Digital chemistry: Paradigm shift, cultural evolution or business as usual?

Jeremy Frey University of Southampton Head Computational Systems Chemistry

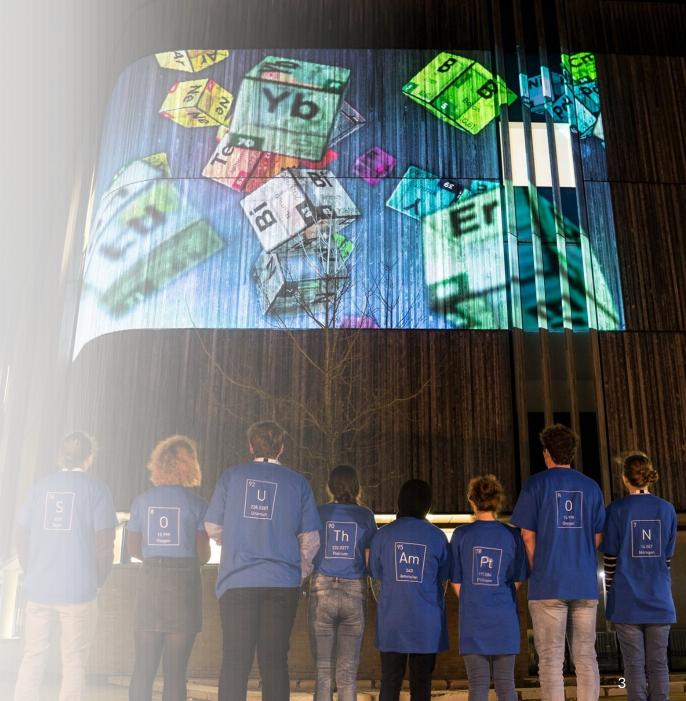


Digital chemistry: Paradigm shift, cultural evolution or business as usual?

Jeremy G. Frey

School of Chemistry & Chemical Engineering

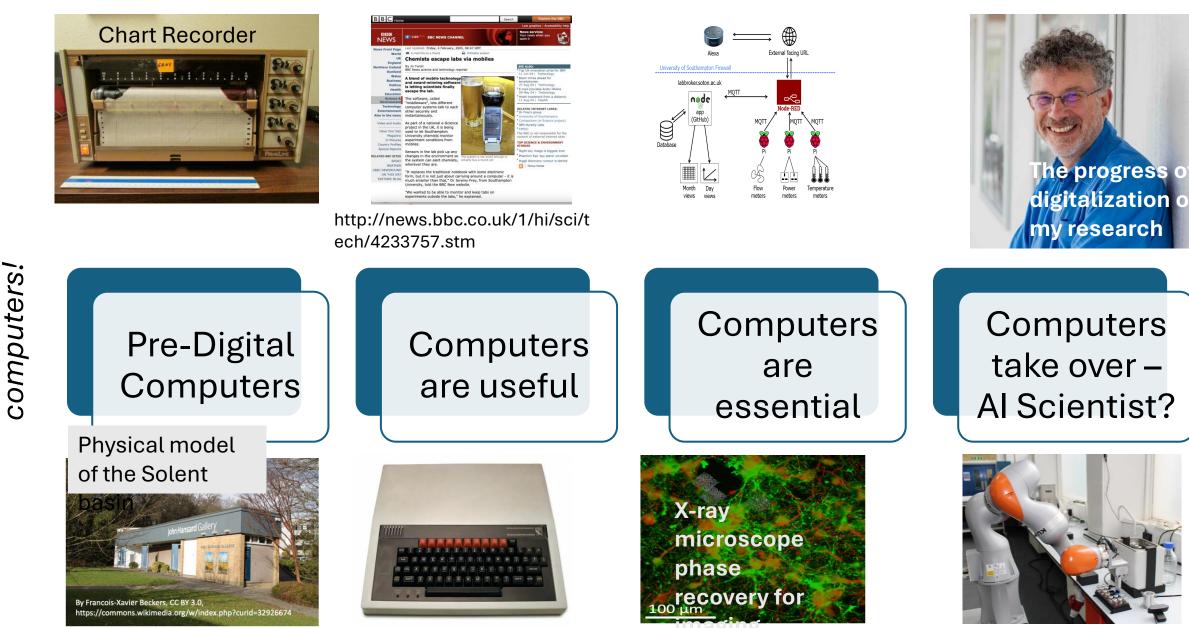
University of Southampton



Science What we investigate

Digital Technoscape How we investigate

Vision



https://en.wikipedia.org/ wiki/BBC_Micro

ChemTalk Basel 2024

25/09/2024

Yes! We did do experiments before

7

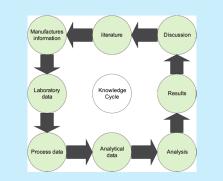
Robotic Labs?

2004

Dark Lab or Smart Lab: The Challenges for 21st Century Laboratory Software

Introduction

Abstract



Dark Lab or Smart Lab: The challenges for 21st Century Laboratory Software, Frey 2004

The Paperless Office or Paperless Laboratory Syndrome

Digital Discovery

PERSPECTIVE

Check for updates Cite this: DOI: 10.1039/d4dd00130c

Digital chemistry: navigating the confluence of computation and experimentation - definition, status quo, and future perspective

Stefan Bräse 🕩 *ab

Digital chemistry represents a transformative approach integrating computational methods, digital data, and automation within the chemical sciences. It is defined by using digital toolkits and algorithms to simulate, predict, accelerate, and analyze chemical processes and properties, augmenting traditional experimental methods. The current status guo of digital chemistry is marked by rapid advancements in several key areas: high-throughput screening, machine learning models, guantum chemistry, and laboratory automation. These technologies have enabled unprecedented speeds in discovering and molecules, materials, and reactions. Digital retrosynthesis and structure-active prediction tools have supported these endeavors. Eurthermore, integrating large-language models and robotics in chemistry labs (e.g. demonstrated in self-driving labs) have begun to automate routine tasks and complex decisionmaking processes. Looking forward, the future of digital and digitalized chemistry is poised for significan growth, driven by the increasing accessibility of computational resources, the expansion of chemica databases, and the refinement of artificial intelligence algorithms. This evolution promises to accelerat more efficient, cost-effective, and environmentally friendly chemical research and production. The challenge lies in advancing the technology itself fostering interdisciplinary collaboration, and ensuring the ethical use of digital tools in chemical research.

