Digi breakfast on Data Science

19 Sep 2014

Aristides Gionis, Aalto SCI

event planned with

Johanna Bragge, Keijo Heljanko, Pekka Malo

Sami Kaski, Mikko Kurimo, Kirsi Virrantaus

data science in Helsinki area

strong competitive advantage

- long tradition in data analysis and machine learning
- centers of excellence, HIIT
- established technology companies
- young companies and startup hub

how to leverage this advantage?

how to become a world-class location for data science?

dimensions

education

research directions

collaborations with companies

- research and education

sharing

- data, tools, expertise, problems, ideas

agenda

Keynote

Heikki Mannila, Academy of Finland

Analytics and data science education

Johanna Bragge, Aalto BIZ

Data camp and hackathon initiatives

Jukka Nurminen, Aalto SCI

Data-driven research spotlights from Aalto

Data hub initiative

Keijo Heljanko, Aalto SCI

Invited talk

Ville Peltola, IBM

Invited talk

Kaisa Salakka, Comptel

Spotlights from Helsinki companies

Questions and discussion

questions

interdisciplinary research centered on data

ways to identify synergies within Aalto

ways to identify complementary expertise

strengthening ties between university and industry

effective ways to collaborate

sharing data

open channel for
comments & questions

http://presemo.aalto.fi/data

keynote

Heikki Mannila

Academy of Finland

SUOMEN AKATEMIA

Data science

- The role of data in science
  - Theory – experimentation – computation – data
  - Changes in the way science is done
- The role of data in industry
  - Changes in the way companies operate
- The role of data in decision-making
- Implications for education
- Huge possibilities for research with fascinating themes and real impact

 © SUOMEN AKATEMIA |

SUOMEN AKATEMIA

Data science

- Some suggestions
- Mainly for the computational people
- Not about organizational issues or about funding

- Not to be taken very seriously

SUOMEN AKATEMIA

Ten suggestions

- Think big
  - Look at areas where computational thinking can really make a difference
- Think small
  - Look at problems where something can be done quickly
- Do not follow the herd
  - There are lots of possibilities in data science – look for someplace that is not so crowded

 © SUOMEN AKATEMIA |

SUOMEN AKATEMIA

Ten suggestions

- Talk to people in other camps
  - Other sciences, start-ups, large companies, …
  - Lots of interesting people to talk to
- Work on a few things
  - But not too many
- Do not assume that only size matters
  - Big data is nice, but …
  - Even small data sets can be very useful and challenging

 © SUOMEN AKATEMIA |

SUOMEN AKATEMIA

Ten suggestions

- Beware and take advantage of technological development
  - Does my problem go away in a few years?
  - What data will be available in a few years?

- Simplify
  - The best methods are the simple ones

- Assume infinite computational power
  - A good first approximation: what could we find if there are no limitations; a test for what the possiblities are

SUOMEN AKATEMIA

Ten suggestions

- Think big
- Think small
- Do not follow the herd
- Talk to people in other camps
- Work on a few things
- Do not assume that only size matters
- Beware and take advantage of technological development
- Simplify
- Assume infinite computational power

- No more than six to be considered
- The last one is mandatory:

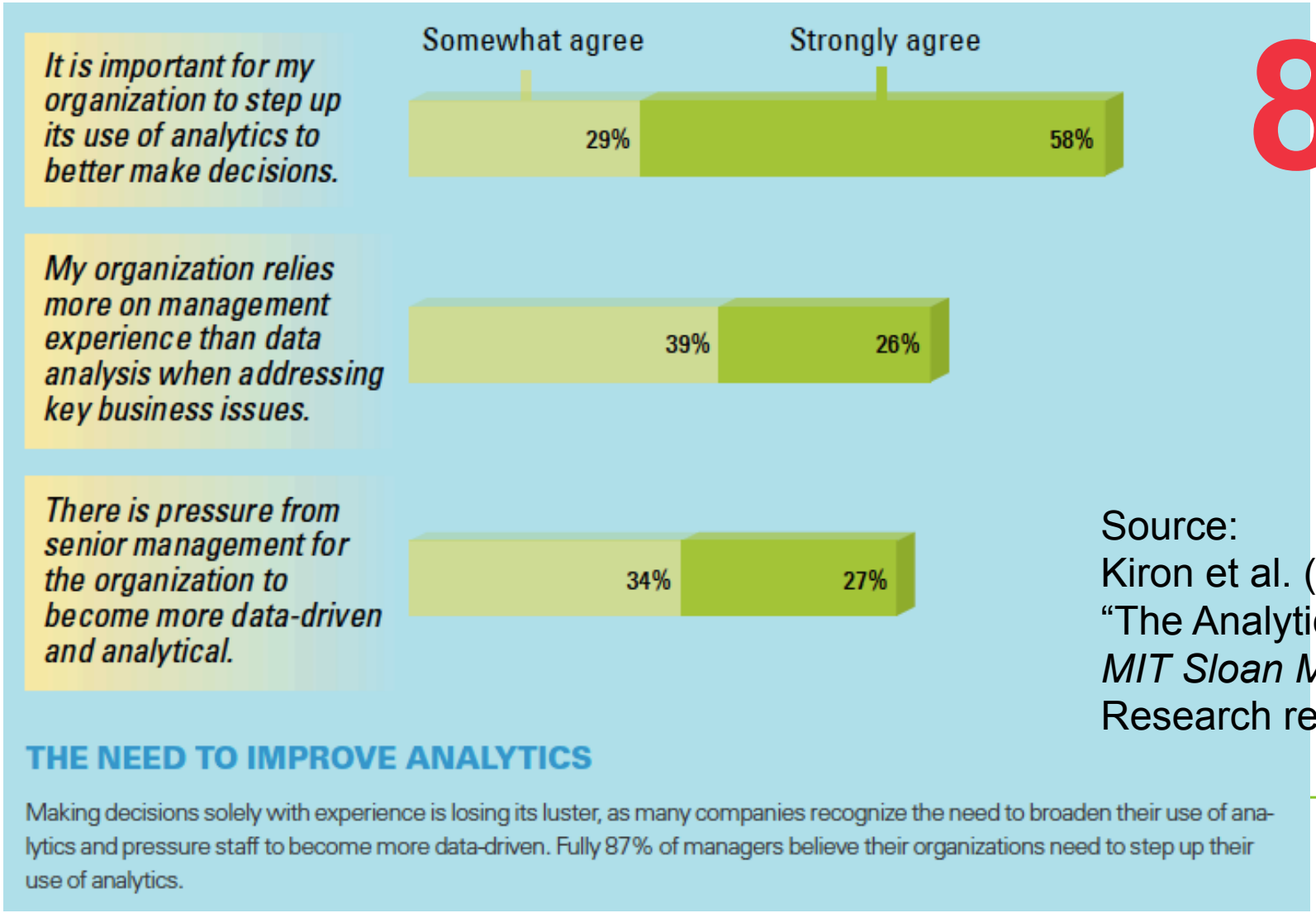
- Have fun

Aalto University
School of Business

Analytics and Data Science Education in Aalto

Johanna Bragge, PhD, Senior University Lecturer
Aalto BIZ / Department of Information and Service Economy

Importance of studying analytics & DS



87%

Source:
Kiron et al. (2014),
"The Analytics Mandate",
MIT Sloan Management Review,
Research report, May, 2014.

Analytics and data science education in Aalto

New minor for all Aalto's MSc students: Analytics and Data Science (ADS)

Current Master's programmes:

Aalto BIZ: Information and Service Management (ISM), specialization area of Business Analytics

Aalto SCI: Machine Learning and Data Mining (Macadamia)

Aalto SCI: Foundations of Advanced Computing (FAdCo)

Forthcoming in Fall 2015:

Renewal of technical MSc programs in IT. Possibility to specialize in Big Data and Large Scale Computing

