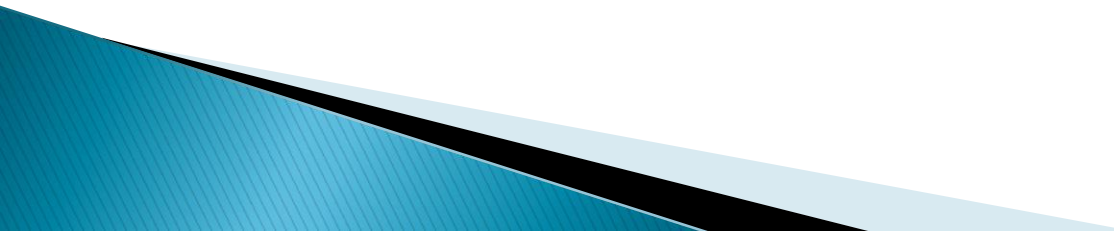


Komplex valószínűségi modellek, következtetésre, tanulásra, adatfúzióra

Bolgár Bence, Antal Péter

antal@mit.bme.hu

Áttekintés

- ▶ Információk a kurzusról
 - ▶ Az adat- és tudásfúziós kihívás
 - ▶ Racionális döntéstámogatás: „bayesi” döntéselmélet
 - Valószínűségszámítás, valószínűség értelmezései
 - Hasznosságelmélet, preferenciák
 - Bayes-tanulás
- 

Alapinformációk

▶ Honlap

- <https://www.mit.bme.hu/oktatas/targyak/VIMIIV20>
(feltöltés alatt)

▶ Tárgyfelelős és oktató

- Bolgár Bence (~bence), bolgar@mit.bme.hu
- Antal Péter (~peter), antal@mit.bme.hu

▶ Időpont, helyszín

- Kedd, 12.15–13.45, IL.405
- Csütörtök, 12.15–13.45, IL.405

Tankönyvek, jegyzetek, segédanyagok

▶ Tankönyvek

- S. Russell and P. Norvig Artificial Intelligence: A Modern Approach Prentice Hall, Second Edition
 - MI Almanach:<https://mialmanach.mit.bme.hu/>
- P. Antal et al.: Valószínűségi döntéstámogató rendszerek, 2014
- P. Antal et al.: Bioinformatika, 2014

▶ Szoftverek

- BayesCube, szoftver+felhasználói kézikönyv
 - <http://bioinformatics.mit.bme.hu/>
- R
 - LearnBayes
- Edward
 - <http://edwardlib.org/>
 - Dustin Tran, Alp Kucukelbir, Adji B. Dieng, Maja Rudolph, Dawen Liang, and David M. Blei. 2016. *[Edward: A library for probabilistic modeling, inference, and criticism](#)*
 - Dustin Tran, Matthew D. Hoffman, Rif A. Saurous, Eugene Brevdo, Kevin Murphy, and David M. Blei. 2017. *[Deep Probabilistic Programming](#)*

Tankönyvek

- ▶ Valószínűségi gráfos modellek, Bayes-hálózatok
 - J. Pearl: „**Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference**”, 1988
 - Cowell, R.G. – Dawid, A.P. – Lauritzen, S.L. – Spiegelhalter, D.J.: „**Probabilistic Expert Systems**”, 1999
 - Friedman, N. – Koller, D.: **Probabilistic Graphical Models**, The MIT Press, 2009
 - Denis, Jean-Baptiste_ Scutari, Marco: **Bayesian Networks With Examples in R**, CRC Press, (2014)
 - Søren Højsgaard, David Edwards, Steffen Lauritzen: **Graphical Models with R**, Springer, (2012)
- ▶ David Bellot: **Learning Probabilistic Graphical Models in R**, –Packt Publishing (2016)
- ▶ Valószínűségi következtetések és bayesi döntéelmélet
 - A.Gelman, J.B.Carlin, H.S.Stern, D.B.Dunson, A.Vehtari, D.B. Rubin: **Bayesian Data Analysis**, Chapman and Hall_CRC, 2014
 - J.Albert: **Bayesian Computation with R**, 2009
 - Jean-Michel Marin, Christian P. Robert: **Bayesian Essentials with R**, Springer-Verlag New York (2014)
 - Richard McElreath: **Statistical Rethinking_ A Bayesian Course with Examples in R and Stan**, Chapman and Hall_CRC (2015)
- ▶ Valószínűségi mintafelismerés, gépi tanulás és mesterséges intelligencia
 - Devroye, Luc, László Györfi, and Gábor Lugosi. **A probabilistic theory of pattern recognition**, Springer, 2013
 - **Christopher M. Bishop–Pattern Recognition and Machine Learning**, Springer (2007)
 - **Kevin P. Murphy: Machine Learning_ A Probabilistic Perspective**, The MIT Press (2012)
 - Kevin B. Korb, Ann E. Nicholson: **Bayesian Artificial Intelligence**, Second Edition–CRC Press, 2010
- ▶ Információelmélet, statisztikai következtetés
 - Cover, Thomas M., and Joy A. Thomas: **Elements of information theory**, John Wiley & Sons, 2012.
 - T.Hastie, R.Tibshirani, J.Friedman: **The Elements of Statistical Learning_ Data Mining, Inference, and Prediction**, 2008

Témák

1. Statisztikai paradigmák, a Bayes–statisztikai paradigma. Bayes–tanulási alapfogalmak, bayesi modell átlagolás. Naive Bayes–háló.
2. Adattípusok: megfigyelési és beavatkozási adatok. Prediktív, generatív, oksági modellek. Következtetés és indukció típusai.
3. Bayesi következtetés analitikus megoldással. A konjugáltság és az exponenciális eloszlások. Bayesi lineáris regresszió.
4. Közelítő módszerek bayesi következtetéshez. Bayesi centrális határeloszlási tétel. Laplace approximáció. Markov Chain Monte Carlo.
5. A variációs bayesi megközelítés.
6. Bayesi logisztikus regresszió/többrétegű perceptron/neurális hálók.
7. Bayesi mátrix faktorizáció.
8. Bayesi neurális hálózatok és bayesi mély struktúrák. I.
9. Esettanulmány: edward I.
10. Bayesi neurális hálózatok és bayesi mély struktúrák. II. GAN
11. Bayesi neurális hálózatok és bayesi mély struktúrák. III. Variational autoencoder.
12. Esettanulmány: edward II.
13. Rekurrens neurális hálózatok (RNN), időszori adatok elemzése.
14. Valószínűségi gráfos modellek (VGM–ek). Függetlenségi modellek.
15. Markov–(véletlen)–mezők. Bayes–hálók. Faktor–gráfok.
16. Rejtett Markov Modellek. Dinamikus Bayes–hálók.
17. Oksági diagrammok.
18. Bayesi becslés– és döntéelmélet. Optimális döntés fogalma, Bayes–faktor, Bayes–döntés, Bayes–hiba. Döntési hálók.
19. Egzakt következtetési módszerek VGM–ekben.
20. Következtetés közelítő módszerekkel VGM–ekben: Az EM algoritmus család. „loopy belief propagation”, „expectation propagation”.
21. Bayes–hálók tanulása
22. Tudástranszfer NN–ekbe és PGM–ekbe, „transfer learning”.
23. Adatvédelmet biztosító elosztott adat– és tudásfúzió (AwE/MA)
24. Aktív tanulás. K–karú rabló. Monte Carlo Tree Search.
25. Megerősítéses tanulás. Mély megerősítéses tanulás.
26. Kiterjesztett sztochasztikus szimulációs eljárások: adaptive és hibrid MCMC módszerek

Követelmények

- ▶ *Szorgalmi időszakban:*
Házi feladat sikeres elkészítése és leadása a félév végéig, amely egy tanulási algoritmus implementálását és egy referencia adathalmazon történő szabványos kiértékelését jelenti.
- ▶ *Vizsgaidőszakban:* Szóbeli vizsga. A vizsgára bocsátás feltétele az elfogadott házi feladat.
- ▶ *Osztályozás:* A vizsga osztályzata a szóbeli vizsgán megszerzett jegy.

Hallmarks of a new AI *era*?

ARTICLE

doi:10.1038/nature16961

Mastering the game of Go with deep neural networks and tree search

David Silver^{1*}, Aja Huang^{1*}, Chris J. Maddison¹, Arthur Guez¹, Laurent Sifre¹, George van den Driessche¹, Julian Schrittwieser¹, Ioannis Antonoglou¹, Veda Panneershelvam¹, Marc Lanctot¹, Sander Dieleman¹, Dominik Grewe¹, John Nham², Nal Kalchbrenner¹, Ilya Sutskever², Timothy Lillicrap¹, Madeleine Leach¹, Koray Kavukcuoglu¹, Thore Graepel¹ & Demis Hassabis¹