Received 13th May 2024 Accepted 15th August 2024 DOI: 10.1039/d4dd00130c rsc.li/digitaldiscovery

(cc) BY

Digital chemistry: navigating the confluence of computation and experimentation definition, status quo, and future perspectives, Bräse 2024

(II) Growth of computational power (the 1970s-1980s):

activity of new compounds before their synthesis. (IV) Rise of high-throughput screening and robotics (1990s-

"Institute of Organic Chemistry, Karlsruhe Institute of Technology, Kaiserstrasse 12, 76131 Karlsruhe, Germany, E-mail: braese@kit.edu Institute for Biological and Chemical Systems - IBCS-FMS, Karlsruhe Institute of Technology, Kaiserstrasse 12, 76131 Karlsruhe, Germany

compounds with their biological activities. This was a major step forward in drug design, allowing for the prediction of the

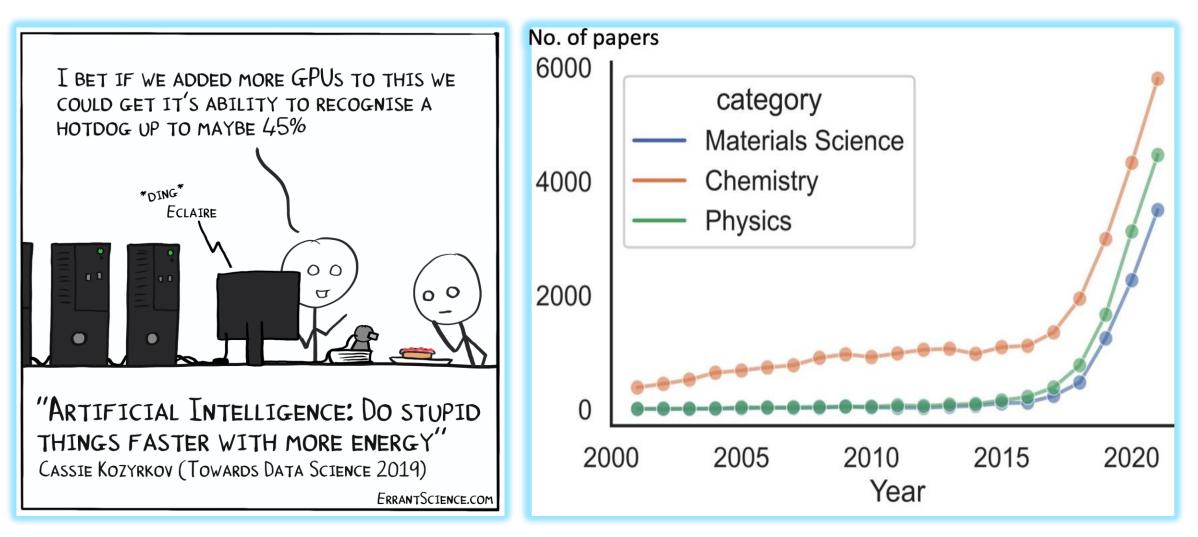
2000s) The integration of high-throughput experimentation (HTE) and screening (HTS) technologies and robotics in the 1990s and

© 2024 The Author(s). Published by the Royal Society of Chemistry

Digital Discovery

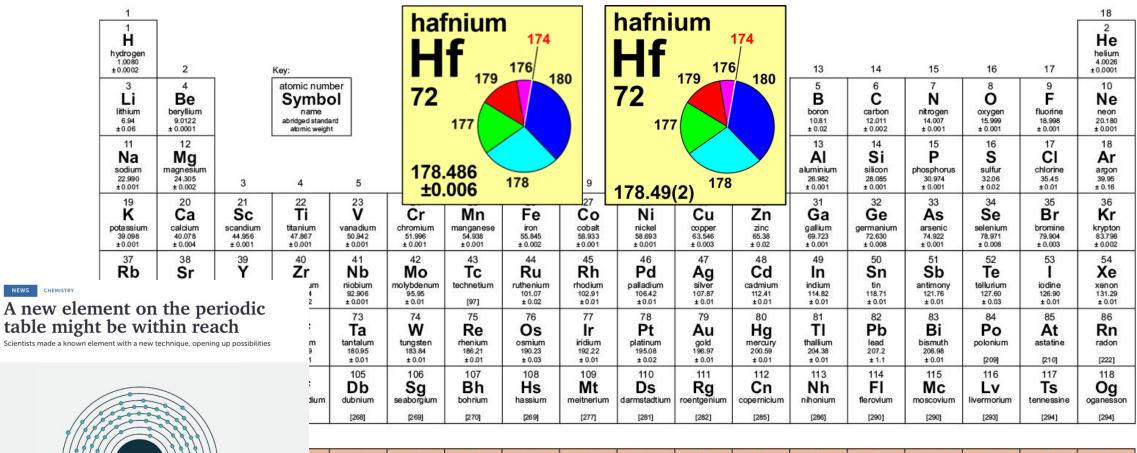
ROYAL SOCIETY OF CHEMISTRY

Augmented Chemical Intelligence?



Data Driven Science

IUPAC Periodic Table of the Elements



L	m	58 Ce cerium 140.12 ± 0.01	59 Pr praseodymium 140.91 ±0.01	60 Nd neodymium 144.24 ±0.01	61 Pm promethium [145]	62 Sm samarium 150.36 ± 0.02	63 Eu europium 151.96 ± 0.01	64 Gd gadolinium 157.25 ± 0.03	65 Tb terbium 158.93 ± 0.01	66 Dy dysprosium 162.50 ± 0.01	67 HO holmium 164.93 ±0.01	68 Er erbium 167.26 ± 0.01	69 Tm thulium 168.93 ± 0.01	70 Yb ytterbium 173.05 ± 0.02	71 Lu lutetium 174.97 ± 0.01
	n	90 Th thorium 232.04 ± 0.01	91 Pa protactinium 231.04 ± 0.01	92 U uranium 238.03 ±0.01	93 Np neptunium [237]	94 Pu plutonium [244]	95 Am americium [243]	96 Cm curium [247]	97 Bk berkelium [247]	98 Cf californium [251]	99 Es einsteinium [252]	100 Fm fermium [257]	101 Md mendelevium [258]	102 No nobelium [259]	103 Lr Iawrencium [262]