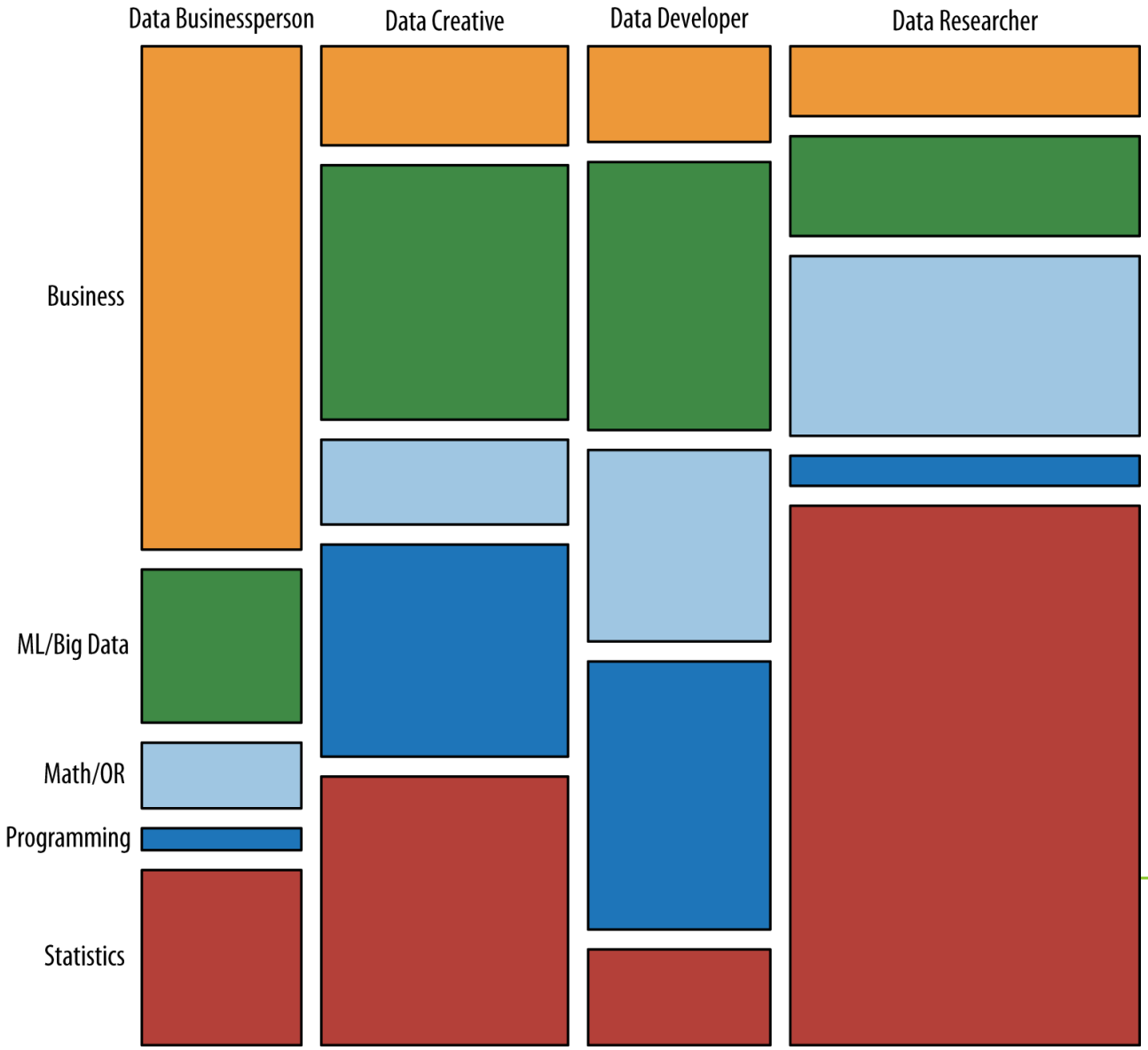
ADS's goal and requirements

The goal of ADS is to educate students on how to become proficient in making sense of big data, and how to apply data analysis skills on their domain of expertise.

Compulsory course				
	ICS- E4010	Introduction to Analytics and Data Science	2 CR	I period
At least one course from Statistical foundations subarea:				
SF	Becs-114.1311	Introduction to Bayesian Statistics	3	III
SF	Becs-114.2601	Bayesian Modeling	5	I-II
SF	MS-C2104	Introduction to Statistical Inference	5	III-IV
SF	MS-C2128	Prediction and Time Series Analysis*	5	II
SF	30E00800	Time Series Analysis* (*alternative to previous)	6	IV-V
At least one course from Computational methods subarea:				
CM	T-61.3050	Machine Learning: Basic Principles	5	I
CM	T-61.5060	Algorithmic Methods of Data Mining	5	I-II
CM	T-61.5010	Information Visualization	5	III
CM	CSE-E5430	Scalable Cloud Computing	5	I-II
CM	T-110.5121	Mobile Cloud Computing	5	I-II
At least one course from Business analytics/Applications:				
BA	Mat-2.3134	Decision Making and Problem Solving	5	I
BA	23E47000	Digital Marketing	6	I, V
BA	30E03000	Data science for Business	6	III
BA	37E01600	Data Resources Management	6	III
BA	57E00500	Capstone: Business Intelligence	6	I
AP	Becs-114.4150	Complex Networks	3-6	II
AP	Becs-E4101	Mathematical Modeling of Social Dynamics	3-6	II 2015
AP	Maa-123.3585	Spatial Data Mining	3-5	V
AP	Maa-123.3530	Visual Analysis	4	II
AP	Mat-2.2103	Design of Experiments and Statistical Models	5	III
AP	Mat-2.4177	Seminar on Case Studies in Operation Research	5	III-IV
AP	S-89.5150	Speech Recognition	5	II

http://studyguides.aalto.fi/minors-guide/2014/en/sci/sci-minors-for-all-aalto-students/analytics-and-data-science.html

4 data scientist clusters & 5 skills groups

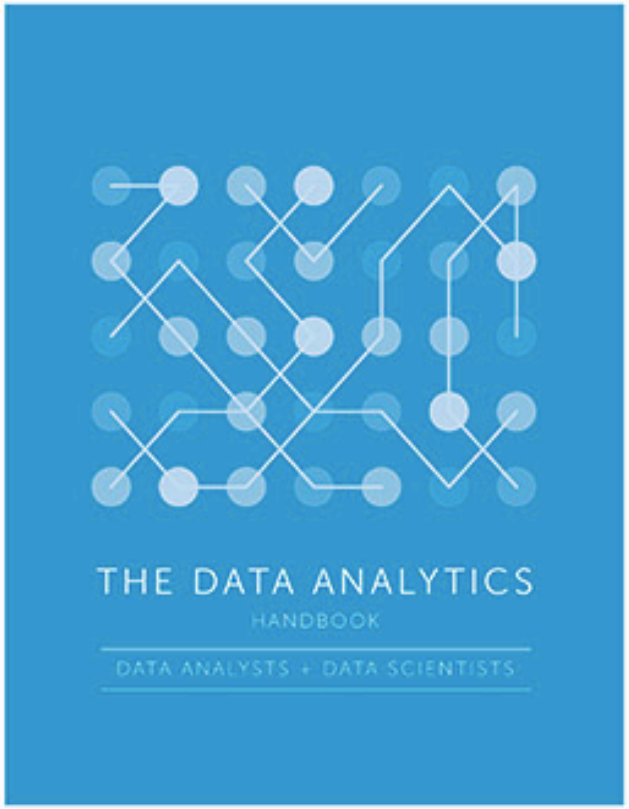


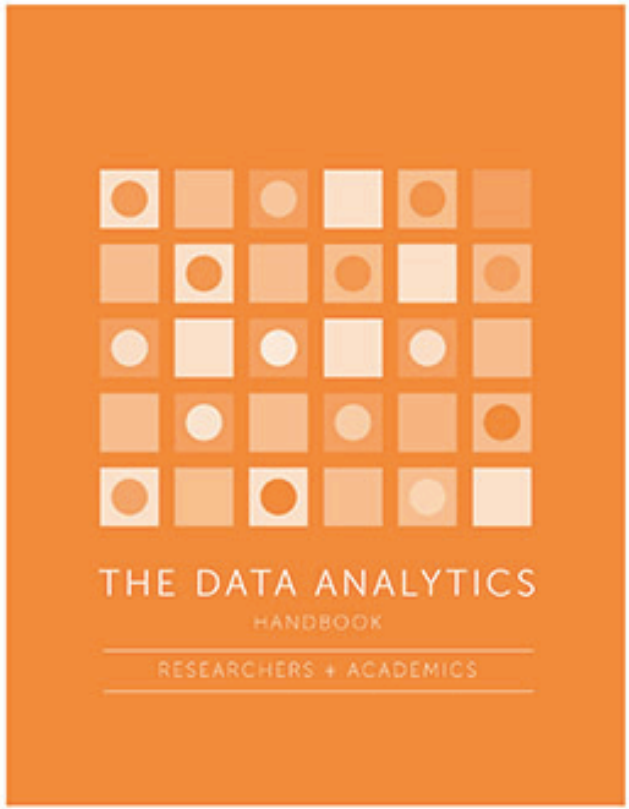
ML = Machine Learning
OR= Operations Research

Source:
Harris, H. D, Murphy, S. P. & Vaisman, M. (2013), “Analyzing the analyzers. An introspective survey of data scientist and their work”. O’Reilly, available at: http://www.oreilly.com/data/free/analyzing-the-analyzers.csp

O'REILLY®

Three data analytics handbooks & “Ask Peter Norvig” - useful resources!





Source: http://blog.teamleada.com/2014/08/ask-peter-norvig/

Aalto University
School of Science

Data camp and hackathon initiatives

Prof. Jukka K. Nurminen
Department of Computer Science and Engineering

17.9.2014

How to Teach Data Science?

- In addition to theoretical knowledge hands-on work is needed
- Partly covered with assignments on different courses
- Need for a project oriented course?
  - Bigger, realistic problems
  - Genuinely new problems without obvious solutions
  - Vaguely formulated
  - Specifying the problem is part of the work

Two Options

Seminar on case studies in data analysis

- Focused topic definition
- Student group “solves” the given problem with the help of industry and academic tutor
- Needed:
  - Problem
  - Dataset(s)
  - Regular guidance
- Main output
  - Solutions

Hackathon

- Theme specific e.g. Green Campus Hackathon
- Student group comes up with an idea and demo of an application (or business idea, new finding, …)
- Needed
  - Multiple datasets
  - Feedback
- Main output
  - New ideas

Consider and Vote – especially industry participants

- Would this kind of course make sense?
- Would you be able to provide the datasets and guidance?

- Please use the Presemo tool to state your view
  - Would you like to contribute to such course?
    - Which option would suit you best? (Seminar vs. Hackathon)
  - Would you be able to provide datasets? Which ones?
  - Other suggestions or ideas?