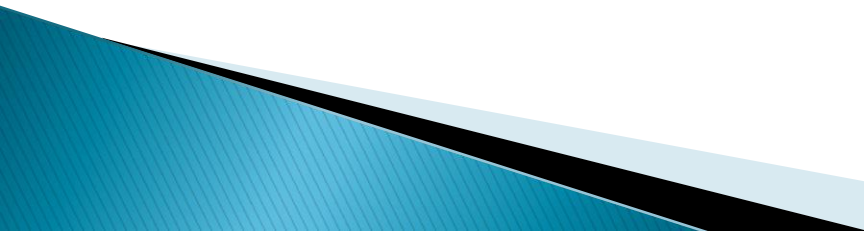
LETTER

doi:10.1038/nature14236

Human-level control through deep reinforcement learning

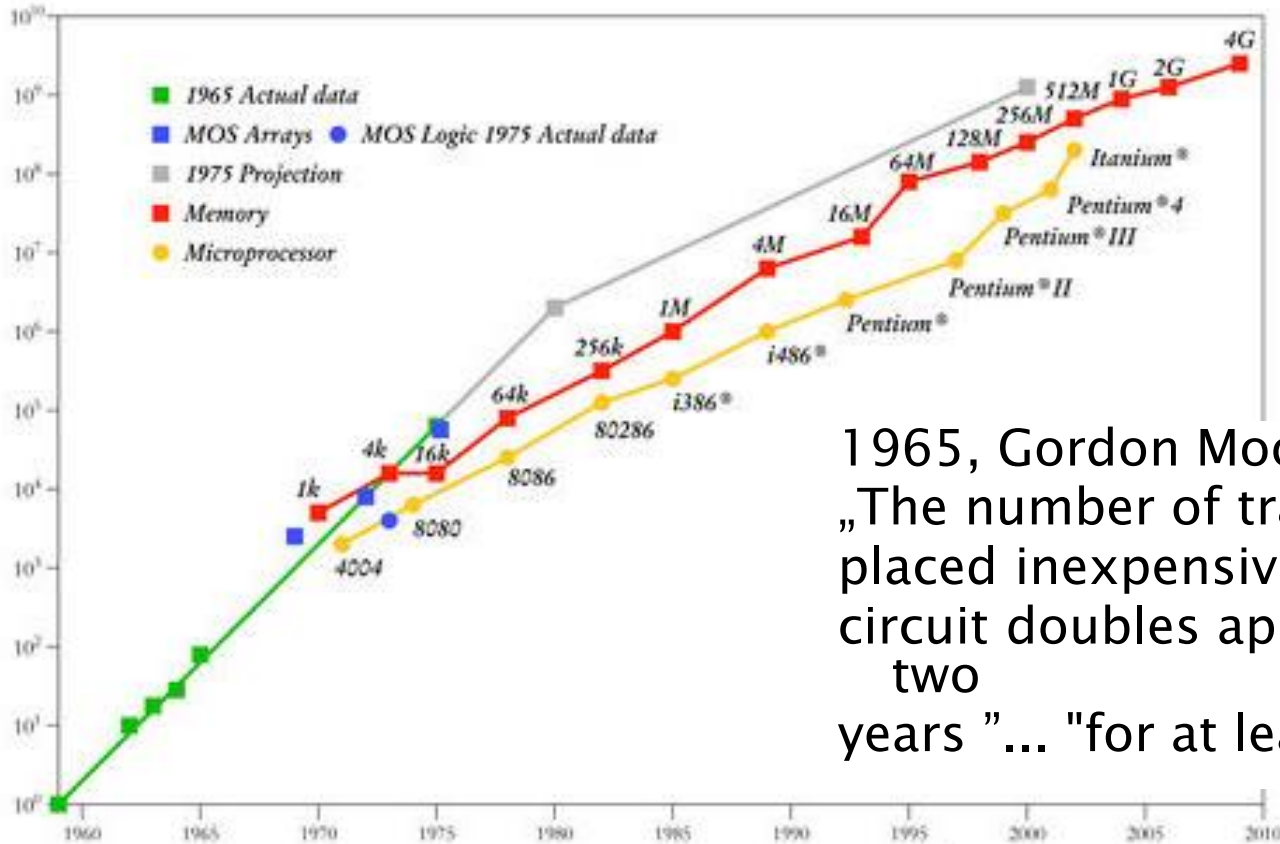
Volodymyr Mnih^{1*}, Koray Kavukcuoglu^{1*}, David Silver^{1*}, Andrei A. Rusu¹, Joel Veness¹, Marc G. Bellemare¹, Alex Graves¹, Martin Riedmiller¹, Andreas K. Fidjeland¹, Georg Ostrovski¹, Stig Petersen¹, Charles Beattie¹, Amir Sadik¹, Ioannis Antonoglou¹, Helen King¹, Dharshan Kumaran¹, Daan Wierstra¹, Shane Legg¹ & Demis Hassabis¹

Factors behind the „A.I./learning hype”

- ▶ New theory?
 - Unified theory of AI?
 - A new machine learning approach?
 - A breakthrough result?
 - ▶ New hardware? (computing power..)
 - GPUs?
 - Quantum computers?
 - ▶ New resources?
 - Data!
 - Knowledge!
 - ▶ Technologies
 - Artificial intelligence? Language understanding?
 - Machine learning? Deep learning?
- 

Computing power: Moore's Law

Transistors
Per Die

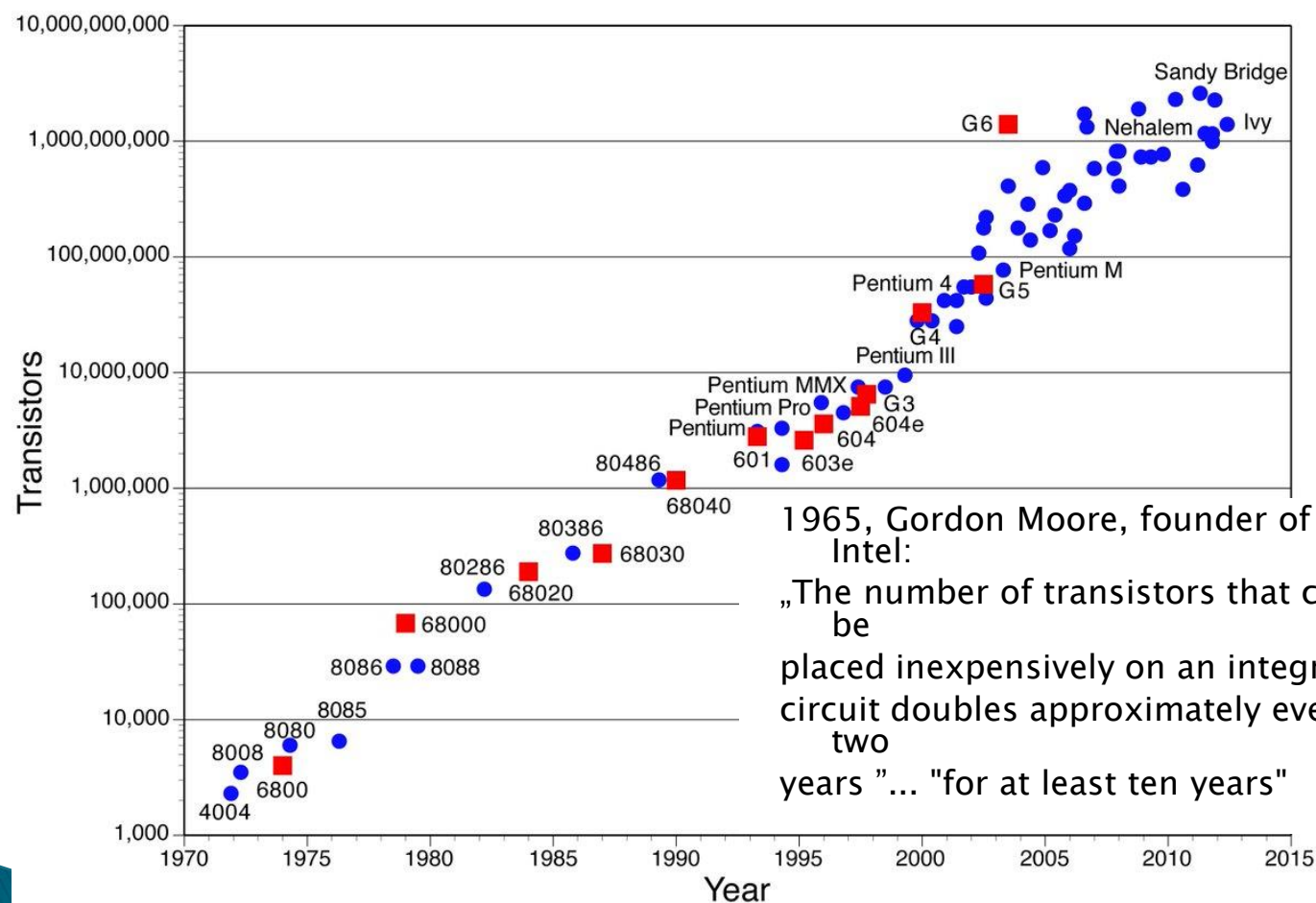


Integration and parallelization wont bring us further. End of Moore's law?

1965, Gordon Moore, founder of Intel:
„The number of transistors that can be placed inexpensively on an integrated circuit doubles approximately every two years “... “for at least ten years”

SCIENCEPHOTOLIBRARY

Computing power: Moore's Law

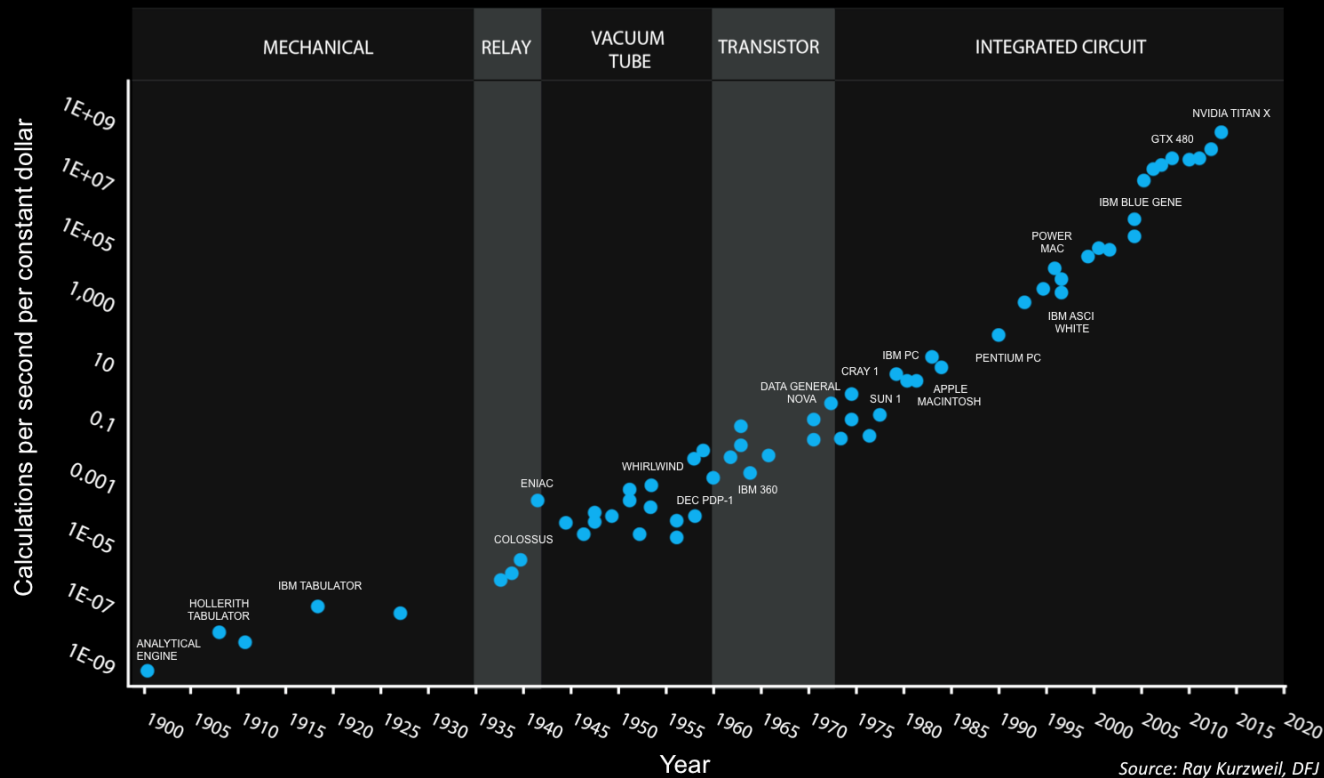


- 10 μm – 1971
- 6 μm – 1974
- 3 μm – 1977
- 1.5 μm – 1982
- 1 μm – 1985
- 800 nm – 1989
- 600 nm – 1994
- 350 nm – 1995
- 250 nm – 1997
- 180 nm – 1999
- 130 nm – 2001
- 90 nm – 2004
- 65 nm – 2006
- 45 nm – 2008
- 32 nm – 2010
- 22 nm – 2012
- 14 nm – 2014
- 10 nm – 2017
- 7 nm – ~2019
- 5 nm – ~2021