25/09/2024

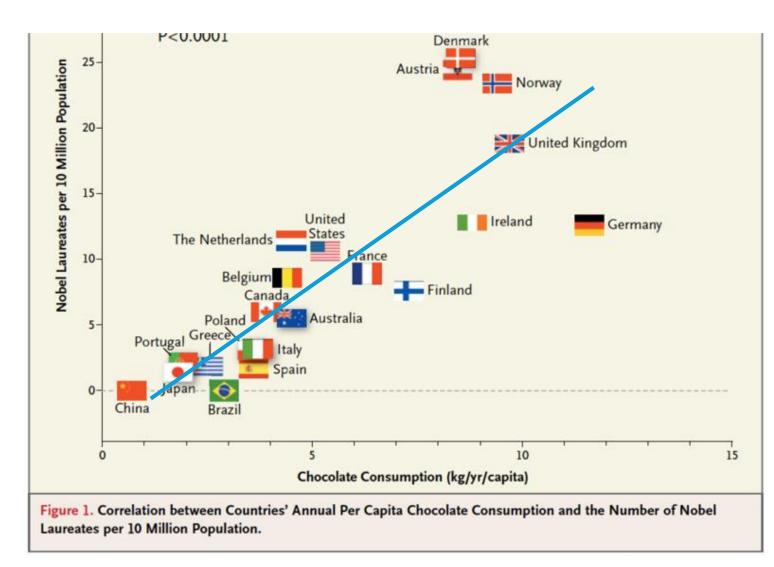
120

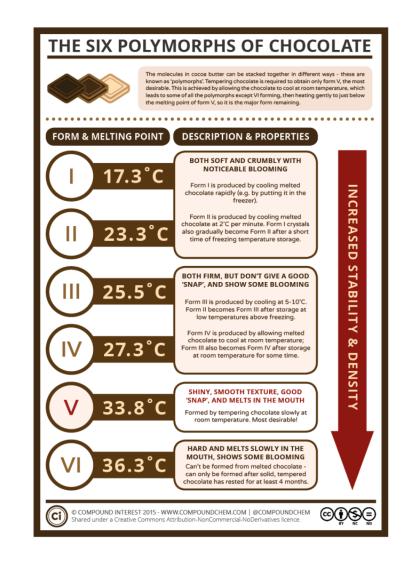
NEWS CHEMISTRY

For notes and updates to this table, see www.iupac.org. This version is dated 4 May 2022. Copyright © 2022 IUPAC, the International Union of Pure and Applied Chemistry. 10

Correlations

Chocolate and Nobel Prizes





PNAS

ZO

25/09/2024

Victoria Buch*,

*Fritz Haber Institute 5250, Centre Nationa Biochemistry, Acade Czech Republic; and

Edited by Mark A. Ra Water autoioniza process of basic i However, pH of (lower, the reduct surface. The evide molecular dynami and OH⁻ ions, der spectroscopic isot rities at the surfac H₃O⁺ does, but O the H₂O surface is reasons, the stren expected to increhave a significant atmosphere. density functional th

n room-temper autoionized, yii ization in crystal i to water, ice is a y

(1). As recently reaction aqueous surface therefore, autoio

• NI	Matter & Energy								
Nature of Water									
> Chemistry									
Inorganic Chemistry									
> Thermodynamics									
> Biochemistry									
N	ucle	ear	En	er	gy				
+ +	+	+	÷	÷	÷	÷	÷	+ D	
-	a		-				-		
	> In > Th > Bi	 Inorg Therr Bioch Nucle 	 Inorgani Thermody Biochem Nuclear 	 Inorganic C Thermodyn Biochemist Nuclear En 	 Inorganic Che Thermodynam Biochemistry Nuclear Energy 	 Inorganic Chem Thermodynamic Biochemistry Nuclear Energy 	 Inorganic Chemist Thermodynamics Biochemistry Nuclear Energy 	 Inorganic Chemistry Thermodynamics Biochemistry 	

Related Topics

R

used

engir

These long-s

Chem

negat

Prof A

Depai "The s chang

ScienceDaily Your source for the latest research news	hité Mixte de Recherche Chemistry and 610 6 Prague,
Mobile: \$ iPhone \$ Android D Web	Follow: S Facebook Twitter S Google+ 006)
HEALTH PHYSICAL/TECH ENVIRONMENT	SOCIETY/EDUCATION QUIRKY constant volume and
Latest Headlines Health & Medicine Mind & Brain	
Featured Research	from universities, journals, and other organizations on was carried out for 1 a slab with two open
Rewrite the textbooks on water's negatively charged by the adsor	surface tension: Air-water interface is to transfer systems
Date: March 19, 2014	Share This utomatically included
Source: University of Melbourne	e high computational
Summary: Researchers in Australi make significant differe water used by the next biophysicists and engin latest investigations hav the physical Chemistry charged by the adsorpt	

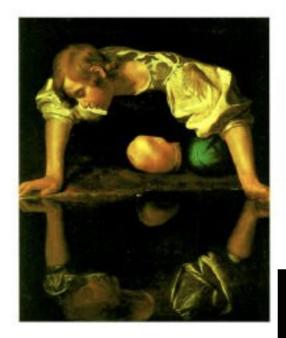
James K. Beattie,* Alex M. Djerdjev and Gregory G. Warr

Received 28th March 2008, Accepted 28th May 2008 First published as an Advance Article on the web 18th September 2008 DOI: 10.1039/b805266b

Problems

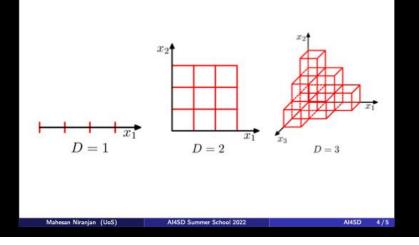
- We need for a map of chemical space
- What do we mean by related?
- What do we mean by "near'?
- In high dimensions useful ideas of distance become less well defined!
- Central Limit Theorem

THE SAME and NOT THE SAME



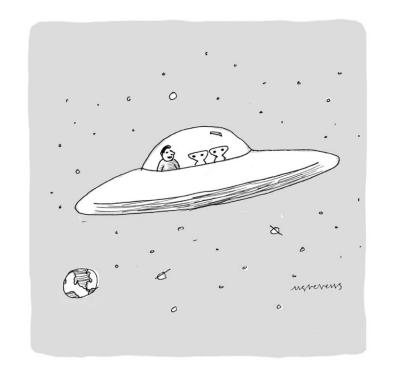






Induction Deduction Abduction

Abduction from Molecules to Chemical Space



How to go from the study of a few molecules to ideas about the entirety of chemical space?