Aalto research spotlights

COIN center of excellence

Sami Kaski, Aalto SCI

Business analytics

Pekka Malo, Aalto BIZ

Spatial data analysis

Kirsi Virrantaus, Aalto ENG

Statistical signal processing

Esa Ollila, Aalto ELEC

ReSoLVE center of excellence

Maarit Mantere, Aalto SCI

Centre of Excellence on Computational Inference Research COIN

Principal Investigators and their groups

Erkki Oja
Professor,
Director of
COIN

Samuel Kaski
Professor,
Deputy Director
of COIN

Erik Aurell
Professor

Jukka Corander
Professor

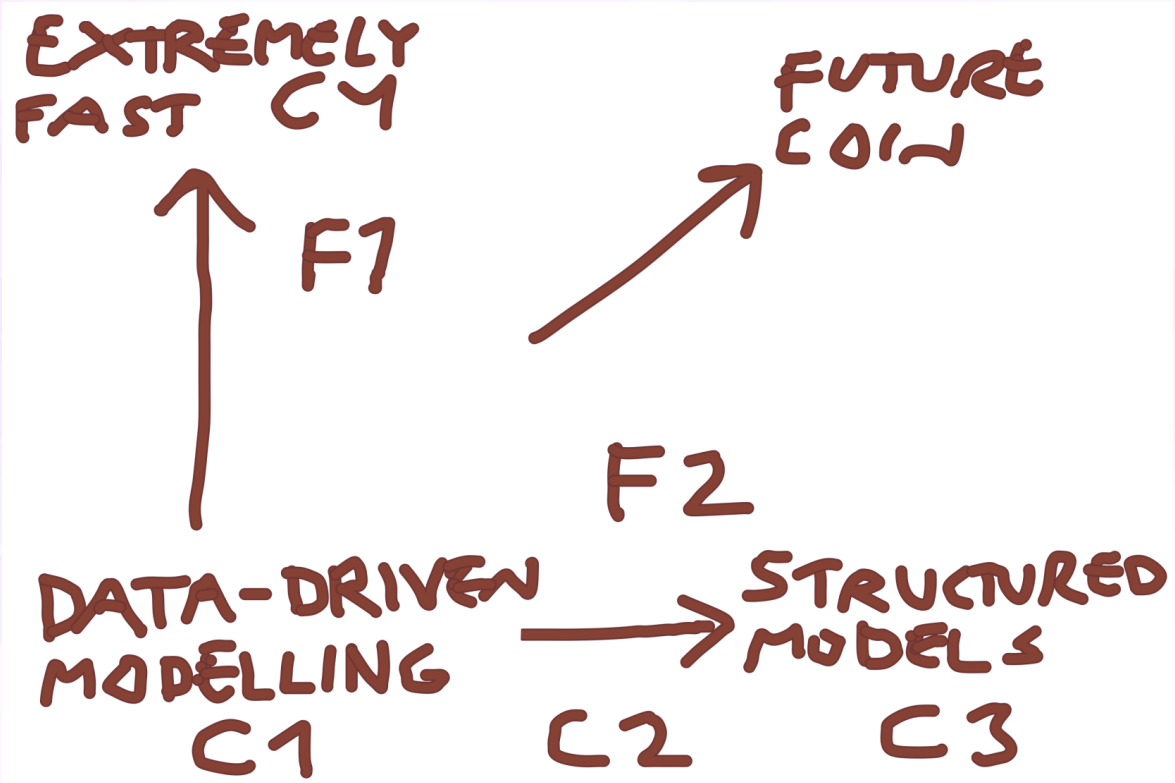
Jorma
Laaksonen
Docent, D.Sc.
(Tech.)

Petri Myllymäki
Professor

Ilkka Niemelä
Professor

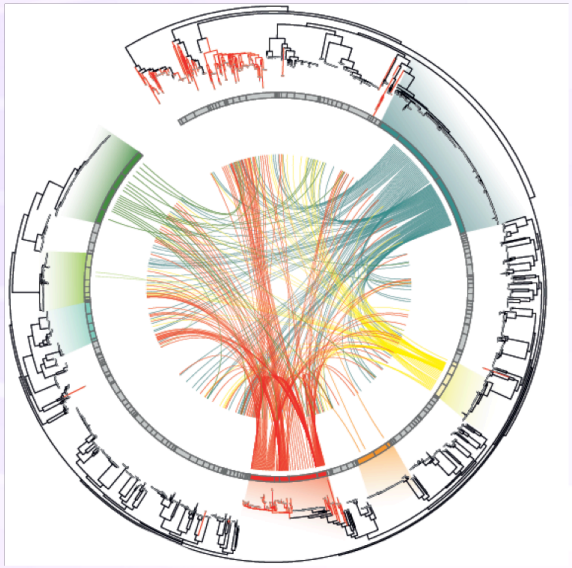
Research Plan: Goal

- Challenge and opportunity: Big Data and complex models

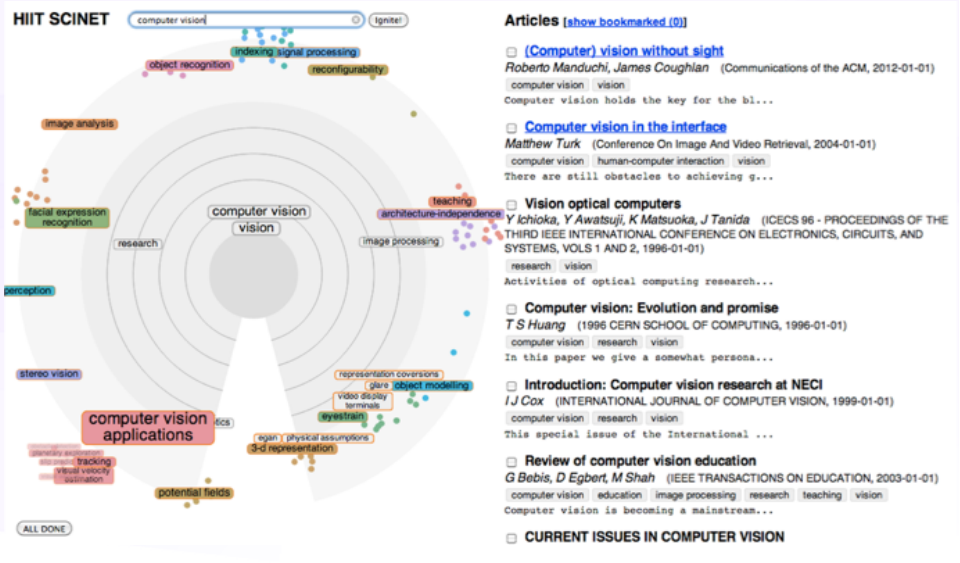


Research Plan: Focus

- Competitive advantage by bringing together central aspects of machine learning that are traditionally studied separately
- Two flagships:



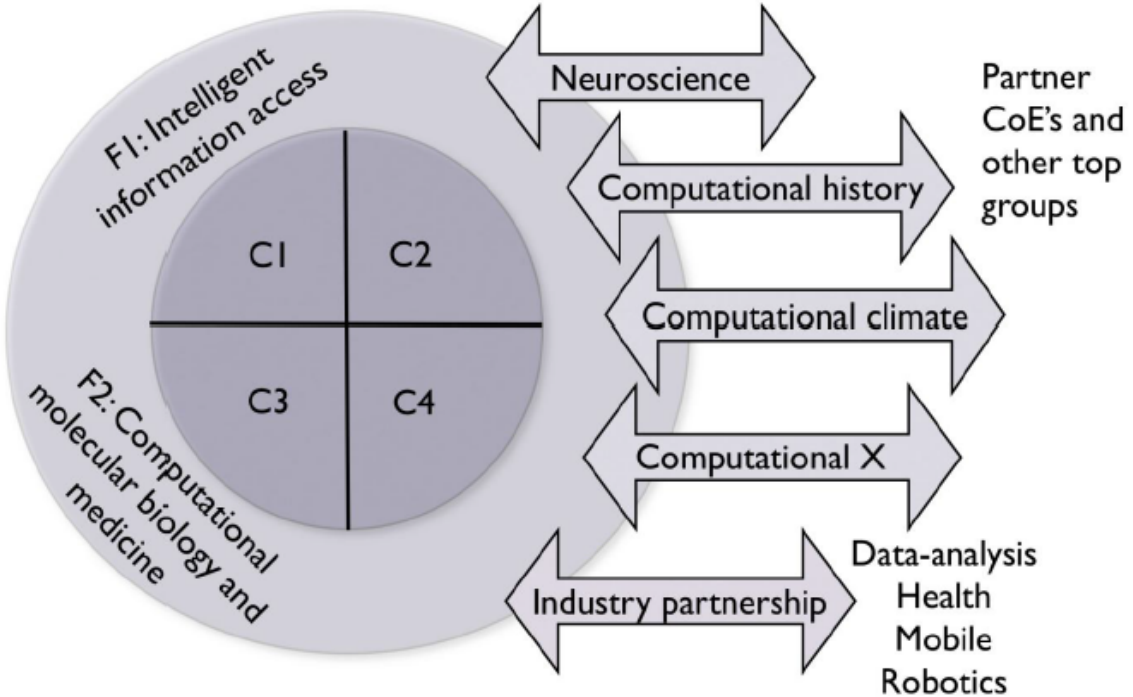
Inference on intractable models, with applications in studies of bacterial evolution



Interactive intent modelling with the SciNet system

Computational Inference Research

Flagships



Core Methodological Challenges

C1: Learning models from massive data
C2: Learning from multiple data sources
C3: Statistical inference in structured stochastic models
C4: Extreme inference engine

Aalto University

UNIVERSITY OF HELSINKI

Aalto University
School of Business

Fox School of Business
TEMPLE UNIVERSITY

FOX | GLOBAL CENTER FOR BIG DATA IN MOBILE ANALYTICS

Research Spotlight:
Marketing meets Wall Street

Pekka Malo
Aalto BIZ / Department of Information and Service Economy

Marketing meets Wall Street

How data science helps to build better metrics for marketing decisions?