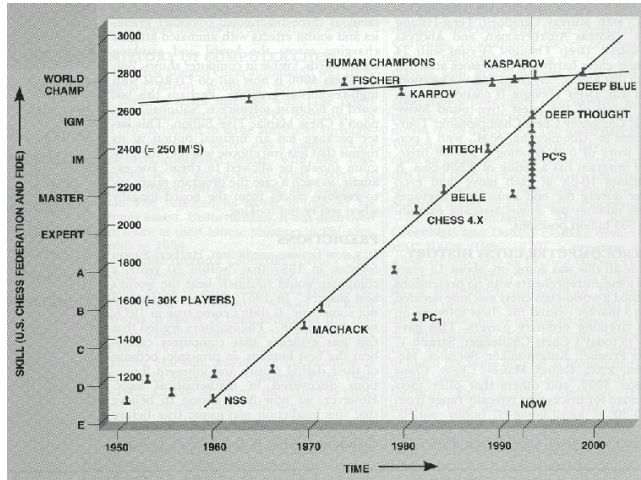
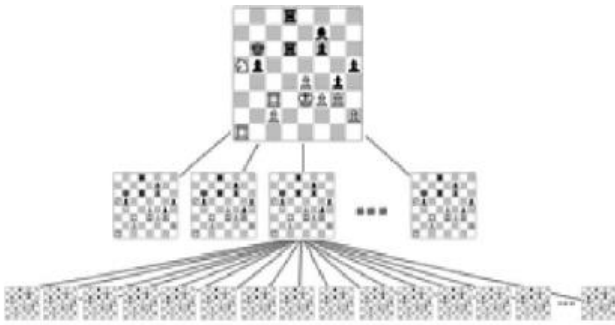
2012: single atom transistor (~0.1n, 1A)

Moore's law: calculation/\$

120 Years of Moore's Law



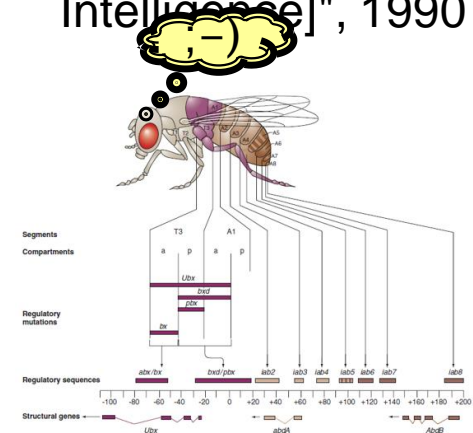
Computing power and search: performance in chess



#	Név	Élőpont
1	SugaR XPrO 1.2 64-bit 4CPU	3415
2	Komodo 11.2 64-bit 4CPU	3402
3	Houdini 5.01 64-bit 4CPU	3382
	IBM Deep Blue (1997)	-



J.McCarthy: "Chess as the Drosophila of AI. [Artificial Intelligence]", 1990

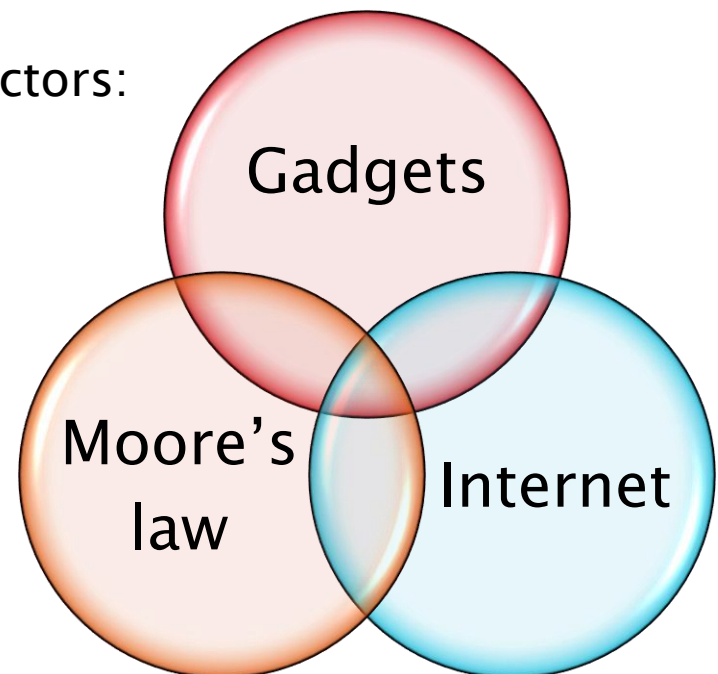


The „big” data

- ▶ Financial transaction data, mobile phone data, user (click) data, e-mail data, internet search data, social network data, sensor networks, ambient assisted living, intelligent home, wearable electronics,...

“The line between the virtual world of computing and our physical, organic world is blurring.” E.Dumbill: Making sense of big data, Big Data, vol.1, no.1, 2013

Factors:



Definitions of „big data”

M. Cox and D. Ellsworth, “Managing **Big Data** for Scientific Visualization,” Proc. ACM Siggraph, ACM, 1997

The 3xV: volume, variety, and velocity (2001).

The 8xV: Vast, Volumes of Vigorously, Verified, Vexingly Variable Verbose yet Valuable Visualized high Velocity Data (2013)

Not „conventional” data: „Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn’t fit the strictures of your database architectures. To gain value from this data, you must choose an alternative way to process it (E.Dumbill: Making sense of big data, Big Data, vol.1, no.1, 2013)

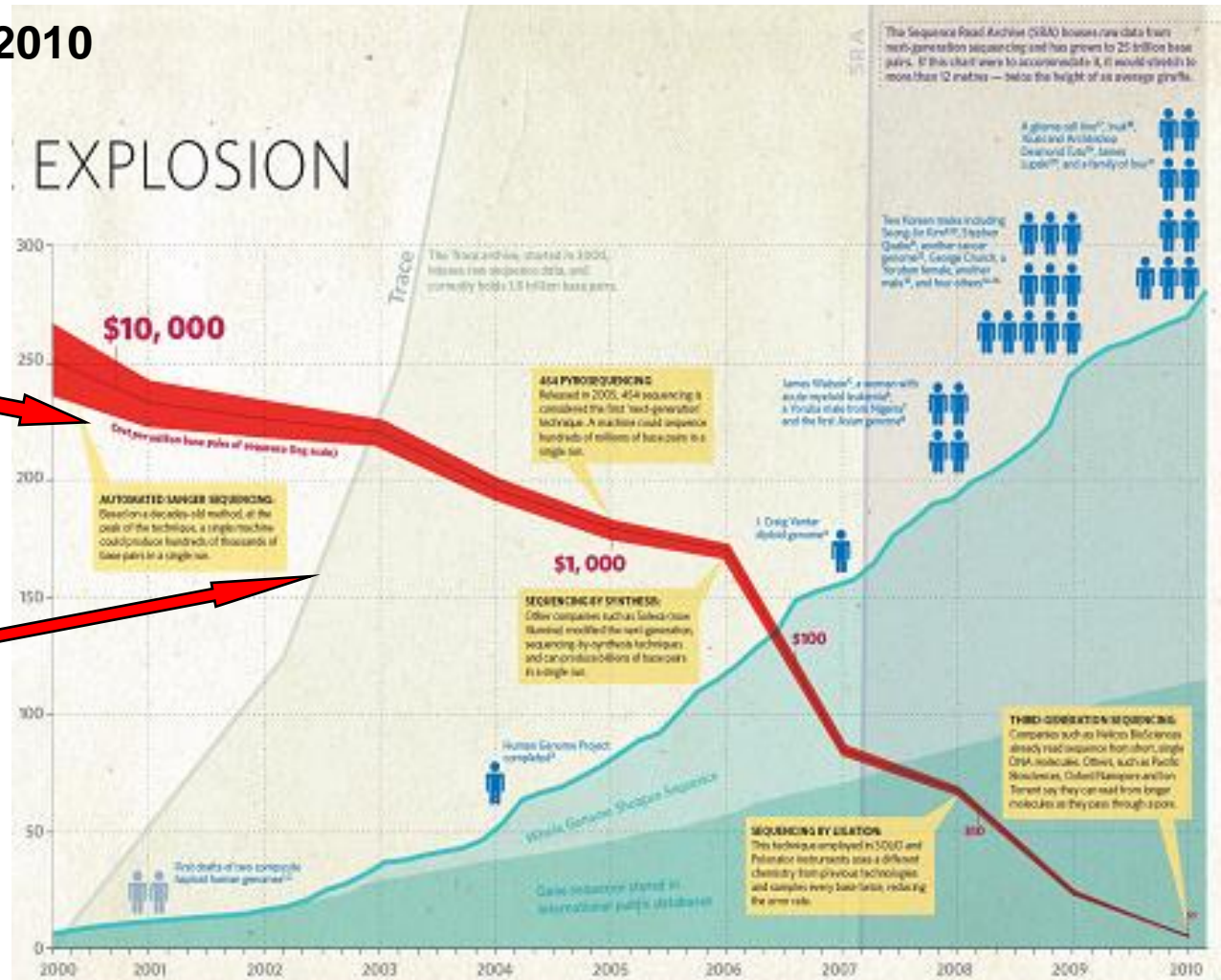
Carlson's Law for Biological Data

NATURE, Vol 464, April 2010

Sequencing costs per mill. base

Publicly available genetic data

- x10 every 2-3 years
- Data volumes and complexity that IT has never faced before...



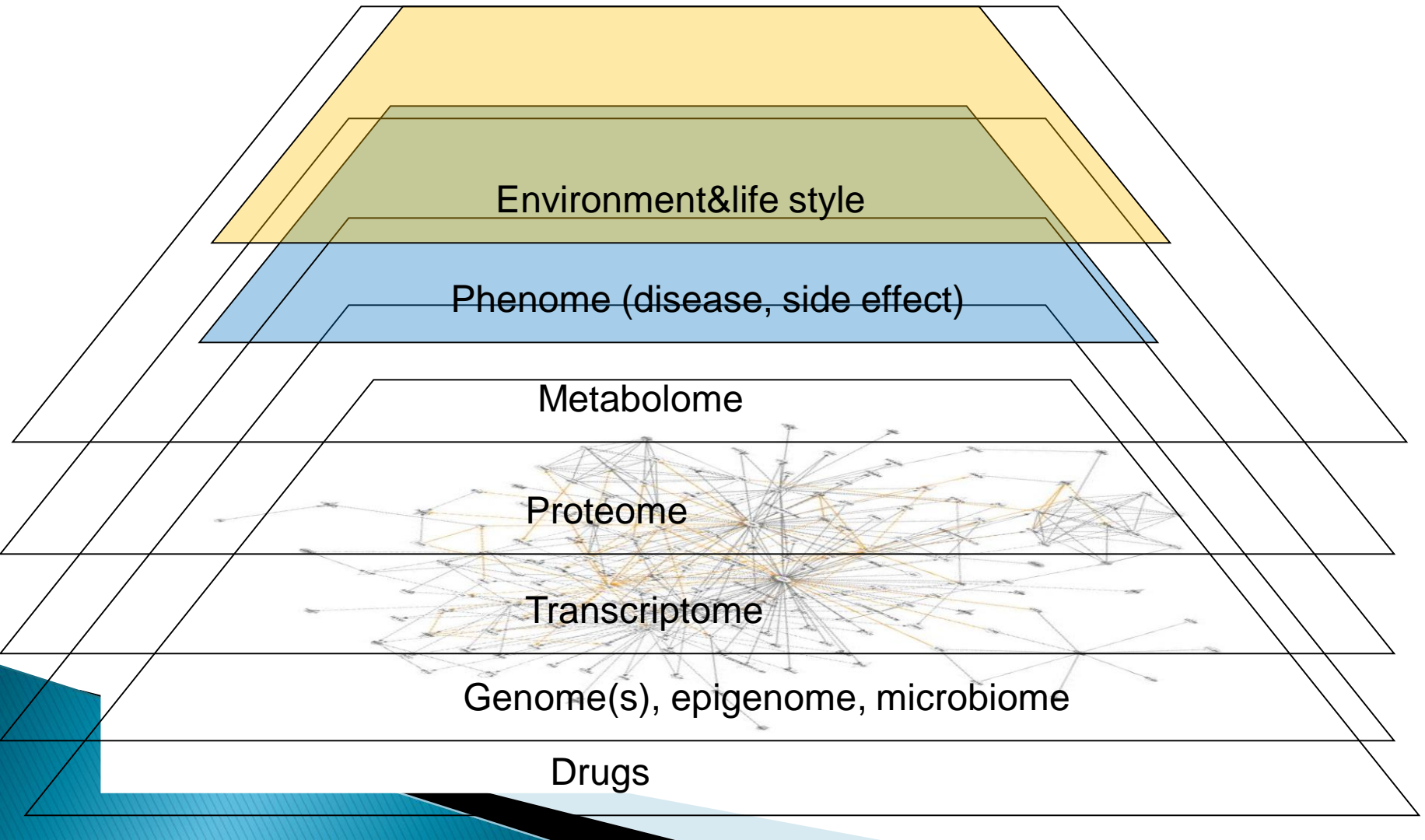
The „omic” definition of „big data”

