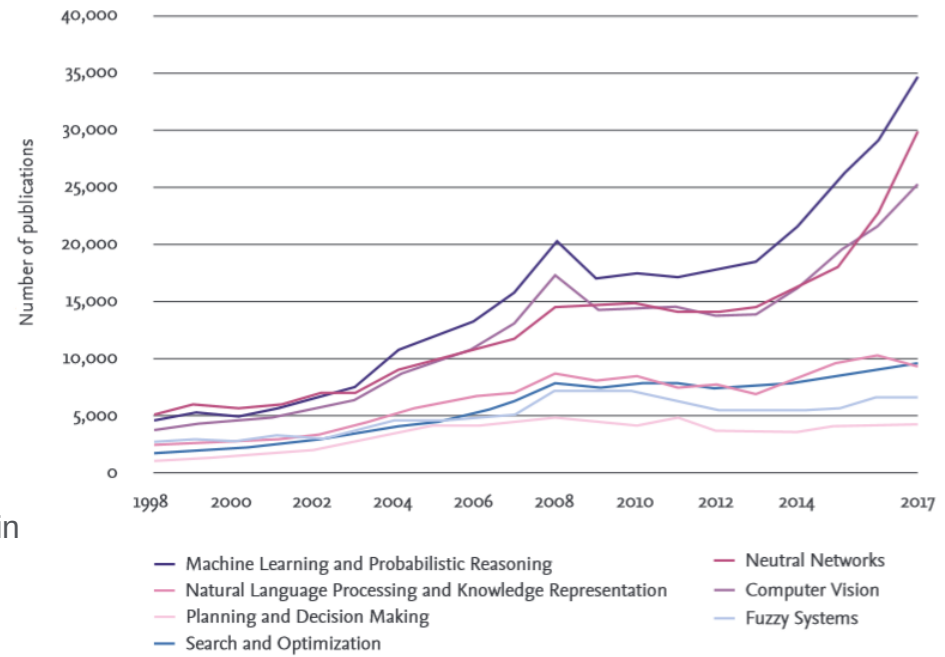


FIGURE 3.2
Annual number of AI publications by keyword co-occurrence cluster (all document types), 1998-2017; sources: Scopus and Elsevier clustering

ELSEVIER

Artificial Intelligence: How knowledge is created, transferred, and used
*2018 Elsevier study comparing AI trends across Europe , USA and China, freely available at; https://www.elsevier.com/research-intelligence/resource-library/ai-report

E -science, Open science, Digital science

- Towards integrated research workflows*, Taverna, Research Objects**, et all … However
  - Continued semantic confusion and labels
  - everyone calls uses different names for stuff? (see useful overview in Plesser 2018)**
  - ACM badges vs Psychology badges
- Elsevier's recent transformative framework agreement with Netherlands helps create new conditions*** to scale progress;
  - *Interoperability: no vendor lock in, researchers and institutions can also use their own tools*
  - *Future proof: system should be flexible to different setups and different agreements*
  - *Vendor/publisher neutral: system development is not limited to any specific vendor*
  - *Researchers and/or institutions own their own research data (not Elsevier)*

> "A major breakthrough for open science in the Netherlands"

ELSEVIER

*Brinckman et al (2019) ; Computing environments for reproducibility: Capturing the "Whole Tale" FGCS https://doi.org/10.1016/j.future.2017.12.029,
Carole Goble at 21st Genomics Standards Consortium meeting 2019,Vienna :https://www.slideshare.net/carolegoble/reproducible-research-how-could-research-objects-help
Reproducibility vs. Replicability: A Brief History of a Confused Terminology, **Hans E. Plesser, Front. Neuroinform., 18 Jan 2018 https://doi.org/10.3389/fninf.2017.00076
***see *https://www.nwo.nl/en/news-and-events/news/2019/12/dutch-research-institutions-and-elsevier-reach-framework-agreement.html

Conclusion; Publisher's Role & Responsibilities

How do we incentivize researchers to make their work (more) reproducible?

- Think beyond "paper behind glass"
  - Code & Data important research outputs
  - Digitalize knowledge implicitly contained in papers and in structures
- Think Scale
  - Use available tech
  - Use available venues
  - Use available organizations
- (academic) Life is complicated enough as it is . Don't redo the wheel again .Work together & learn from each other to implement standards and best research practices. Make it easy and fun.
- Systematic trial and error makes for progress . Usually more error . But we are not alone

ELSEVIER

Thank you

Sweitze Roffel & Georgios Tsatsaronis
Elsevier