25/09/2024



88886

AI Has Created a Battle Over Web Crawling > Training data may wind up in short supply as websites restrict crawler bots

BY ELIZA STRICKLAND | 31 AUG 2024 | 7 MIN READ | 🗍

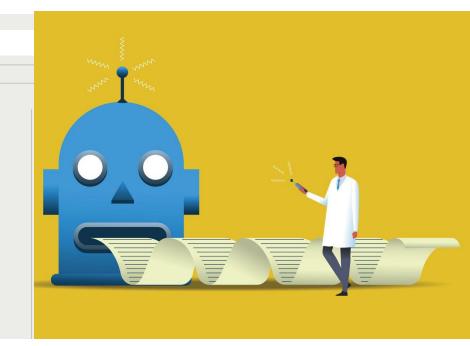
INTERVIEW ARTIFICIAL INTELLIGENCE

Will the "AI Scientist" Bring Anything to **Science?** > A tool to take over the scientific process continues a controversial trend

BY ELIZA STRICKLAND | 09 SEP 2024 | 6 MIN READ | 🗍

Eliza Strickland is a Senior Editor at IEEE Spectrum covering AI and biomedical engineering.

Eliza Strickland is a Senior Editor at IEEE Spectrum covering AI and biomedical engineering.

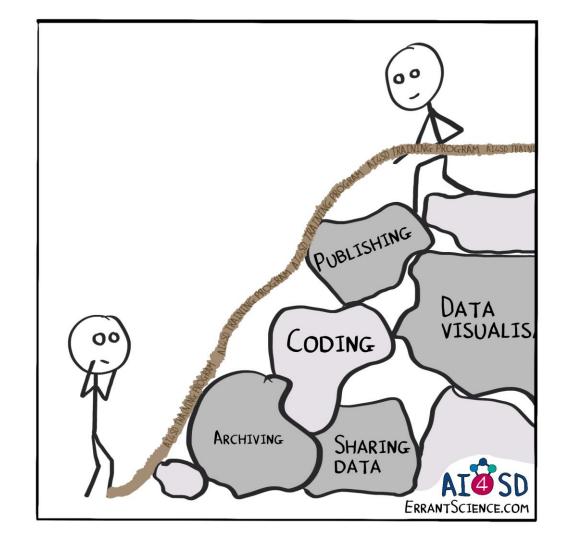




```
IEEE Spectrum
```

NEWS

ARTIFICIAL INTELLIGENCE

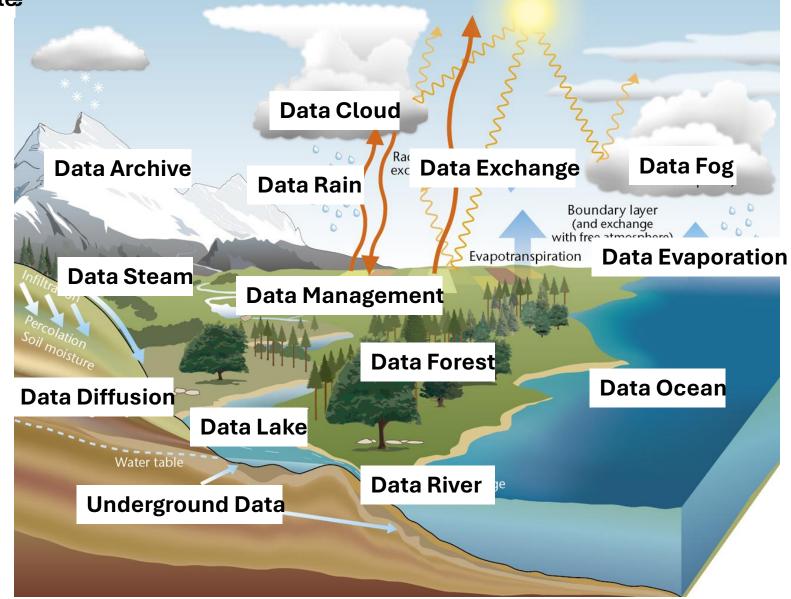


Data, data everywhere but not enough to model

ChemTalk Basel 2024

]

https://www.metoffice.gov.uk/weather/learnabout/weather/how-weather-works/water-cycle