.. [data] is often big in relation to the phenomenon that we are trying to record and understand. So, if we are only looking at 64,000 data points, but **that represents the totality or the universe of observations. That is what qualifies as big data. You do not have to have a hypothesis in advance before you collect your data. You have collected all there is—all the data there is about a phenomenon.**



Biomedical omic data/big data

2010<: “Clinical phenotypic assay”/drugome: open clinical trials, adverse drug reaction DBs, adaptive licensing, Large/scale cohort studies (~100,000 samples)



Big health data streams

New "Omics" Data Streams	Traditional Data Streams	Quantified Self Data Streams
Genome -SNP mutations ✓ -Structural variation -Epigenetics	Personal and Family Health History ✓	Self-reported data: health, exercise, food, mood journals, etc. ✓
Microbiome ✓	Prescription History ✓	Mobile Application Data ✓
Transcriptome	Lab Tests: History and Current ✓	Quantified Self Device Data ✓
Metabolome	Demographic Data ✓	Biosensor Data Objective Metrics
Proteome	Standardized Instrument Response ✓	
Diseasome ✓		
Environmentome ✓		
Legend: Consumer-available ✓		

M.Swan: THE QUANTIFIED SELF: Fundamental Disruption in Big Data Science and Biological Discovery, Big data, Vol 1., No. 2., 2013

UK Biobank



UK Biobank is a national and international health resource with unparalleled research opportunities, open to all bona fide health researchers. UK Biobank aims to improve the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses – including cancer, heart diseases, stroke, diabetes, arthritis, osteoporosis, eye disorders, depression and forms of dementia. **It is following the health and well-being of 500,000 volunteer participants and provides health information**, which does not identify them, to approved researchers in the UK and overseas, from academia and industry. Scientists, please ensure you read the [background materials](#) before registering. **To our participants, we say thank you for supporting this important resource to improve health.** Without you, none of the research featured on this website would be possible.

Elektronikus Egészségügyi Szolgáltatási Tér (EESZT)?



EESZT a betegellátásban

Szeretné tudni, hogy miként segítik az EESZT szolgáltatásai a betegellátást? Kovács úr történetén keresztül megismerheti azokat az előnyöket, amik a gyorsabb és hatékonyabb gyógyítást szolgálják....



Elektronikus Egészségügyi Szolgáltatási Tér (EESZT)

Az Elektronikus Egészségügyi Szolgáltatási Teret (EESZT) európai uniós támogatásból hozta létre az Állami Egészségügyi Ellátó Központ vezette konzorcium, hogy az egészségügyi szolgáltatók összekapcsolásával és a nekik létrehozott egységes kommunikációs tér biztosításával kiaknázza az e-egészségügyben rejlő lehetőségeket. Így az egészségügyi ellátók és ellátottak is szorosabb kapcsolatba kerülnek, leegyszerűsödik az ellátás folyamán keletkezett adatok és dokumentumok elérhetősége, biztosítottá válik az egészségügyi ágazat számára a rendelkezésre álló adatok egységes kezelése és hatékony elemzése, ami egyaránt segítség a betegnek és az orvosnak is.

<https://e-egeszsegugy.gov.hu/fooldal>

EESZT: MENTA



22



Egészségnaplók

Vezesse egy helyen összes fontos egészségügyi adatait! Feljegyzései többé sosem vesznek el.



Telemedicina

Össze meg egyetlen gombnyomással az Ön által otthon mért értékeket kezelőorvosával!



Egészségterv

Tűzze ki egészségével kapcsolatos céljait! Interaktív alkalmazásunk segít Önnek elérnie azokat.



Információs központ

Érdeklí a személyre szabott tájékoztatás? Kövesse naprakész híreinket rendszeresen!



Szolgáltatáskereső

Egészségügyi szolgáltatást venne igénybe? Gyors és hiteles információkat talál országszerte.



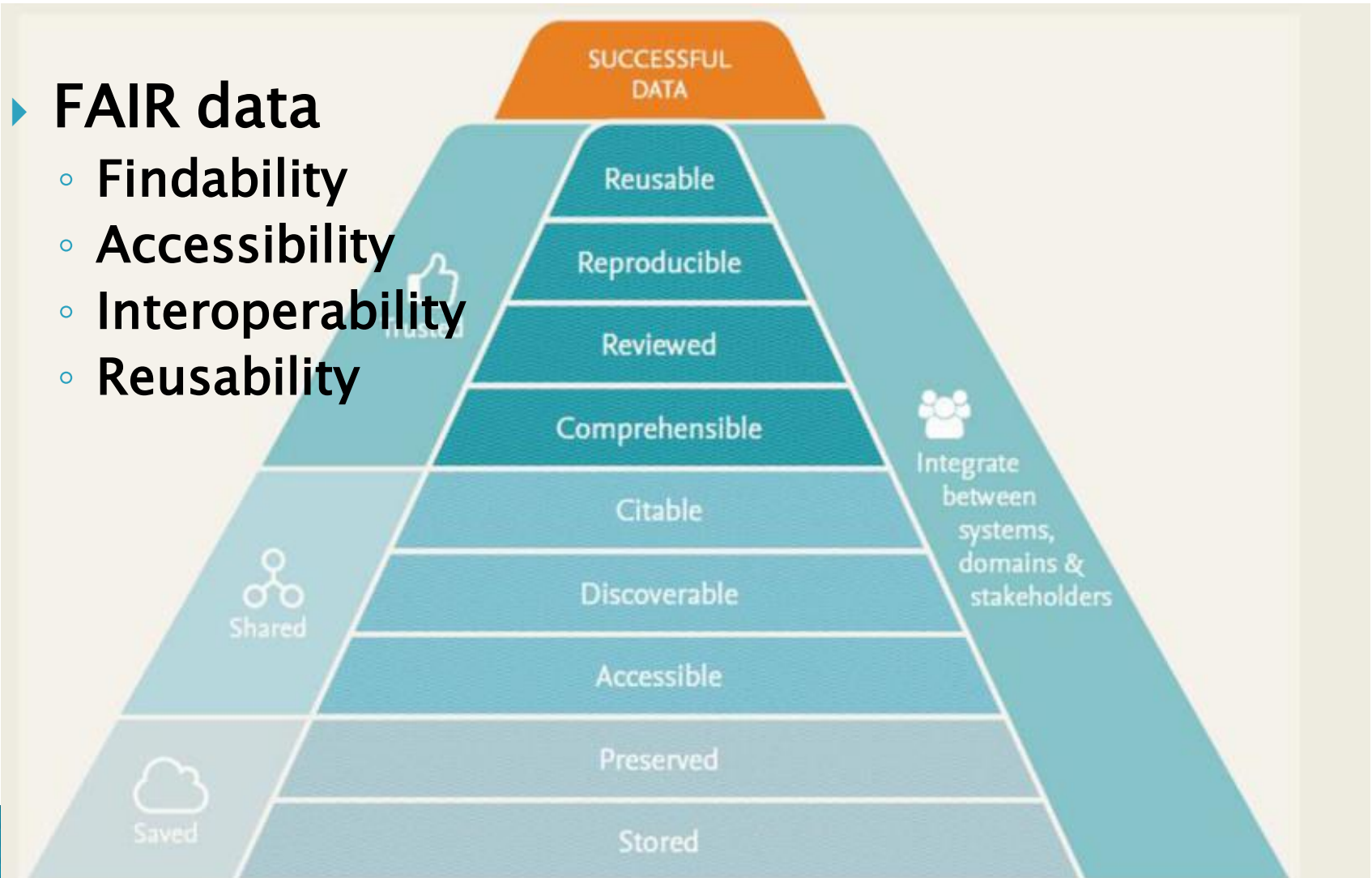
Fejlesztőknek

Csatlakozási lehetőséget biztosítunk az egészségügy területén zajló fejlesztéseknek.

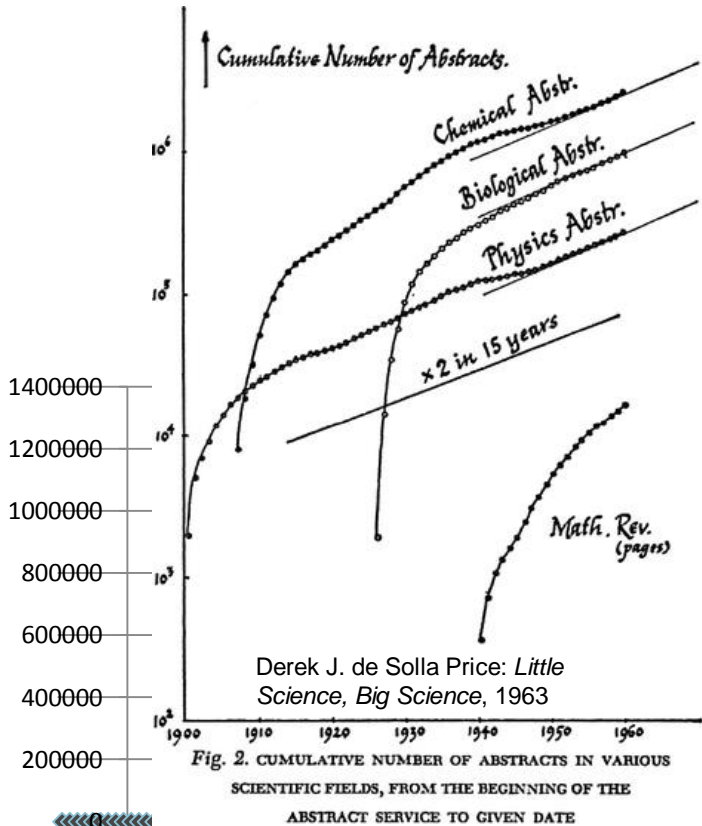
Nyílt adat

▶ FAIR data

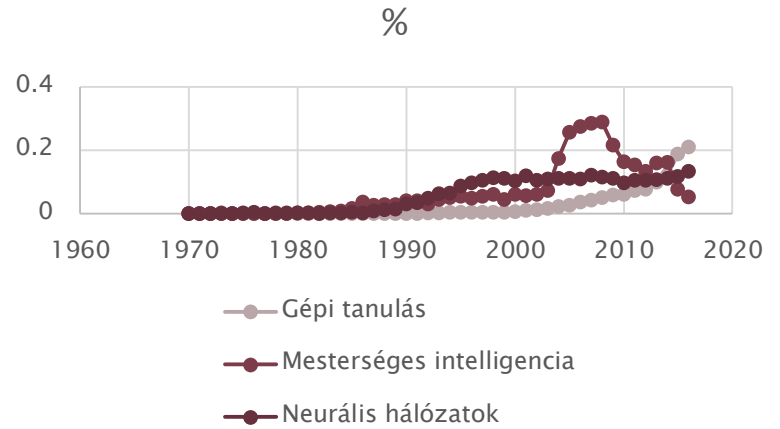
- Findability
- Accessibility
- Interoperability
- Reusability