Objectives of our project:

- To understand what is the role of product-market based assets (e.g., brand, customer, channel equity, and marketing actions) in driving financial performance and shareholder value

- To develop new tools for making marketing decisions using a blend of quantitative and unstructured data sources (e.g., news, social media)

Marketing meets Wall Street

Company feedback is essential

- What measures would company managers like to see on their dashboards?

External data is good, but internal data is needed too

- Do firms sacrifice long-term value creation to meet earnings targets?
- How “suboptimal” marketing strategies are reflected in company performance?
- What is the impact of quality of earnings (e.g., repeat customers vs. new customers) on valuation?
- What is the impact of uncertain payoff on marketing activities / resource allocation?

Data scientists are welcome to join our project!

- We have a few onboard already, but more is needed!

Research Group for Geoinformatics

Department of Real Estate, Planning and Geoinformatics
Aalto ENG
Kirsi Virrantaus, professor

- Spatial Data Analysis
- Spatial Statistics and Spatial Data Mining
- Visual Analytics
- Map Design
- Cognition and Visual Perception Processes
- Spatio-Temporal Knowledge Management
- Situational Awareness
  - Safety and Defence, Urban Design, Traffic

Topics of Potential Co-Operation

- Spatio - temporal data analysis method development for various applications
- Our special interest and competence is on
  - safety, crisis management, urban and traffic applications

- Visualizations for spatio-temporal data
- Visual design of user interfaces

- Development of GIS applications' prototypes and demonstrators

Statistical Signal Processing Theory and Methods for Data Science @ Aalto ELEC, Department of Signal Processing and Acoustics

Research team

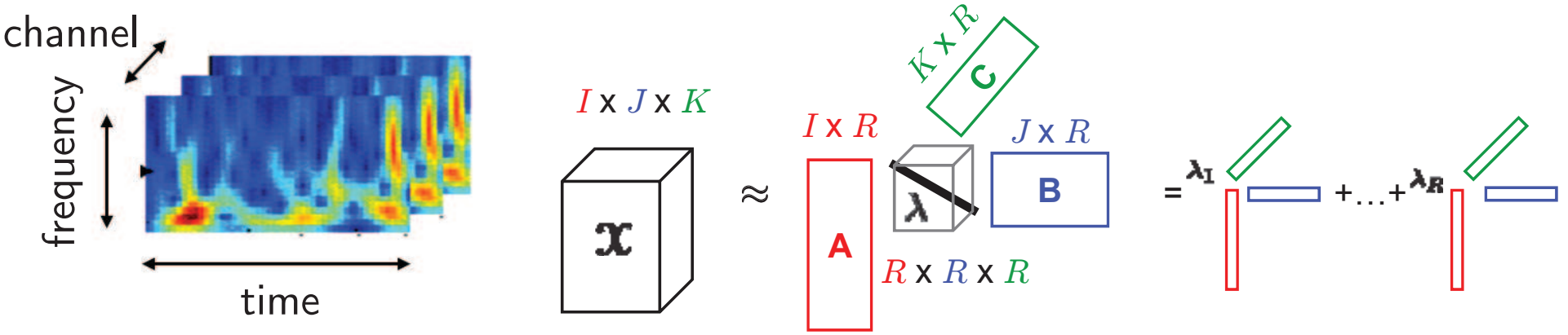
- Academy Professor Visa Koivunen
- Academy Research Fellow Esa Ollila (Regularized Estimation of High-dimensional Covariance Matrices, ...)
- Dr. Hyon-Jung Kim (Tensor Data Analysis for Large Scale Data)
- Doctoral student Shahab Basiri (Big Data Bootstrap)
- Master students (Compressive Sensing Methods, Statistical Inference in Smart Grids, High-dimensional Sensor Array SP)

Collaboration

- Princeton University and Rutgers University (USA)
- The Hebrew University of Jerusalem (Israel)
- Supelec (France)
- Yonsei University (Korea)

Tensor Methods for Large Scale Data

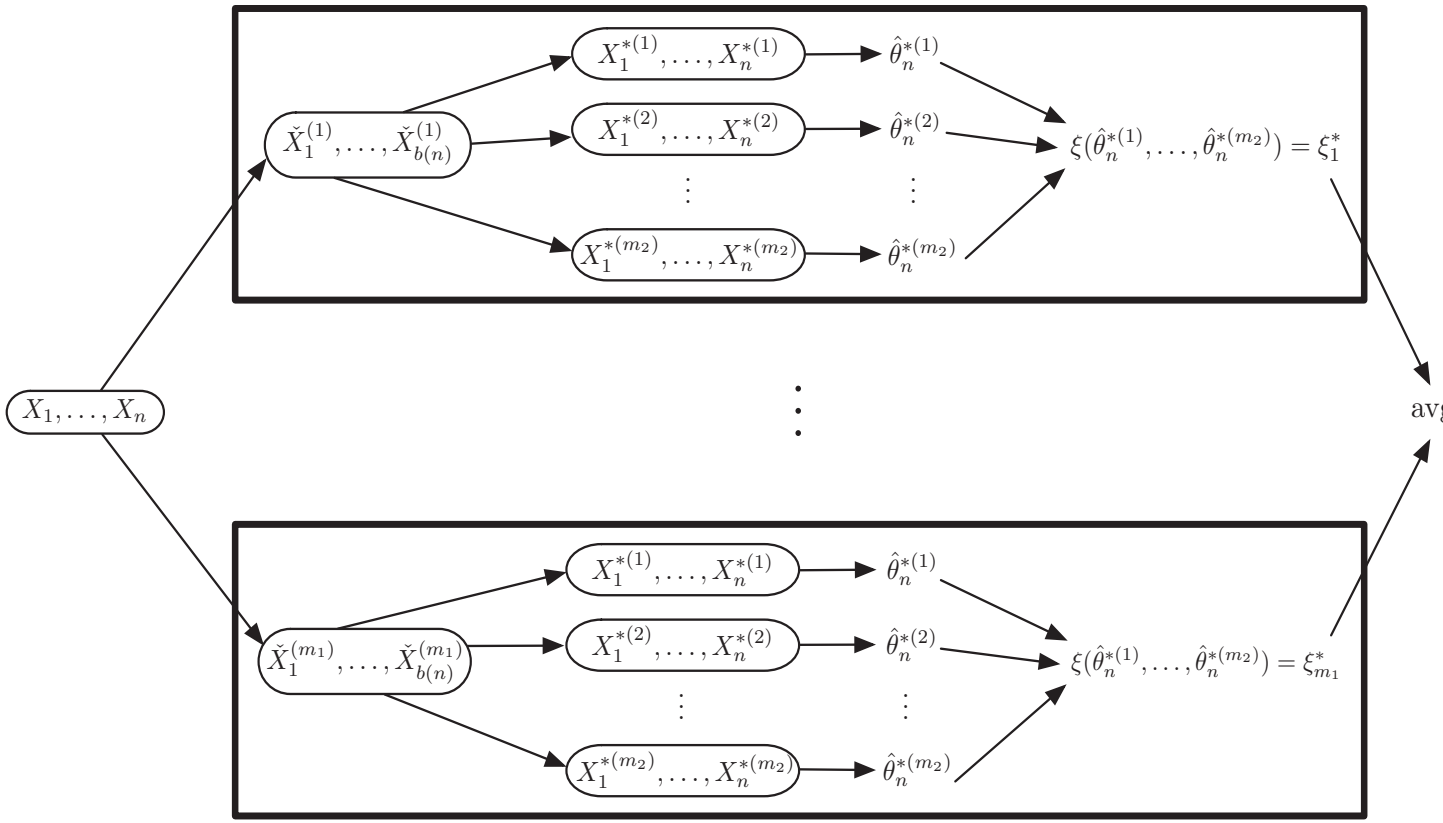
- Tensor models (PARAFAC, Tucker and Tensor Train) provide a unified representation, simplified notation and algebra for *multidimensional* data.
- We develop computational methods for tensor data
- EEG, data from smart phones, health data, business data, multiantenna radio channel data



Our research

- *Past*: Statistically robust methods for extracting relevant information. Exploiting sparseness and low-rank structure for better accuracy/stability
- *Future*: Analysis of large scale real-world data.

Scalable and Distributed Statistical Inference for Big Data: Fast and Robust Bootstrapping for Subsets of the Data



subsampling w/o replacement smaller sets

***

distributed storage and computation

***

scalable for large data sets

Data $X_1, \ldots, X_n \rightarrow$ estimator $\hat{\theta}(X_1, \ldots, X_n) \rightarrow$ accuracy ξ (error bar)

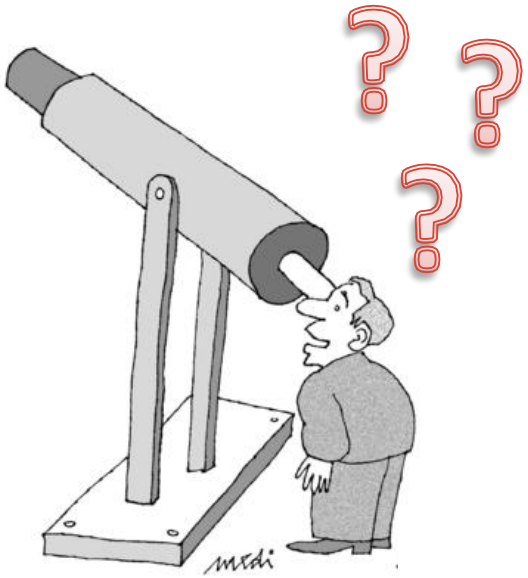
- Distributed and scalable bootstrapping for big data yields error bars of parameter estimates or confidence intervals (e.g., for hypothesis testing).