Data-logical Cycle

25/09/2024

ChemTalk Basel 2024

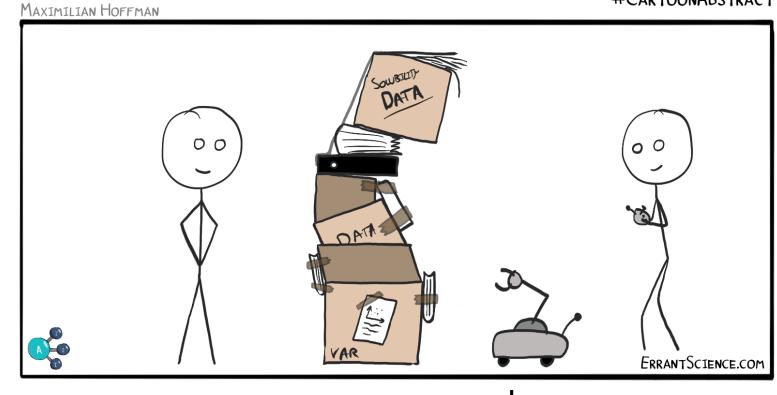
Making Haystacks to find needles



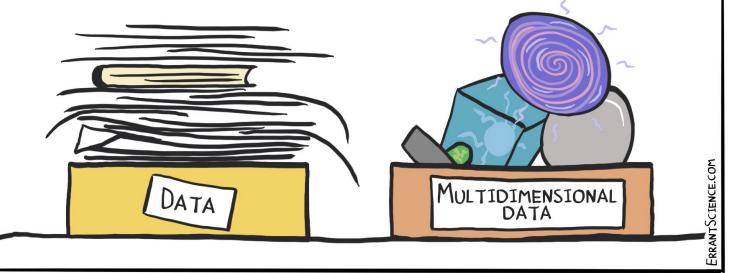


CREATING A MERGED DATASET AND ITS EXPLORATION WITH DIFFERENT MACHINE LEARNING ALGORITHMS

#SKILLS4SCIENTISTS #CARTOONABSTRACT

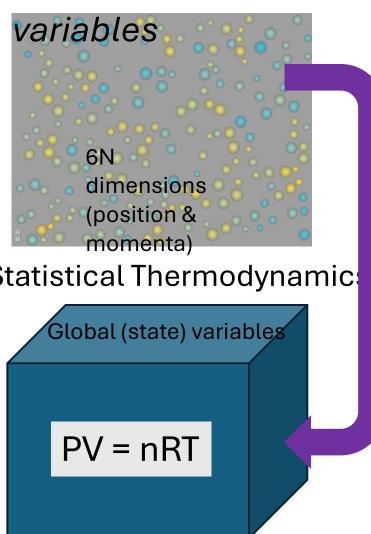


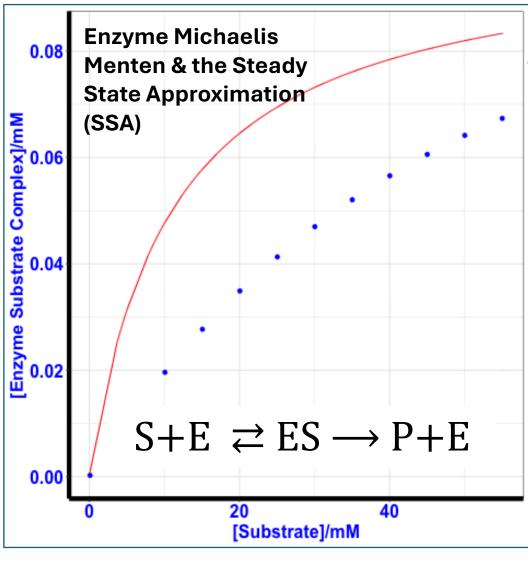
Dimensional Reduction



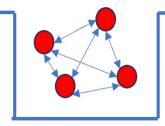
25/09/2024

individual particles to collective

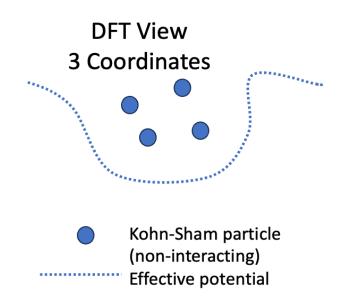




Schrodinger View 3N Coordinates



Formally Equivalent

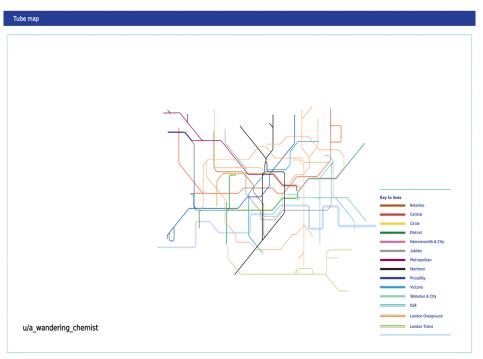


25/09/2024

The Shape of Data – Topological Data Analysis Does Chemical Space have holes?



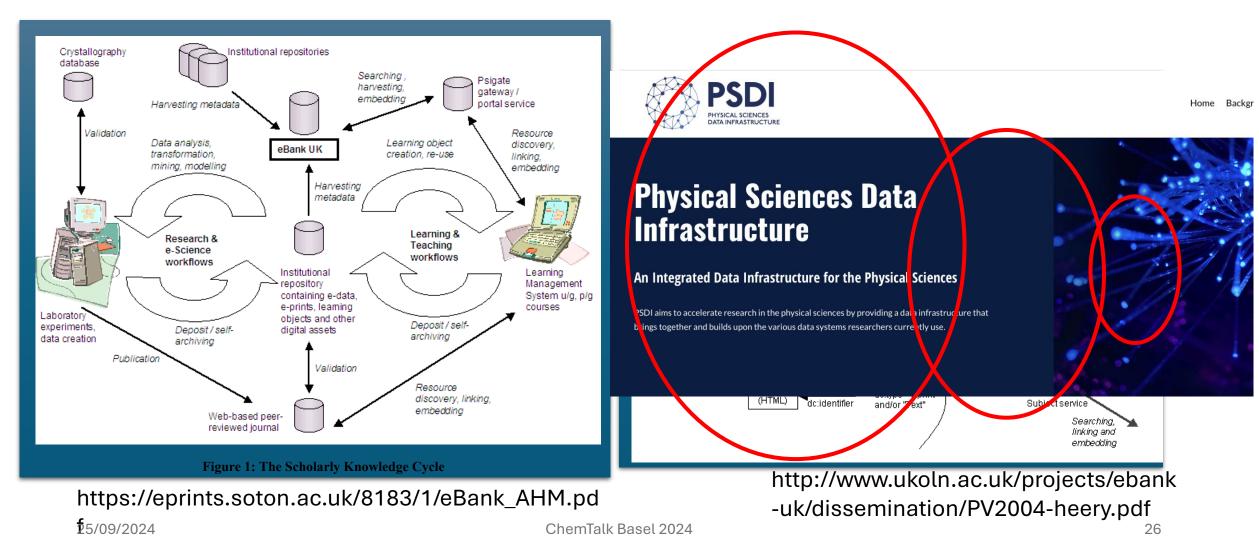




https://pixabay.com/photos/foo d-eat-diet-plain-bagel-2202382/

https://pixabay.com/photos/okto berfest-pretzel-mealgastronomy-963852/ https://www.reddit.com/r/datais beautiful/comments/b8ihhr/com parison_between_the_london_tu be_map_and_its/

At least share you haystack! eScience Data Infrastructures: From eBank to PSDI

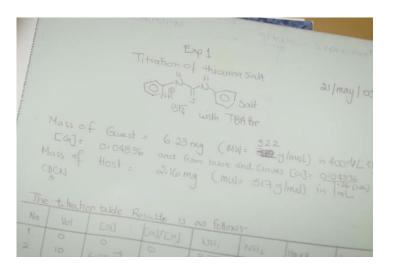


DATA REVIVAL - Recover old inaccessible data





s



https://www.data-revival.com

- UoS stored Chemistry notebooks
- ~ 2000 Chemists years
- ~10 Tb scanned data



https://www.youtube.com/watch?v=kJalCEEYHQY&t=365

Wenhao Sun, Michigan, Ann Arbor, USA "Extracting details of material synthesis from the literature, tells us more about what experimental researchers think about synthetic possibilities, than what is synthetically possible."