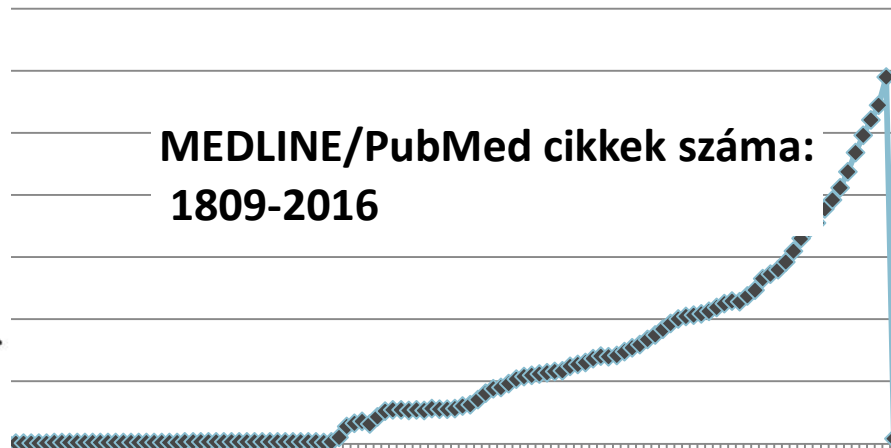
Írottan tárolt intelligencia: cikkek



It will be noted that after an initial period of rapid expansion to a stable growth rate, the number of abstracts increases exponentially, doubling in approximately 15 years.



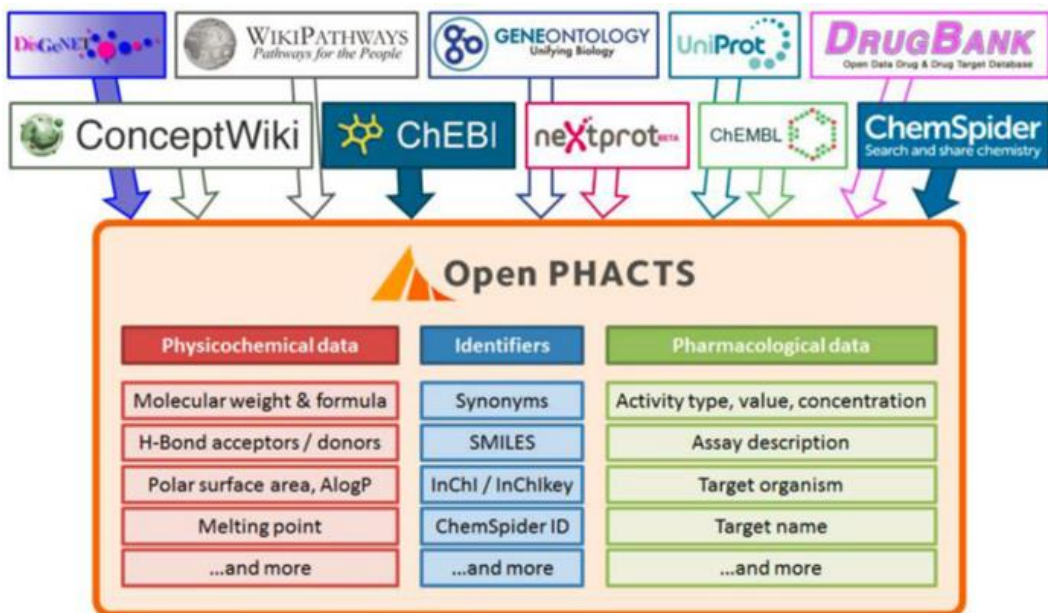
**MEDLINE/PubMed cikkek száma:
1809-2016**



1809 1816 1823 1830 1837 1844 1851 1858 1865 1872 1879 1886 1893 1899 1906 1913 1920 1927 1934 1941 1948 1955 1962 1969 1976 1983 1990 1997 2004 2011 2016

Nyílt tudás

Szemantikus adattárak - Szemantikus publikálás



Williams, Antony J., et al. "Open PHACTS: semantic interoperability for drug discovery." *Drug discovery today*, 2012

Dumontier, Michel, et al. "Bio2RDF release 3: a larger connected network of linked data for the life sciences, EUR-WS, 2014.

[OPENBEL:] Hofmann-Apitius, Martin, et al. "Towards the taxonomy of human disease." *Nature reviews. Drug discovery*, 2015

M. Gerstein, "E-publishing on the Web: Promises, pitfalls, and payoffs for bioinformatics," *Bioinformatics*, 1999

M. Gerstein: **Blurring the boundaries between scientific 'papers' and biological databases**, *Nature*, 2001

P. Bourne, "Will a biological database be different from a biological journal?," *Plos Computational Biology*, 2005

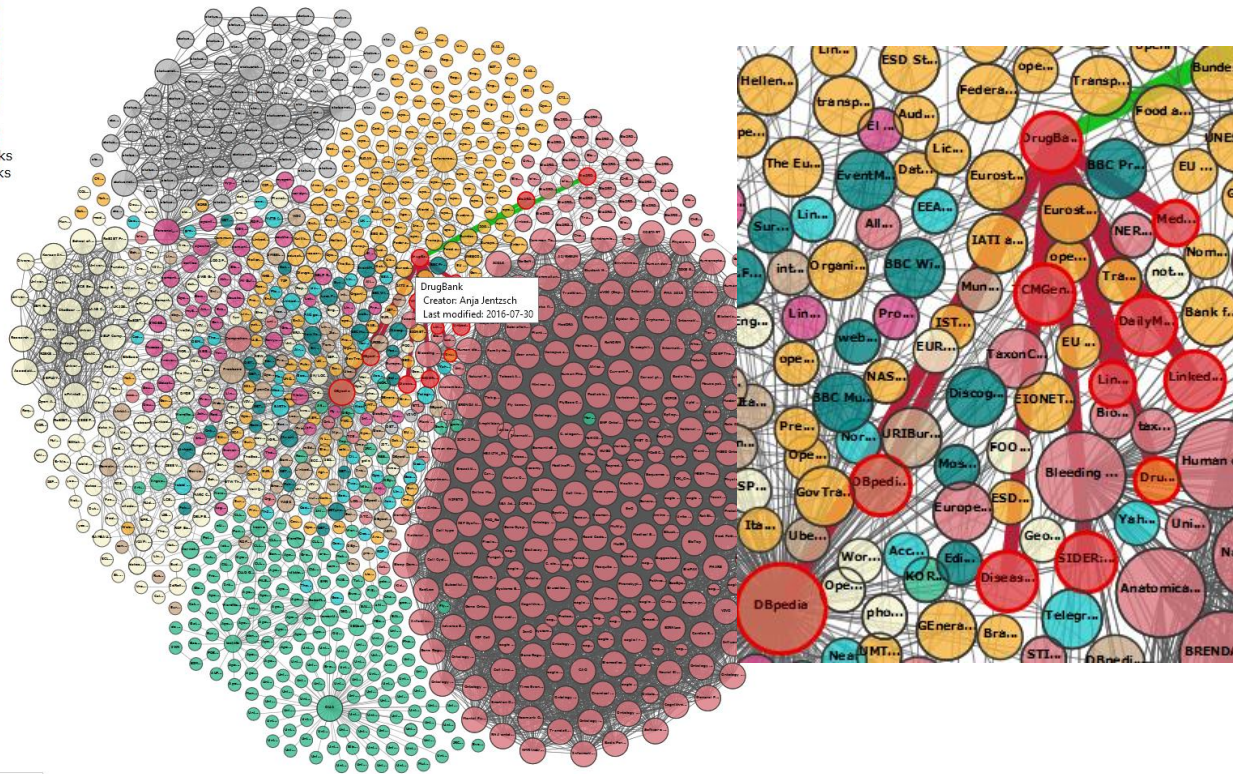
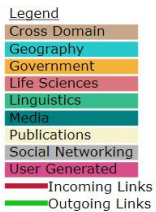
M. Gerstein et al: "**Structured digital abstract makes text mining easy**," *Nature*, 2007.

M. Seringhaus et al: "Publishing perishing? Towards tomorrow's information architecture," *Bmc Bioinformatics*, 2007.

M. Seringhaus: "Manually structured digital abstracts: A scaffold for automatic text mining," *Febs Letters*, 2008.

D. Shotton: "**Semantic publishing**: the coming revolution in scientific journal publishing," *Learned Publishing*, 2009

Összekapcsolt nyílt tudás (Linked Open Data)

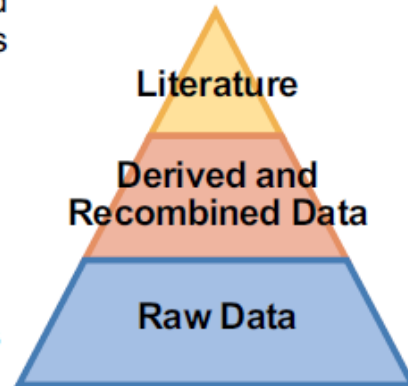


Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak. <http://lod-cloud.net/>

E-science, data-intensive science, the fourth paradigm

All Scientific Data Online

- Many disciplines overlap and use data from other sciences
- Internet can unify all literature and data
- Go from literature to computation to data back to literature
- Information at your fingertips for everyone-everywhere
- Increase Scientific Information Velocity
- Huge increase in Science Productivity