Research on SOlar Long-term Variability and Effects

Courtesy NASA/SOHO

When the Sun killed an Australian cow….

…huge societal value



Further enforce co-operation with private sector

Collaboration with academic partners vital for efficient and intelligent data analysis methods

MISSION

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

Aalto Data Hub Initiative

Assoc. Prof. Keijo Heljanko
Keijo.heljanko@aalto.fi

Aalto University

UNIVERSITY OF HELSINKI

Aalto Data Hub Initiative

- Aalto has a wide range of experts on analytics and data science
- The new minor on analytics and data science gives Aalto students access to a wide variety of courses on the topic
- How can we provide similar access for companies to Aalto data science researchers and students?

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

Aalto Data Hub – Idea

- A networking group for both Companies and Aalto researchers working on data science
- A contact point to find Aalto researchers working on different subareas of data science
- A way to share best practices on Big Data analytics technology

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

Sharing Your Data

- Data science needs datasets from Companies
  - Data for Data Science major student projects
  - Data for researchers – PhD and MSc students need data sets for Theses work, as do research projects

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

Sharing Big Data Processing Best Practices

- Aalto Researchers can share best practices on Big Data platforms and analytics
- Companies can network and share knowledge with both researchers and other companies
- Is there sufficient interest in creating an “Aalto Big Data Platform”?

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

Aalto Data Hub – Getting Involved

- Think if there are data sets that you can share with us
  - Aalto student projects
  - Data science researchers
- Do you want to share Big Data best practices?
- Get in touch: keijo.heljanko@aalto.fi
- Join the LinkedIn Aalto Data Hub Group

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

agenda

Keynote

Heikki Mannila, Academy of Finland

Analytics and data science education

Johanna Bragge, Aalto BIZ

Data camp and hackathon initiatives

Jukka Nurminen, Aalto SCI

Data-driven research spotlights from Aalto

Data hub initiative

Keijo Heljanko, Aalto SCI

Invited talk

Ville Peltola, IBM

Invited talk

Kaisa Salakka, Comptel

Spotlights from Helsinki companies

Questions and discussion

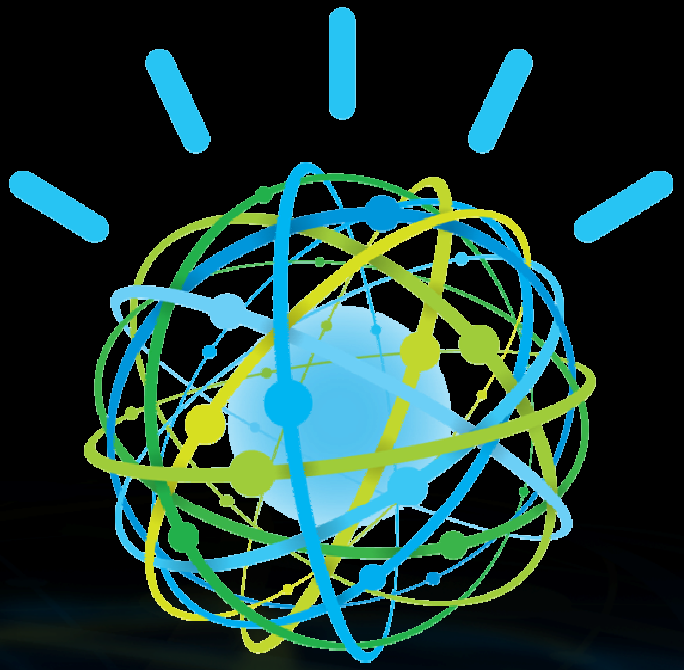
invited talk

Ville Peltola, IBM

IBM

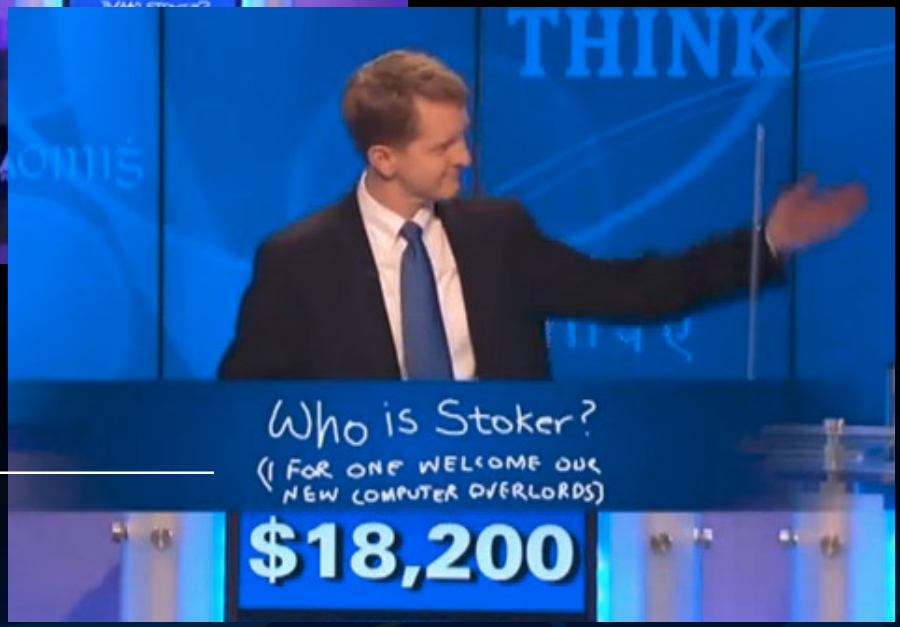
Creative Discovery in Data Science

Ville Peltola, Innovation Director



IBM

Watson and the Jeopardy grand challenge 2011



"I for one welcome our new computer overlords"

IBM

Putting Watson to work to address the world's pressing issues

IBM Research Project *(2006 –)*

R&D

Jeopardy! Grand Challenge *(Feb 2011)*

Demonstration

Watson for Healthcare *(Aug 2011 –)*

Commercialization

Watson for Financial Services *(Mar 2012 –)*

Expansion

Watson Industry Solutions *(2012 –)*

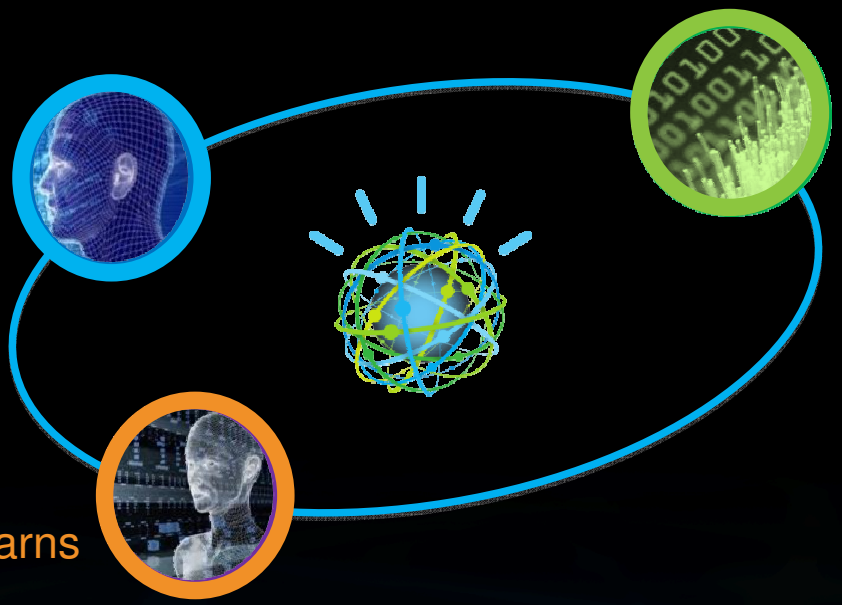
Cross-industry Applications

IBM

IBM Watson combines transformational technologies

1 Understands natural language and human communication

2 Generates and evaluates evidence-based hypothesis



3 Adapts and learns from user selections and responses

Watson enables three classes of cognitive services

Ask

- Ask questions for greater insights
- Natural language dialogue

Discover

- Find the rationale for given responses
- Prompt for inputs to yield improved responses

Decide

- Ingest and analyze domain sources, info models
- Evidence-based decisions with greater confidence