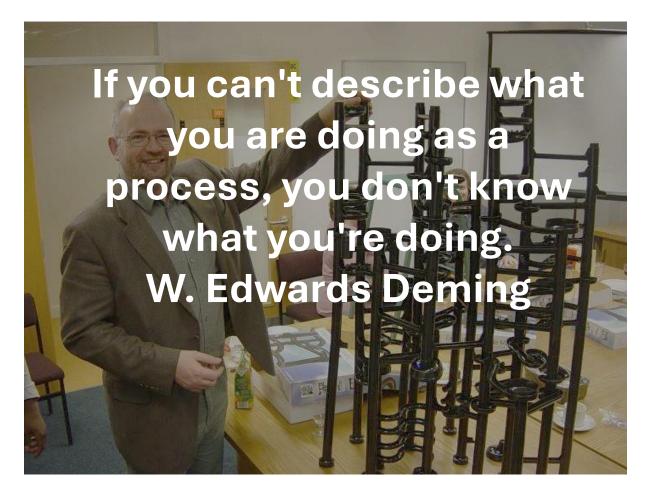
Faraday Discussion, Data Driven Science, Trinity College Oxford, Sept 2024

"What theoreticians think about how experimental researchers work, and the constraints that exist, rarely align with the realities of the laboratory"



How do we communicate models and process?

- Surprisingly difficult to explain what a process involves
- Much of the detail is assumed to be understood and not explicitly discussed
- This is where the missunderstandings usually arise.



Methods are as important as the data

The "ClimateGate" affair - the publication of e-mails and

documents hacked or leaked from one of the world's leading





A weekly series of thought-provoking opinion pieces on environmental topics



Reviving Rio The second Earth Summit holds the promise to cure the world's environmental ills

Your comments

RECENT ARTICLES

EU must increase climate ambition

- Counting the cost of alien invasions
- Has CITES had its day?
- Curious case of Kiwi hedgehogs
- Restoring natural capital
- It's still real and it's still a problem
- Hailing the arrival of alien predators

y/8388485.stm

Alexandrea a Cale aniaire at the alive aniale term

http://news.bbc.co.uk/1/hi/technolog

climate research institutions - is being intensely debated on the web. But what does it imply for climate science? Here, Mike Hulme and Jerome Ravetz say it shows that we need a more concerted effort to explain and engage the public in understanding the processes and practices of science and scientists.

As the repercussions of <u>ClimateGate</u> reverberate around the virtual community of global citizens, we believe it is both important and urgent to reflect on what this moment is telling us about the practice of science in the 21st Century.



Business

Education

Science &

Technology

Environment

Entertainment

Video and Audio

Have Your Say

Country Profiles

Special Reports

Magazine

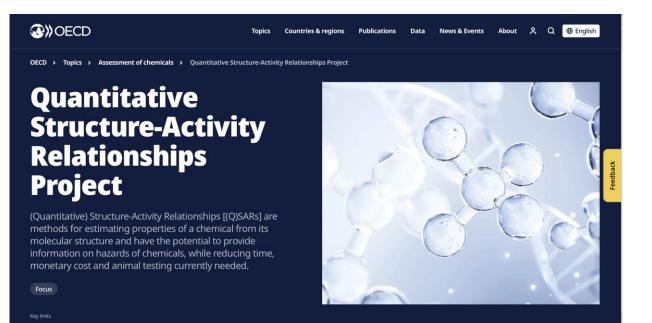
In Pictures

Also in the news

Health

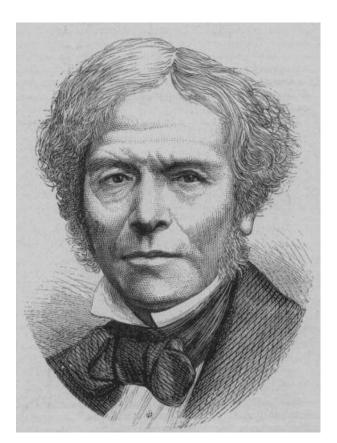
OECD Principles of QSAR (ML) Modelling from 2004

- Add a Principle '0' Characterize the data
- ... the original OECD principles did not call out a specific principle to capture the importance of data aggregation and curation prior to modelling,
- Transparency in Modeling through Careful Application of OECD's QSAR/QSPR Principles via a Curated Water Solubility Data Set,
- C. N. Lowe, *et al*, Chemical Research in Toxicology 2023 36 (3), 465-478



ON BEING A SCIENTIST

A GUIDE TO RESPONSIBLE CONDUCT IN RESEARCH



Faraday's laboratory notebooks are also remarkable in the amount of detail that they give about the design and setting up of experiments, interspersed with comments about their outcome and thoughts of a more philosophical kind. All are couched in plain language, with many vivid phrases of delightful spontaneity....

Peter Day, 'The Philosopher's Tree: A Selection of Michael Faraday's Writings'



Welcome

Welcome to the Davy Notebooks Project website. Our site – especially the Blog section – is updated frequently. Project updates will be posted to our social media feeds (Twitter feed embedded, and buttons, below). To join the project mailing list or to send any queries to the project team, please use our Contact page.

- The aim is for the IUPAC standards to facilitate the smooth interaction of people, computers and equipment on a global scale.
- Producers and Consumers of data must not be like the noncommunicating paths of a double spiral staircase (DNA)



University of Reading, Department of Chemistry

Chemistry is a Social Machine

Can Scientific Discovery Be Automated?

Progress in the sciences can only move as fast as humans can think outsourcing to A.I. could change that.