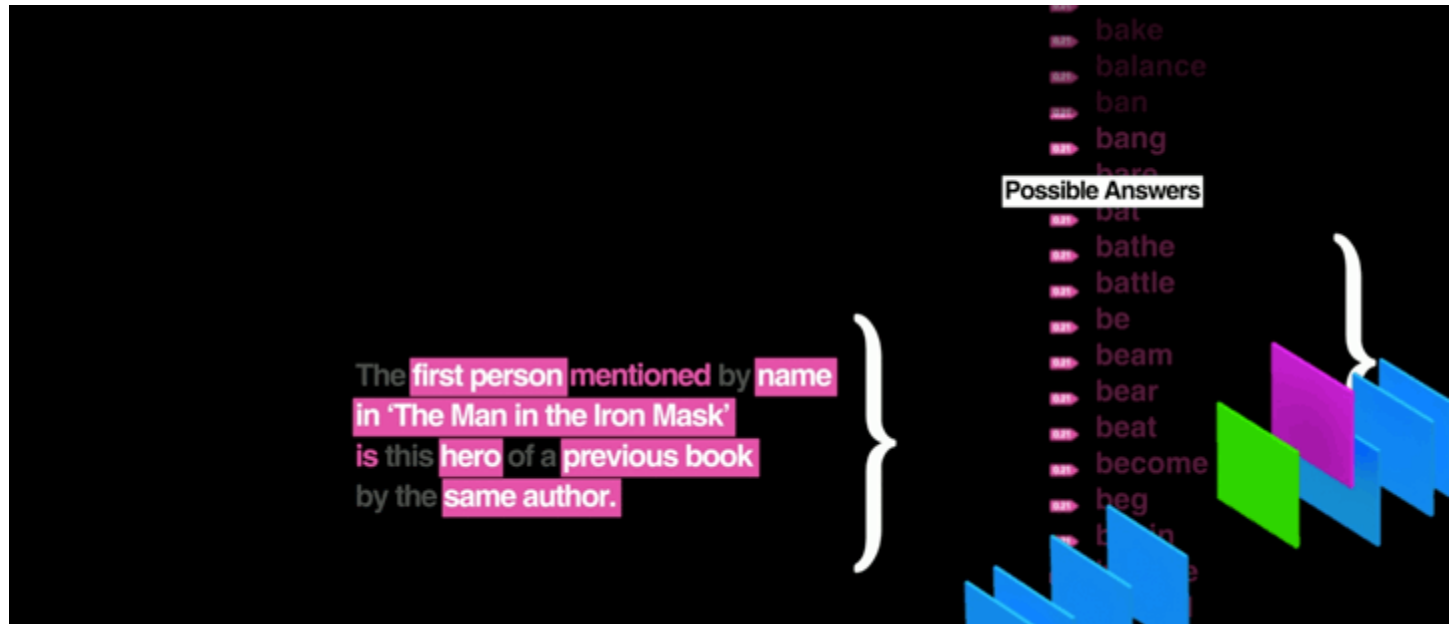


The FOURTH PARADIGM

DATA-INTENSIVE SCIENTIFIC DISCOVERY

EDITED BY TONY HEY, STEWART TANSLEY, AND KRISTIN TOLLE

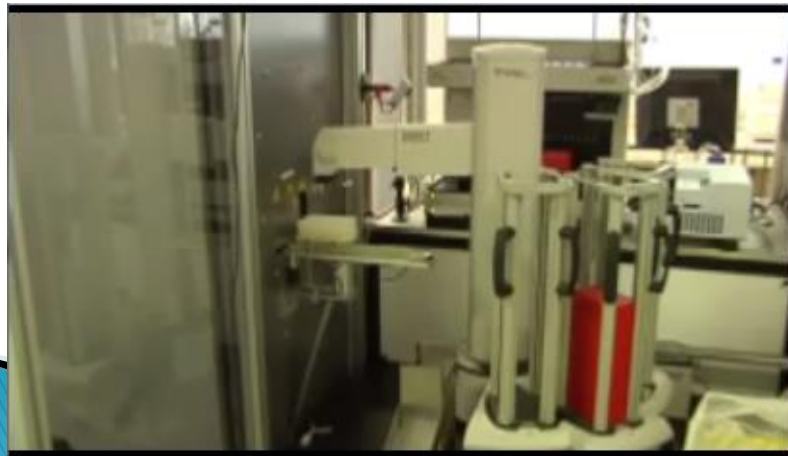
The Science Behind an Answer



- ▶ <http://www-03.ibm.com/innovation/us/watson/what-is-watson/science-behind-an-answer.html>

Gépi felfedező rendszerek

- Langley, P. (1978). Bacon: A general discovery system.
- ...
- ...
- R.D.King et al.: **The Automation of Science**, Science, 2009
- Sparkes, Andrew, et al.: Towards **Robot Scientists** for autonomous scientific discovery, 2010



„Ada”



„Eve”

„Machine science”

- Swanson, Don R. "Fish oil, Raynaud's syndrome, and undiscovered public knowledge." *Perspectives in biology and medicine* 30.1 (1986): 7-18.
- Smalheiser, Neil R., and Don R. Swanson. "Using **ARROWSMITH**: a computer-assisted approach to formulating and assessing scientific hypotheses." *Computer methods and programs in biomedicine* 57.3 (1998): 149-153.
- D. R. Swanson et al.: **An interactive system for finding complementary literatures: a stimulus to scientific discovery**, Artificial Intelligence, 1997



AB



BC

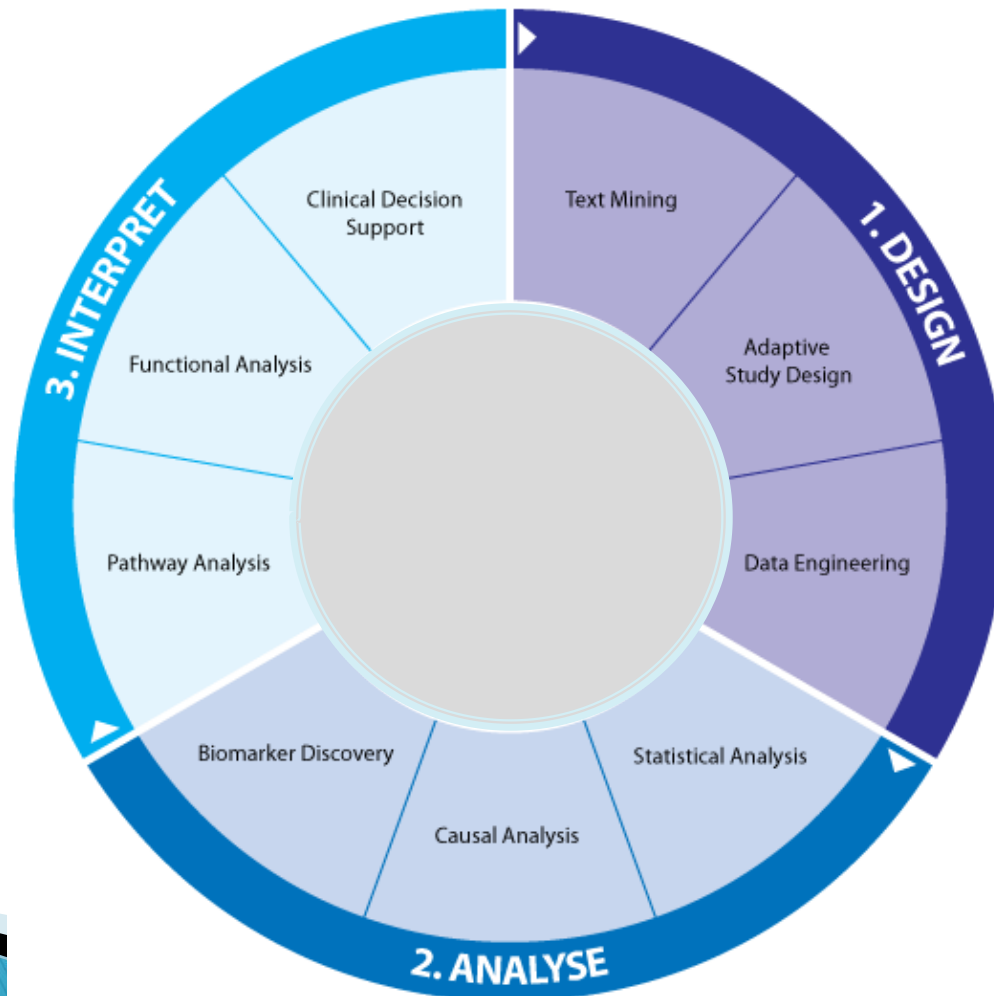


ABC

- James Evans and Andrey Rzhetsky: **Machine science**, Science, 2013

„Soon, computers could generate many useful hypotheses with little help from humans.”

A teljes tudományos kutatási ciklus automatizálása?



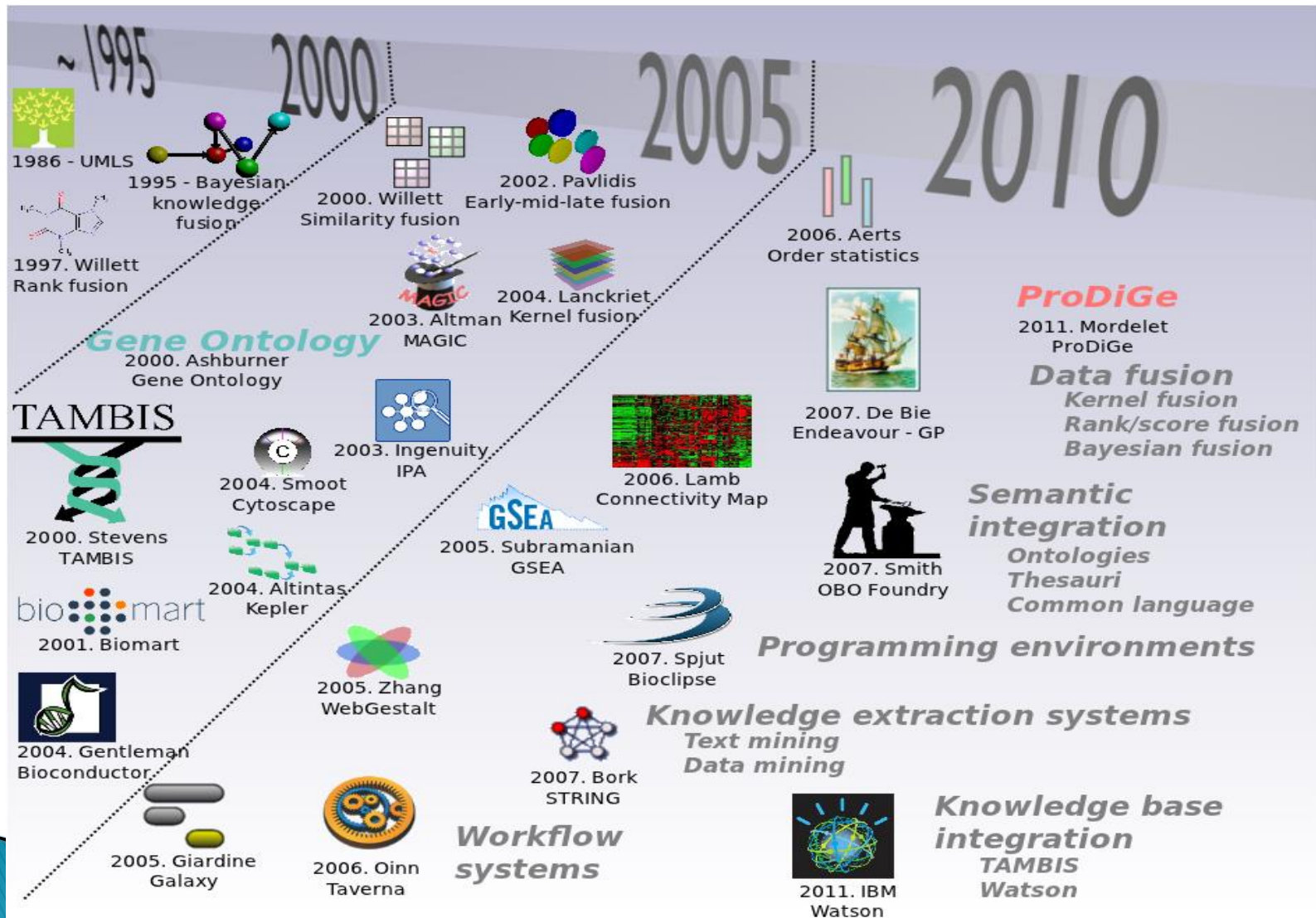
Adat- és tudásfúzió:~2000

▶ Kombinálhatóak-e:

- szakértői és szakirodalmi forrásokból:
 - ismeretek kvalitatív összefüggésekről,
 - különböző normál tartományok,
 - ismert, kvantitatív részleges statisztikák,
- statisztikai adatok
 - egy centrum adatai (egyetlen minőségbiztosítással),
 - több centrumú adatok a szubjektív mérések, standardizálásával,
 - különböző protokollal gyűjtött adatok.