IBM

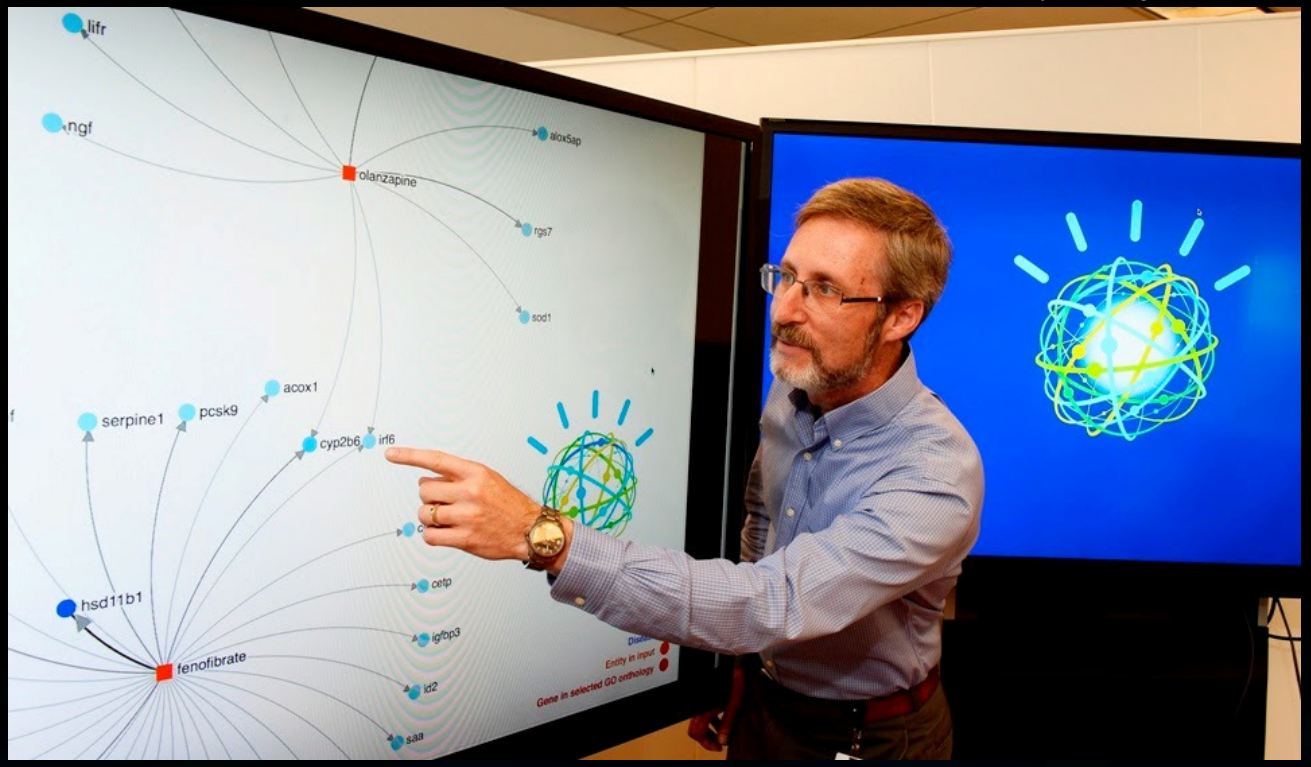
Could a computer discover new food recipes?

IBM

Chef Watson

Watson Discovery Advisor

Bayer College of Medicine



Watson evaluated and analyzed nearly 70,000 scientific articles on p53 to predict proteins that turn on or off p53's activity. As a result, cancer researchers recently found six potential proteins to target for new research, a dramatic increase from the average one protein discovery per year.

Have at it, programmers: IBM makes Watson available via API

by Derrick Harris NOV. 14, 2013 - 12:43 AM PST

8 Comments +1

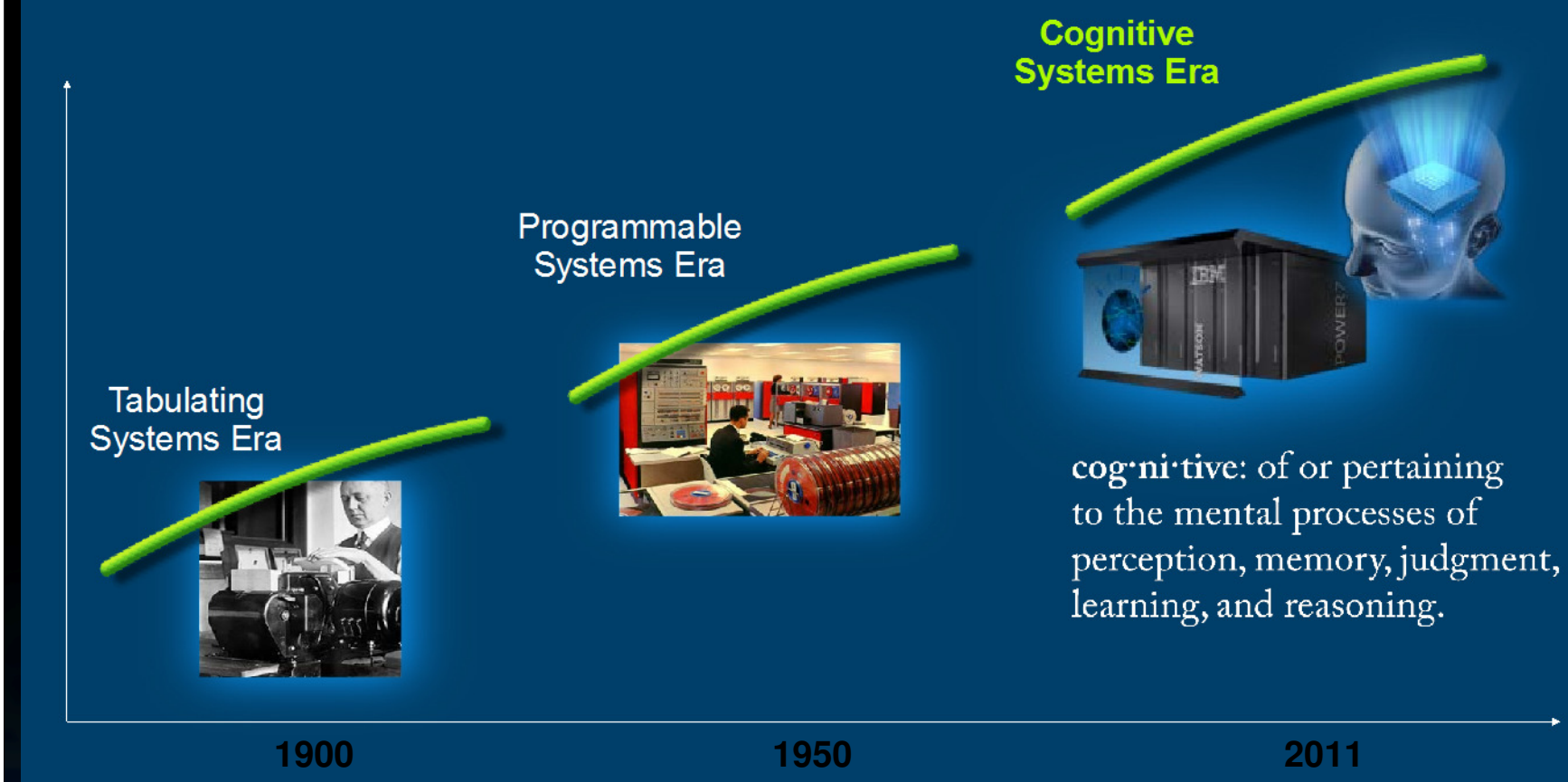
A A

SUMMARY: *IBM has upped the ante in the API game by making its Watson question-answering system available as a service. That's right, Watson could soon power your smartphone app.*

photo: IBM

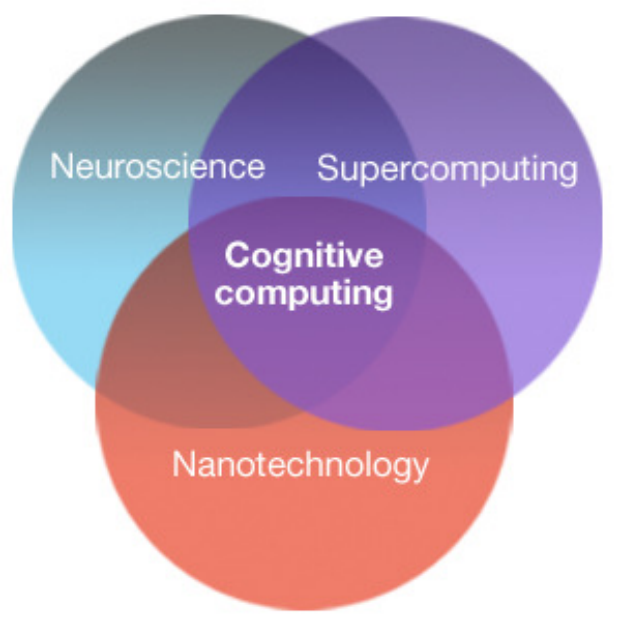
IBM

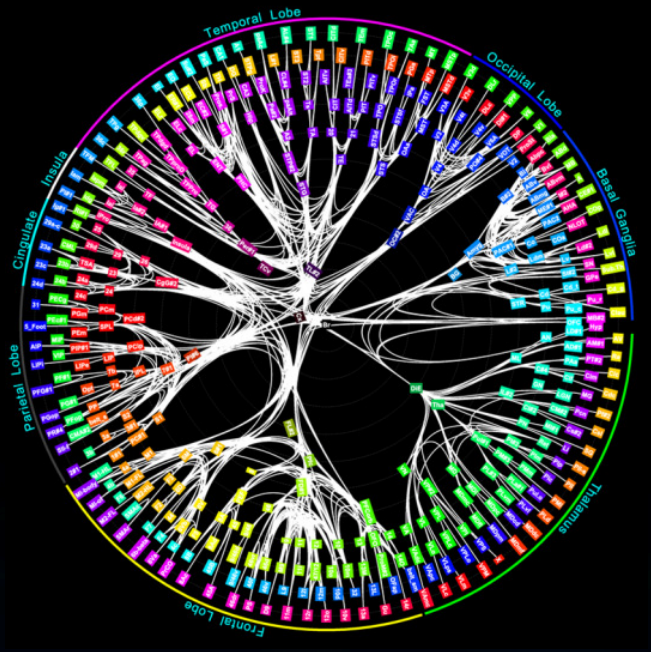
Watson is ushering in a new era of computing . . .