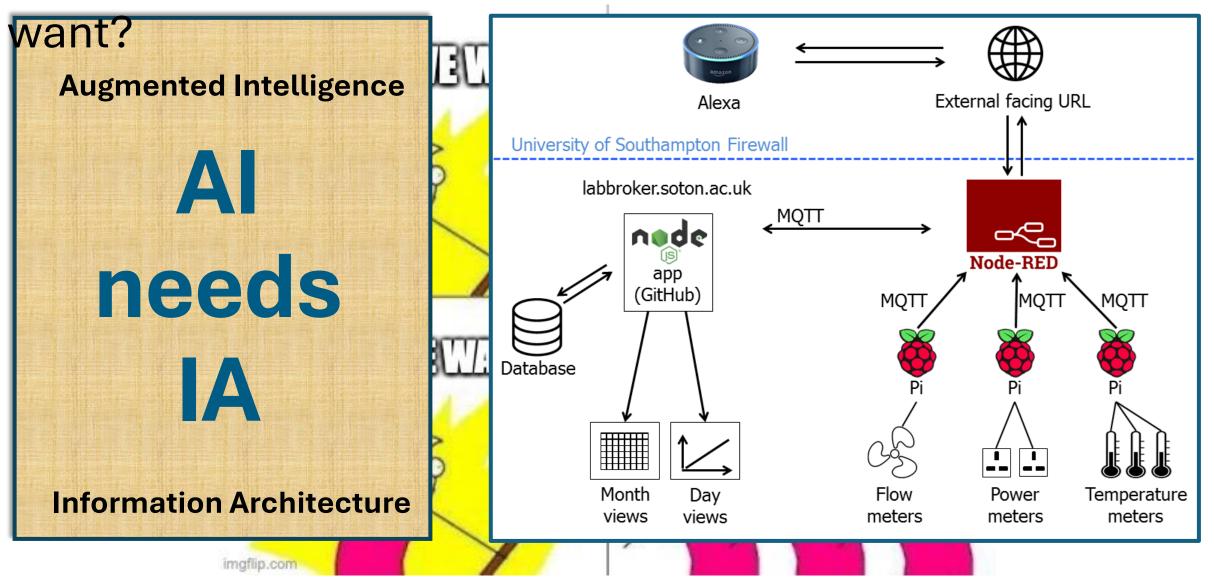
AHMED ALKHATEEB APR 25, 2017

SCIENCE

The twin challenges of too much quantity and too little quality are rooted in the finite neurological capacity of the human mind. Scientists are deriving hypotheses from a smaller and smaller fraction of our collective knowledge and consequently, more and more, asking the wrong questions, or asking ones that have already been answered. Also, human creativity seems to depend increasingly on the stochasticity of previous experiences-particular life events that allow a researcher to notice something others do not. Although chance has always been a factor in scientific discovery, it is currently playing a much larger role than it should.

ChemTalk Basel 2024

What do we need to do to get the data we





25th Anniversary of MQTT

By Jo Twist BBC News science and technology reporter

A blend of mobile technology and award-winning software is letting scientists finally escape the lab.

The software, called "middleware", lets different computer systems talk to each other securely and instantaneously.

As part of a national e-Science Video and Audio project in the UK, it is being Have Your Say used to let Southampton University chemists monitor Magazine experiment conditions from In Pictures mobiles. Country Profiles Special Reports

RELATED BBC SITES SPORT WEATHER CBBC NEWSROUND ON THIS DAY EDITORS' BLOG

Scotland

Business

Politics

Health

Education

Science &

Environment

Technology

Entertainment

Also in the news

Wales

Sensors in the lab pick up any changes in the environment so The system is not smart enough to the system can alert chemists, actually buy a round yet wherever they are.

"It replaces the traditional notebook with some electronic form, but it is not just about carrying around a computer - it is much smarter than that," Dr Jeremy Frey, from Southampton University, told the BBC New website.

"We wanted to be able to monitor and keep tabs on experiments outside the labs," he explained.

Friday, 4 February, 2005, 08:47 GMT

SEE ALSO: Top UK innovation prize for IBM 11 Jun 04 | Technoloay Boom times ahead for smartphones 27 Aug 04 | Technology E-mail provides Arctic lifeline 09 May 04 | Technology

Heart treatment from a distance 11 Aug 04 | Health

RELATED INTERNET LINKS: Dr Frey's group University of Southampton Combechem (e-Science project) IBM Hursely Labs EPRSC

The BBC is not responsible for the content of external internet sites

TOP SCIENCE & ENVIRONMENT STORIES

Night-sky image is biggest ever Phantom Eye 'spy plane' unveiled Higgs discovery rumour is denied News feeds

It could also realise the promise of controlling home environments from phones easily and seamlessly.

But there are many other potential applications for the monitoring jobs done by the power industries, healthcare professions and other labs.

IBM won the Royal Academy of Engineering's MacRobert prize which rewards technological and engineering innovation for the program in June last year.



Chemists enjoy a drink at the bar while keeping and eye on the lab

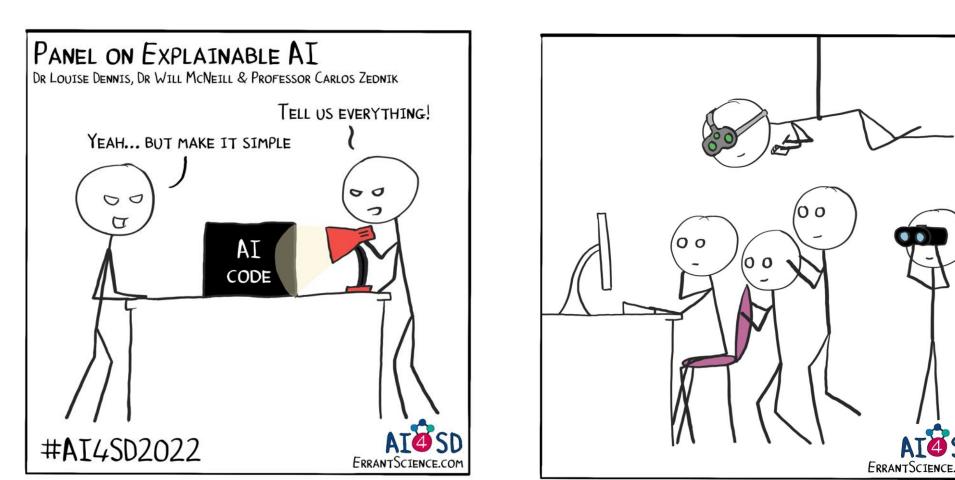
Used by top global banks, the WebSphere MQ family is a decade old.

It has transformed e-commerce because of its ability to allow data transfers across computer systems and different platforms without extra coding.