(P. Antal: Integrative Analysis of Data, Literature, and Expert Knowledge, Ph.D. dissertation, K.U.Leuven)

Data and knowledge fusion landscape by 2010



Keret adat- és tudásfúzióra: Bayes-döntés és -tanulás

▶ Vélekedések és döntési preferenciák egységes kvantitatív kerete

- Thomas Bayes (c. 1702 – 1761)
- Bayesi értelmezése a valószínűségnek
- Bayes-szabály
- Bayes-statisztika
- Bayes-döntés
- Bayesi modellátlagolás
- Bayes-hálók
- Önkalibráló...



$$p(\text{Modell} \mid \text{Adat}) \propto p(\text{Adat} \mid \text{Modell}) p(\text{Modell})$$

(G.E.P.Box: „all models are wrong, but some are useful”)

Tool for fusion: Bayesian networks

Directed acyclic graph (DAG)

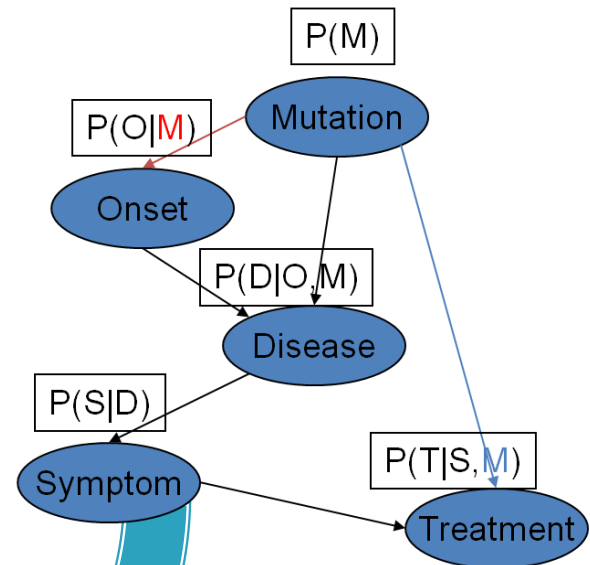
- nodes – random variables/domain entities
- edges – direct probabilistic dependencies (edges – causal relations)

Local models – $P(X_i | Pa(X_i))$

3. Concise representation of joint distributions

$$P(M, O, D, S, T) =$$

$$P(M)P(O | M)P(D | O, M)P(S | D)P(T | S, M)$$



1. Causal model

$M_P = \{I_{P,1}(X_1; Y_1 | Z_1), \dots\}$
2. Graphical representation of (in)dependencies

Uncertainty

- ▶ .A.Einstein: „God does not play dice..”

<https://arxiv.org/ftp/arxiv/papers/1301/1301.1656.pdf>

- ▶ Einstein–Podolski–Rosen paradox / Bell Test

- ▶ S. Hawking: „Does god play dice?”

<http://www.hawking.org.uk/does-god-play-dice.html>

- ▶ The BIG Bell Test (Nov30, 2016)

- http://index.hu/tudomany/2016/12/03/szazezren_bizonyitottak_einstein_tevedeset/
- <http://bist.eu/100000-people-participated-big-bell-test-unique-worldwide-quantum-physics-experiment/>



Interpretations of probability

- ▶ Sources of uncertainty
 - inherent uncertainty in the physical process;
 - inherent uncertainty at macroscopic level;
 - ignorance;
 - practical omissions;
- ▶ Interpretations of probabilities:
 - combinatoric;
 - physical propensities;
 - frequentist;
 - personal/subjectivist;
 - instrumentalist;

$$\lim_{N \rightarrow \infty} \frac{N_A}{N} = \lim_{N \rightarrow \infty} \hat{p}_N(A) = p(A) ? p(A | \xi)$$

A chronology

- ▶ [1713] *Ars Conjectandi* (The Art of Conjecture), Jacob Bernoulli
 - **Subjectivist interpretation** of probabilities
- ▶ [1718] *The Doctrine of Chances*, Abraham de Moivre
 - the first textbook on probability theory
 - **Forward predictions**
 - „given a specified number of white and black balls in an urn, what is the probability of drawing a black ball?”
 - his own death
- ▶ [1764, posthumous] *Essay Towards Solving a Problem in the Doctrine of Chances*, Thomas Bayes
 - **Backward questions:** „given that one or more balls has been drawn, what can be said about the number of white and black balls in the urn”
- ▶ [1812], *Théorie analytique des probabilités*, Pierre–Simon Laplace
 - General Bayes rule
- ▶ ...
- ▶ [1933]: A. Kolmogorov: *Foundations of the Theory of Probability*

Bayes' rule

$$p(X | Y) = \frac{p(Y | X)p(X)}{p(Y)} = \frac{p(Y | X)p(X)}{\sum_x p(Y | X)p(X)}$$

Principles for induction

- ▶ Epicurus' (342? B.C. – 270 B.C.) principle of multiple explanations which states that one should *keep all hypotheses that are consistent with the data*.
- ▶ The principle of Occam's razor (1285 – 1349, sometimes spelt Ockham). Occam's razor states that when inferring causes *entities should not be multiplied beyond necessity*. This is widely understood to mean: Among all hypotheses consistent with the observations, choose the simplest. In terms of a prior distribution over hypotheses, this is the same as giving simpler hypotheses higher a priori probability, and more complex ones lower probability.

Bayes rule, Bayesianism

George E.P. Box: „all models are wrong, but some are useful”

- ▶ Prediction without model identification?
 - Bayesian model averaging
 - „Kernel” methods

A scientific research paradigm

$$p(\textit{Model} | \textit{Data}) \propto p(\textit{Data} | \textit{Model}) p(\textit{Model})$$

A practical method for inverting causal knowledge to diagnostic tool.

$$p(\textit{Cause} | \textit{Effect}) \propto p(\textit{Effect} | \textit{Cause}) \times p(\textit{Cause})$$

Frequentist vs Bayesian prediction

In the frequentist approach: Model identification (selection) is necessary

$$p(\textit{prediction} | \textit{data}) = p(\textit{prediction} | \textit{BestModel}(\textit{data}))$$

In the Bayesian approach models are weighted

$$p(\textit{prediction} | \textit{data}) = \sum_i p(\textit{pred.} | \textit{Model}_i) p(\textit{Model}_i | \textit{data})$$

Note: in the Bayesian approach there is no need for model selection



Full Bayesian learning

View learning as Bayesian updating of a probability distribution over the **hypothesis space**

H is the hypothesis variable, values h_1, h_2, \dots , prior $\mathbf{P}(H)$ j th observation d_j gives the outcome of random variable D_j training data $\mathbf{d} = d_1, \dots, d_N$

Given the data so far, each hypothesis has a posterior probability:

$$P(h_i|\mathbf{d}) = \alpha P(\mathbf{d}|h_i)P(h_i)$$

where $P(\mathbf{d}|h_i)$ is called the **likelihood**

Predictions use a likelihood-weighted average over the hypotheses:

$$\mathbf{P}(X|\mathbf{d}) = \sum_i \mathbf{P}(X|\mathbf{d}, h_i)P(h_i|\mathbf{d}) = \sum_i \mathbf{P}(X|h_i)P(h_i|\mathbf{d})$$

No need to pick one best-guess hypothesis!

Bayesian model averaging

View learning as Bayesian updating of a probability distribution over the hypothesis space

H is the hypothesis variable, values h_1, h_2, \dots , prior $P(H)$

j th observation d_j gives the outcome of random variable D_j
training data $\mathbf{d} = d_1, \dots, d_N$

Given the data so far, each hypothesis has a posterior probability:

$$P(h_i|\mathbf{d}) = \alpha P(\mathbf{d}|h_i)P(h_i)$$

where $P(\mathbf{d}|h_i)$ is called the likelihood

Predictions use a likelihood-weighted average over the hypotheses:

$$P(X|\mathbf{d}) = \sum_i P(X|\mathbf{d}, h_i)P(h_i|\mathbf{d}) = \sum_i P(X|h_i)P(h_i|\mathbf{d})$$

No need to pick one best-guess hypothesis!

Bayesian Model Averaging example

Suppose there are five kinds of bags of candies:

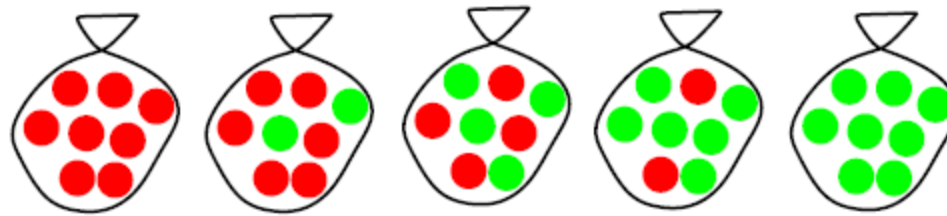
10% are h_1 : 100% cherry candies

20% are h_2 : 75% cherry candies + 25% lime candies

40% are h_3 : 50% cherry candies + 50% lime candies

20% are h_4 : 25% cherry candies + 75% lime candies

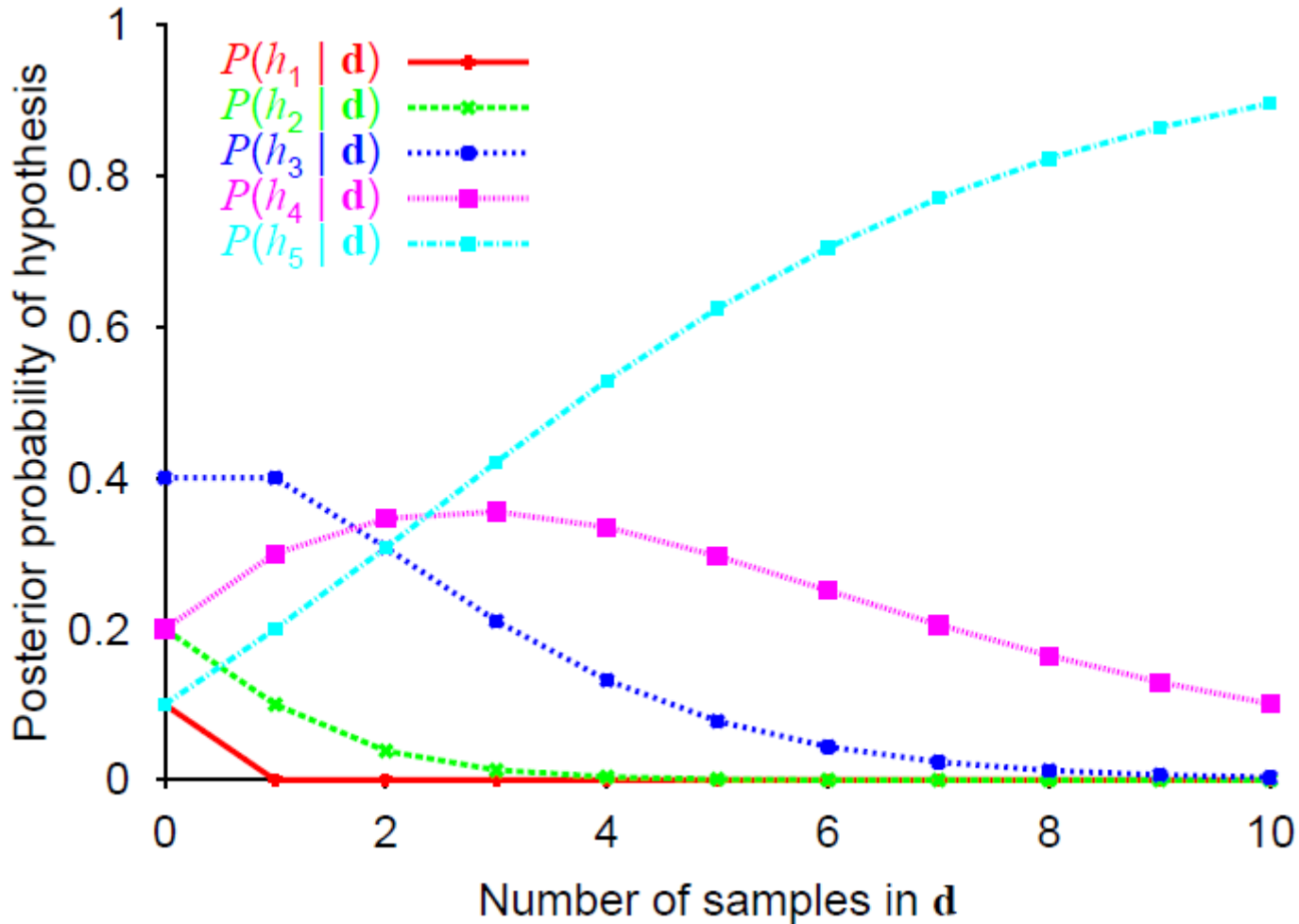
10% are h_5 : 100% lime candies



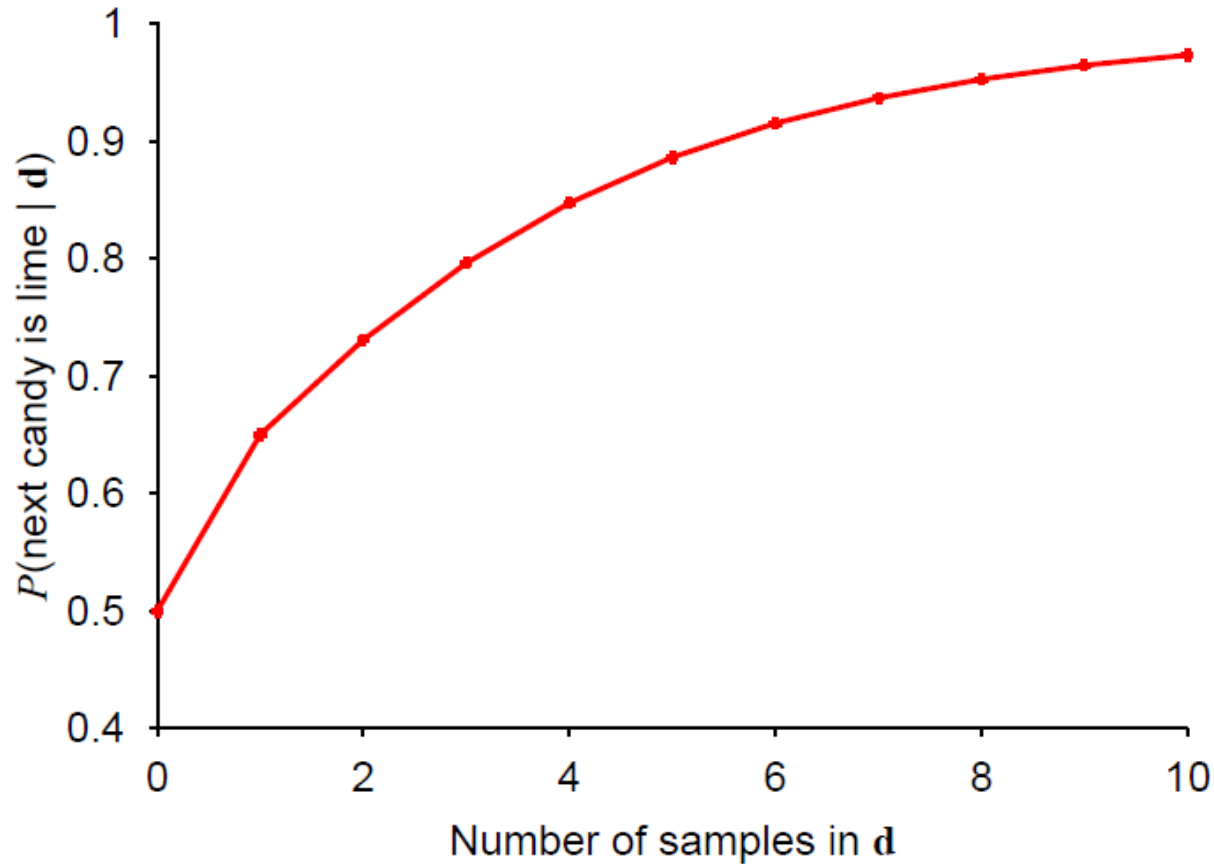
Then we observe candies drawn from some bag: ●●●●●●●●●●

What kind of bag is it? What flavour will the next candy be?

Learning rate for models



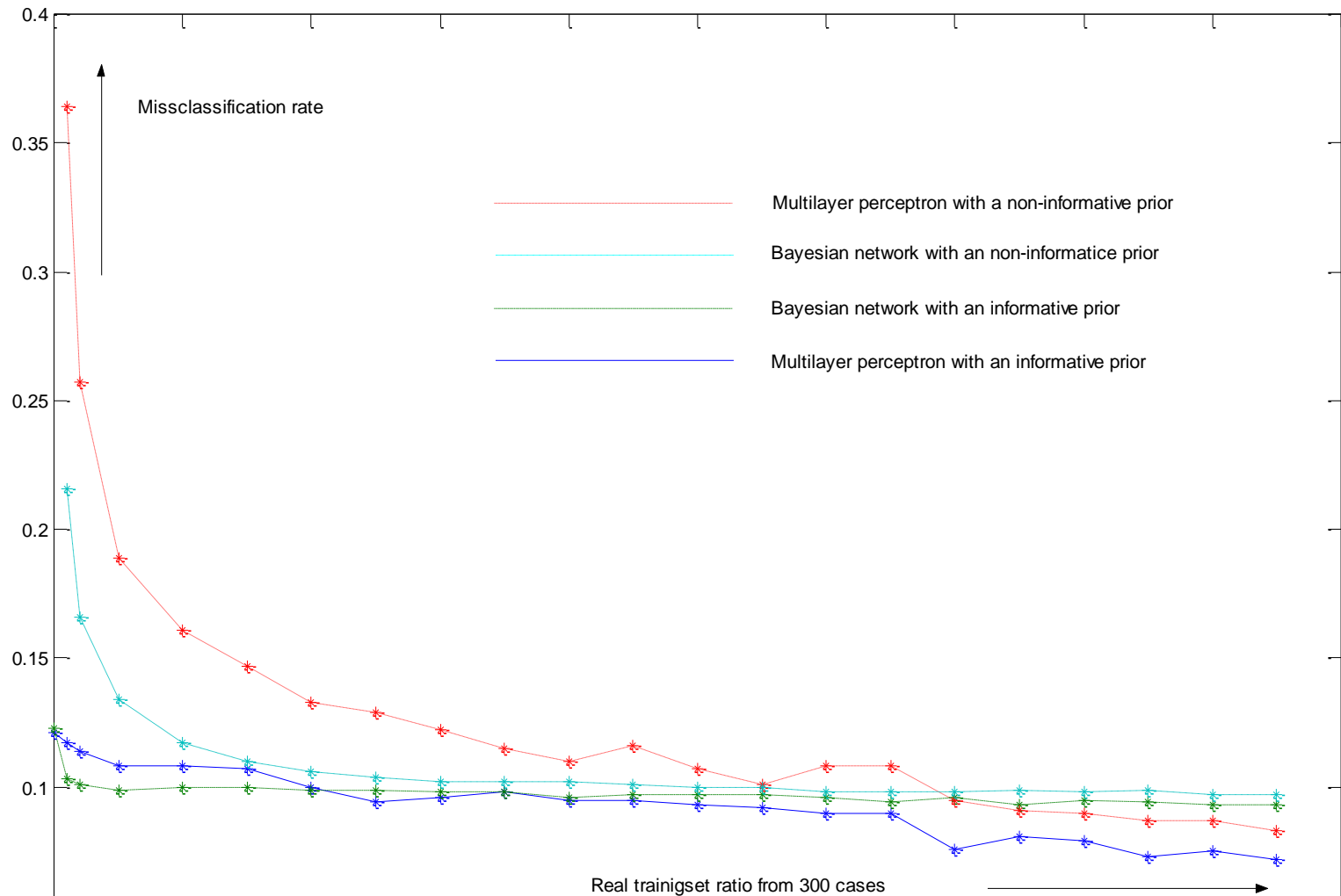
Learning rate for model predictions



Two main challenges in fusion

- ▶ Informative priors
 - Non-degenerative (do not exclude good models)
 - Transformation invariance (Jeffrey's prior)
 - Transparent (interpretable, provable)
 - Complexity regularization as well
 - [Conjugate (hyperparameters, closed for learning)]
- ▶ Bayesian model averaging

„Informált” neurális hálózatok II.



P. Antal, G. Fannes, D. Timmerman, Y. Moreau, B. De Moor: Bayesian Applications of Belief Networks and Multilayer Perceptrons for Ovarian Tumor Classification with Rejection, *Artificial Intelligence in Medicine*, vol. 29, pp 39-60, 2003

Summary

- ▶ Factors behind the AI/machine learning „hype”
 - Data, knowledge, computation
- ▶ Complex models in fusion
 - Functions, probability distributions, causal models, decision models
- ▶ Complex process of fusion
 - Study design, data engineering, ...
- ▶ A coherent framework for fusion
 - Bayesian decision theory, Bayesian learning