IBM

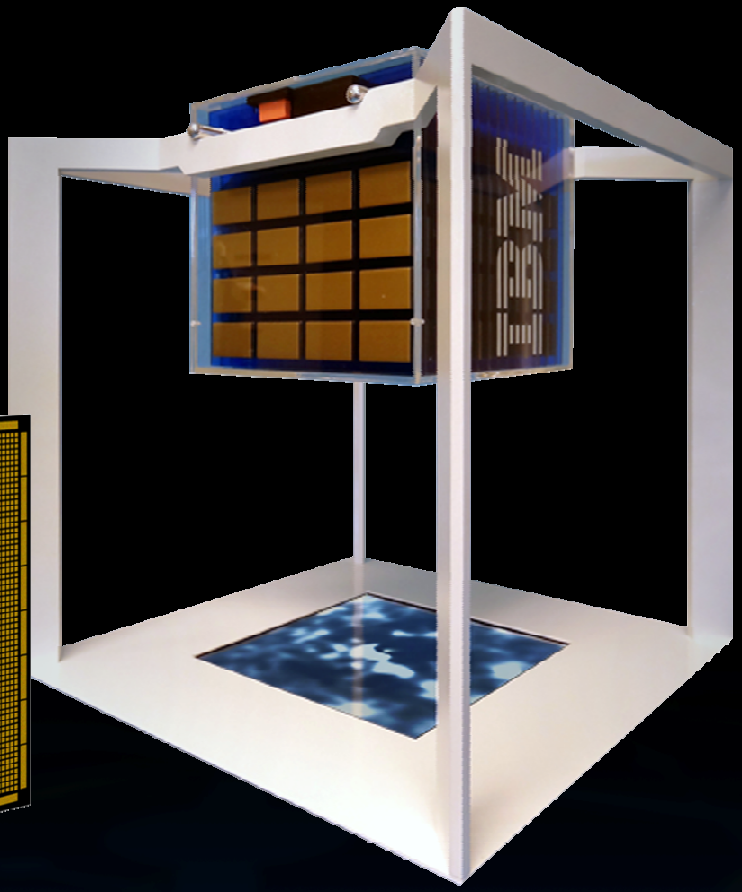
Systems of Neuromorphic Adaptive Plastic Scalable Electronics (SyNAPSE)

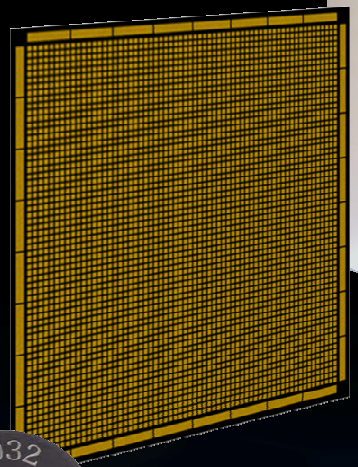




IBM

TrueNorth - Field-programmable Neurosynaptic Supercomputer



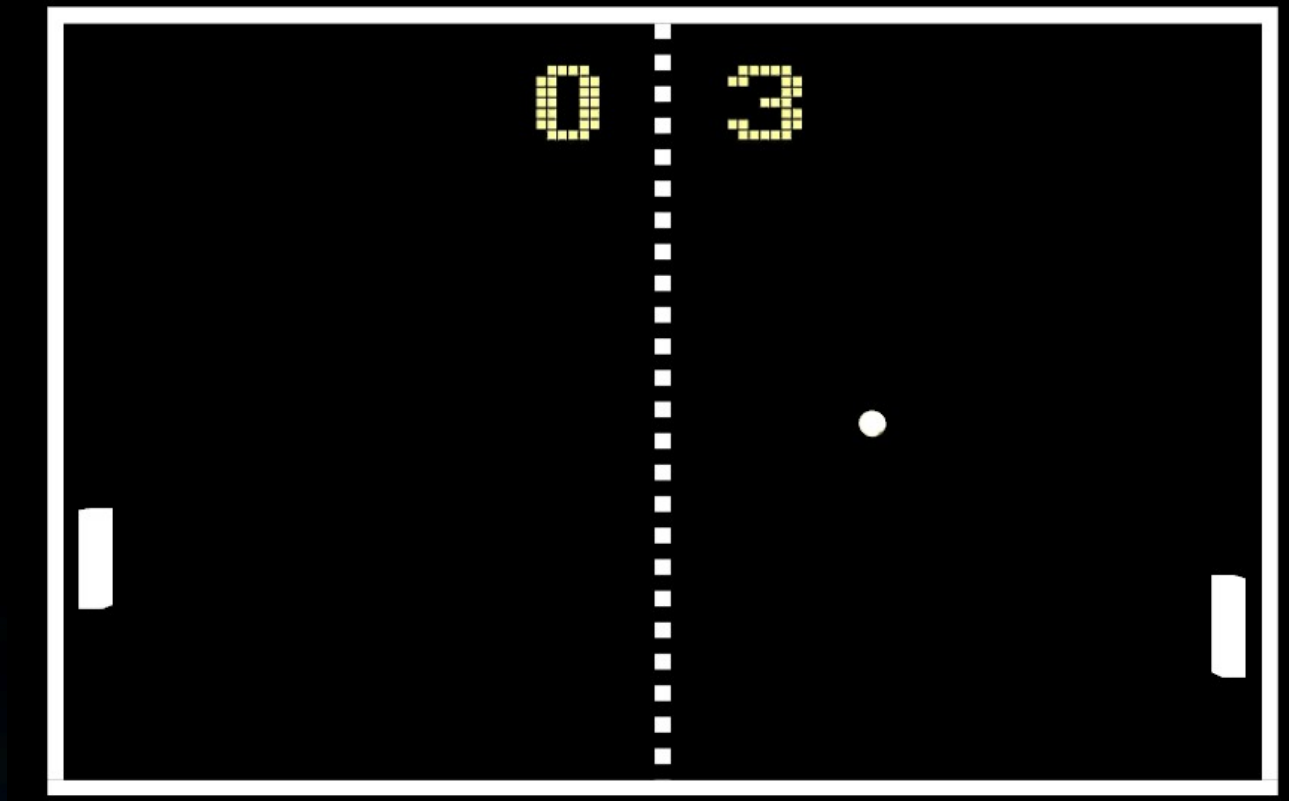


- 4096 cores
- 10^6 neurons
- 2.56×10^8 synapses
- 28 nanometer
- 50 milliWatts
- 4 cm^2
- parallel, distributed, scalable, multimodal, multitasking, real-time
- non-von Neumann

"You don't do linear programming, you train the hardware"

IBM

We have now teached it to for example play pong…



Thank you

Twitter: @villepeltola

invited talk

Kaisa Salakka, Comptel

Collaboration Between University and Industry

Kaisa Salakka, Senior Product Manager, Analytics

kaisa.salakka@comptel.com

© Comptel Corporation 2014

JOINT CONFIDENTIAL

Task at Hand

”

- What are the ways to enhance collaboration between the university and industry?
- How to get students involved in data science projects?
- How to get real and relevant data to university?
- What do you see as major opportunities in this field?

© Comptel Corporation 2012

Comptel - More than 290 Customers Across 85 Countries

Source: Comptel stock exchange releases, press releases, or annual reports

© Comptel Corporation 2012

Major Opportunities

1. Internet of things
   - Data privacy is not as big issue as with consumer analytics
   - Critical alarm prediction case example
2. Open source technologies
   - Access to use and contribute analytics tools
   - Growing fast
3. Other forms of collaboration
   - Joint projects between university, analytics vendor and customer
   - Aalto ES Code It competition
   - Trainee programs
   - Thesis work

© Comptel Corporation 2014
JOINT CONFIDENTIAL

Internet of Things

SENSORS

We are giving our world a digital nervous system

Acceleration/tilt, electric/magnetic, leaks / levels, flow, humidity / moisture, temperature, chemical / gas, motion / velocity / displacement…

CONNECTIVITY

These inputs are digitalized and placed onto networks

LTE, 3G - GPS / GPRS, 2G / GSM , WiFi, Bluetooth, Z-wave, Zigbee, NFC, ANT, RFID, Powerline, Ethernet, Printed, WAN, MAN, LAN, PAN…

PEOPLE & PROCESSES

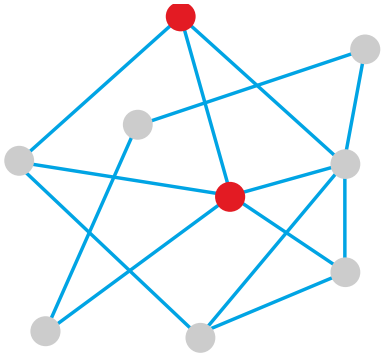
The data flow is combined into bi-directional systems that integrate data, people, processes and systems for better decision making

© Comptel Corporation 2014

JOINT CONFIDENTIAL

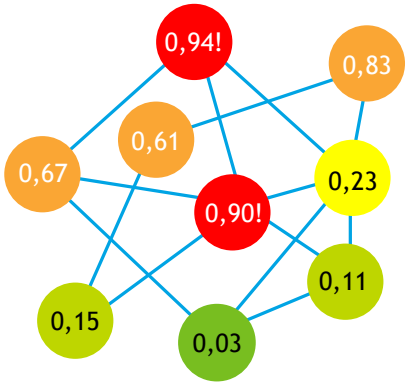
Know Your Network – and What Will Happen

ANALYSE DATA



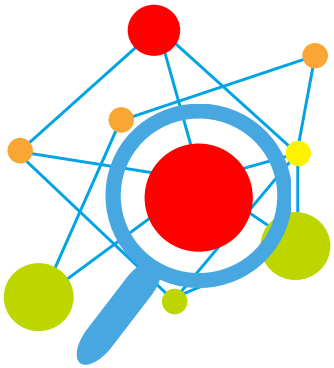
Input
Real-time and historical data of every network element.