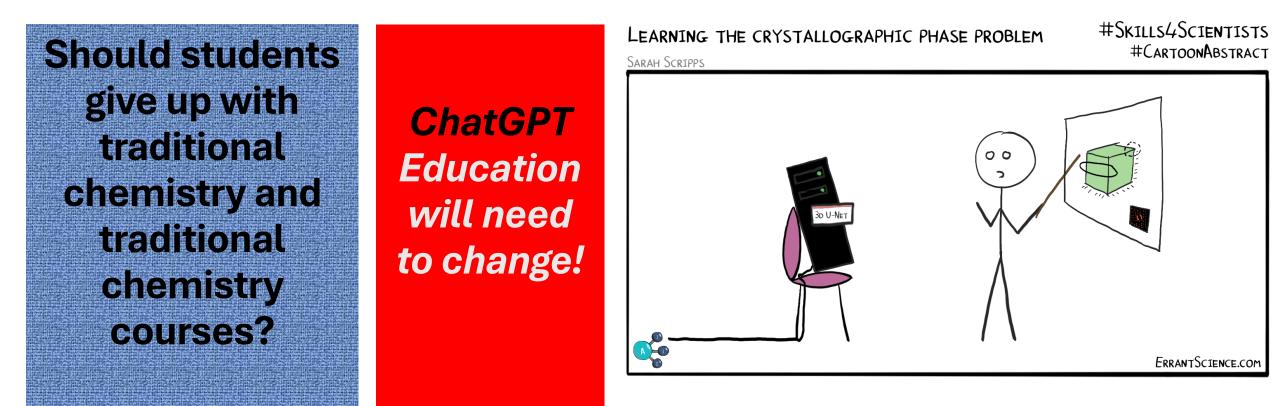
25/09/2024

ChemTalk Basel 2024

Ethical & Explainable Al Needed for scientific discovery



Derek Lowe In the Pipeline Chemistry World ... Do you know who will find these things out? Not our AI and ML systems, although I'm sure they'll help whenever possible. No, it is going to be us. Just like it always has been. The law of conservation of data



Limits of Chemical Computation

nature

Explore content \checkmark Journal information \checkmark Publish with us \checkmark

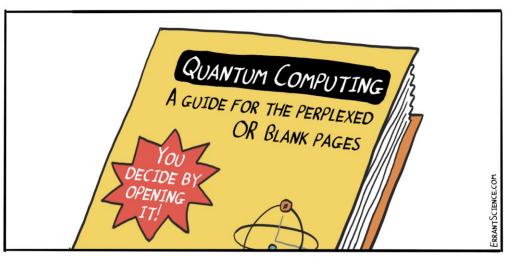
nature > articles > article

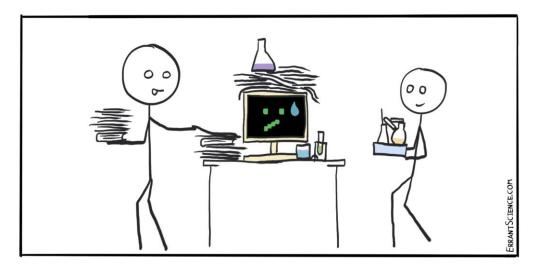
Published: 09 December 2015 Undecidability of the spectral gap

Toby S. Cubitt 🖂, David Perez-Garcia & Michael M. Wolf

Nature 528, 207–211 (2015) | Cite this article

There are computationally undecidable problems in quantum chemistry!



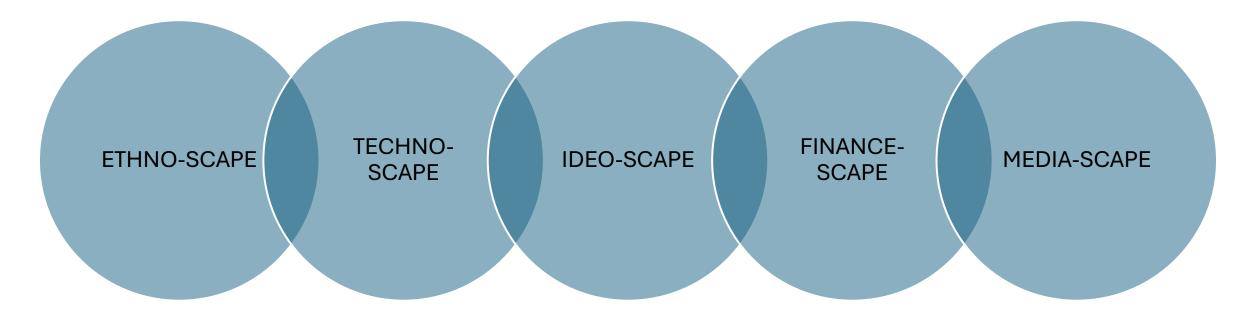




Data & Decisions

Data Should be FAIR	Decisions should be Intelligently Open					
Findable	Accessible					
Accessible	Comprehensible Useable					
Interoperable						
Reusable	Assessable					
Barend Mons etal 2016	Onora O'Neil 2002					

Appadurai's 5-scapes for globalisation: Global (Chemical) Cultural Flow



Chemists, Cooperation, Equipment, Society, Industry, Funding, Dissemination, Data

Limits of Chemical Computation

784 8 T.L. 284 828 D.C.

nature

Explore content V Journal information V Publish with us V

nature > articles > article

Published: 09 December 2015 Undecidability of the spectral gap

Toby S. Cubitt 🖂, David Perez-Garcia & Michael M. Wolf

Nature 528, 207-211 (2015) | Cite this article

There are computationally undecidable problems in quantum chemistry!

ChemTalk Basel 2024

11

Godel

Turing

We are still in a Liminal period



25/09/2024

Dissemination is part of the research But beware of "Engrooved behaviour"

concept highlighted in teaching by Paul Trowler

http://www.brad.ac.uk/sustainableuniversities/media/sustainableuniversities/Keyn ote-Prof-Paul-Trowler.pdf



http://www.todayandtomorrow.net/2010/02/22/quarter-mile-groove/

The Future (is ours to see)

When chemistry becomes a discipline, mathematical chemists will design new materials, predict their properties, and tell engineers how to make them — without ever entering a laboratory. We've got a long way to go on that one!

Robert A. Heinlein, "Where to?" 1950



Thank you for listening

and all the brilliant members of my research group and my excellent colleagues, that made this work



Digital chemistry: Paradigm shift, cultural evolution or business as usual?

Jeremy Frey University of Southampton Head Computational Systems Chemistry