PREDICT ALARMS



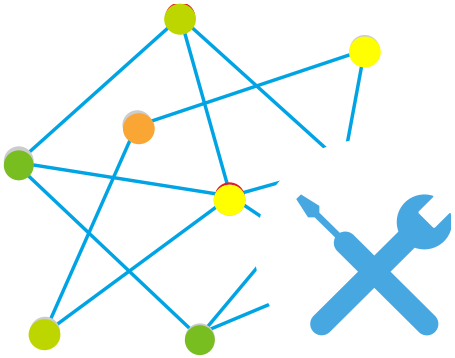
Output
Automated detection through machine learning, with propensity scores assigned for all elements.

KNOW MORE



Insights
Individual element level information through root cause analysis.

OPTIMISE PRACTICES



Outcome
Ability to prevent failures before they occur and automate processes increasingly.

© Comptel Corporation 2014

Plan to Release Part of the Comptel Analytics Product in an Open Source Repository

Tool to Automate
Big Data Analytics

© Comptel Corporation 2014
JOINT CONFIDENTIAL

Comwort: Comptel Workflow Tool

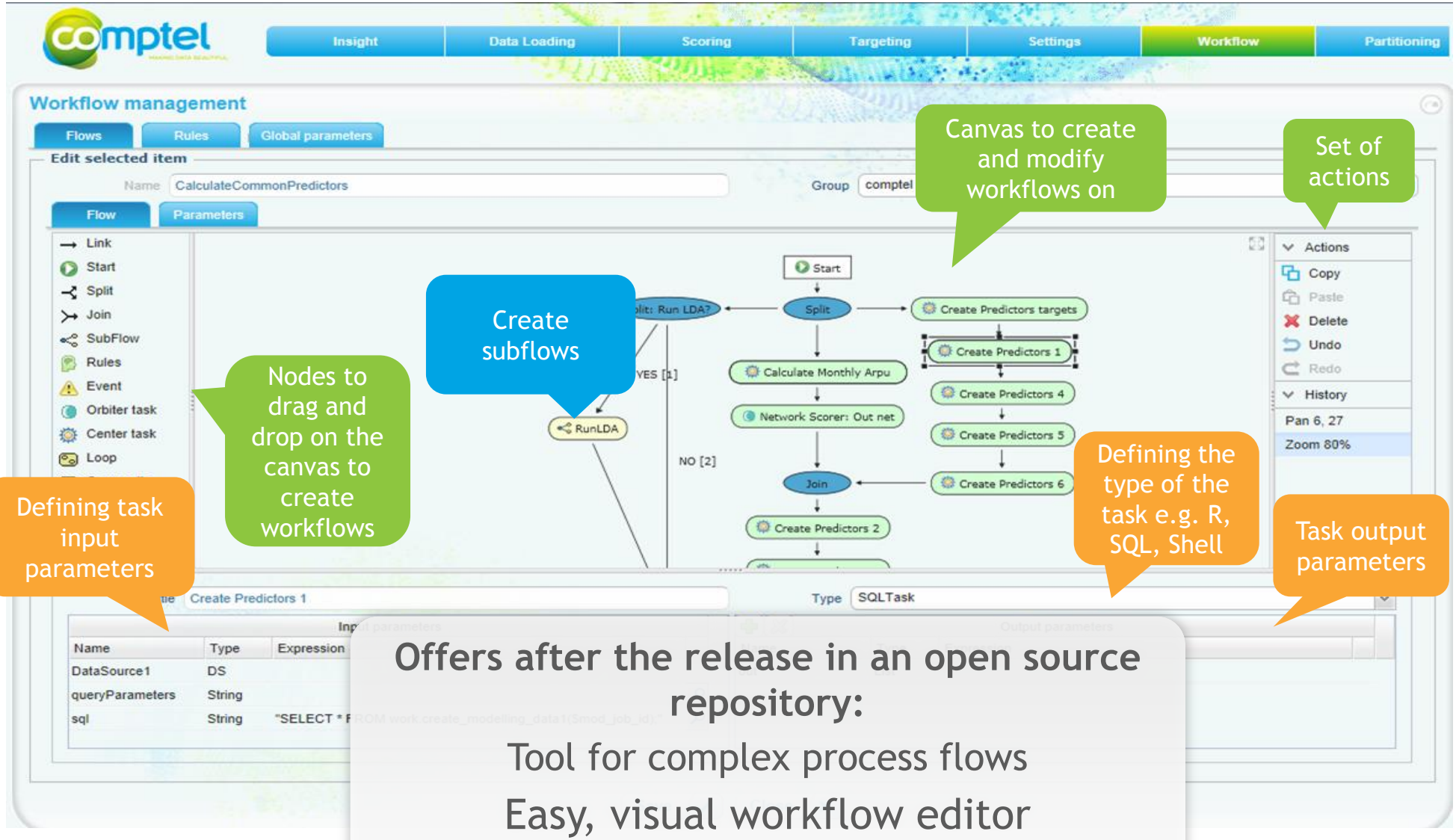
"We love R.
We love Hadoop and SQL.
We love shell scripts.

We love also many other excellent tools
that solve specific problems well.

And we love to combine all of them to automate daily operations where everything should work smoothly."

© Comptel Corporation 2014
JOINT CONFIDENTIAL

Workflow editing environment



Offers after the release in an open source repository:

Tool for complex process flows

Easy, visual workflow editor

Excellent parameter handling capabilities

© Comptel Corporation 2014
JOINT CONFIDENTIAL

Other Forms of Collaboration

Trainee programs

Thesis work

Joint projects between university, analytics vendor and customer

Aalto ES Code It competition

© Comptel Corporation 2014
JOINT CONFIDENTIAL

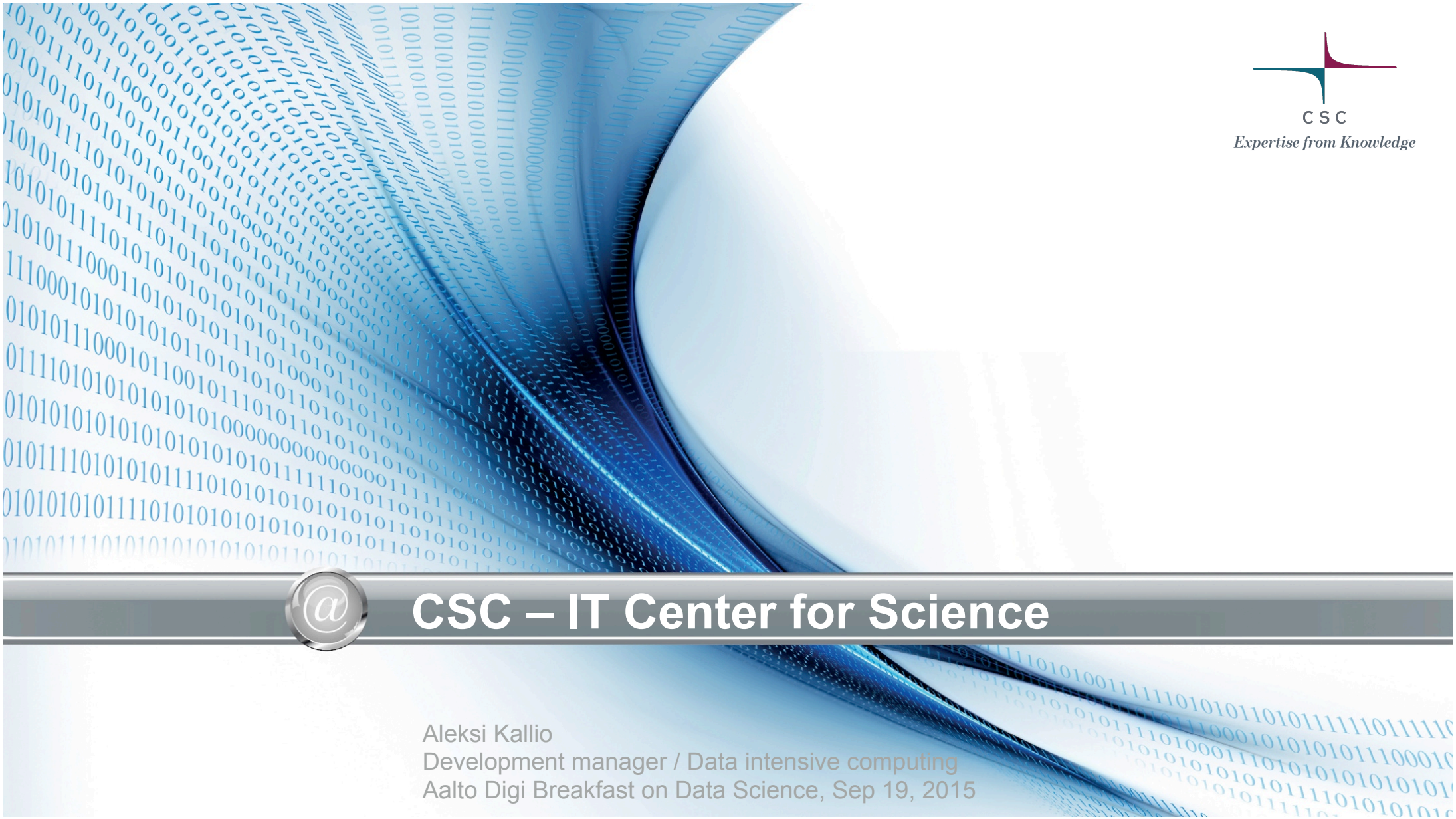
company spotlights

Aleksi Kallio, CSC

Harri Valpola, ZenRobotics

Hannes Heikinheimo, Reaktor

Markus Virtanen, Elisa



CSC
Expertise from Knowledge

CSC – IT Center for Science

Aleksi Kallio
Development manager / Data intensive computing
Aalto Digi Breakfast on Data Science, Sep 19, 2015

CSC in nutshell

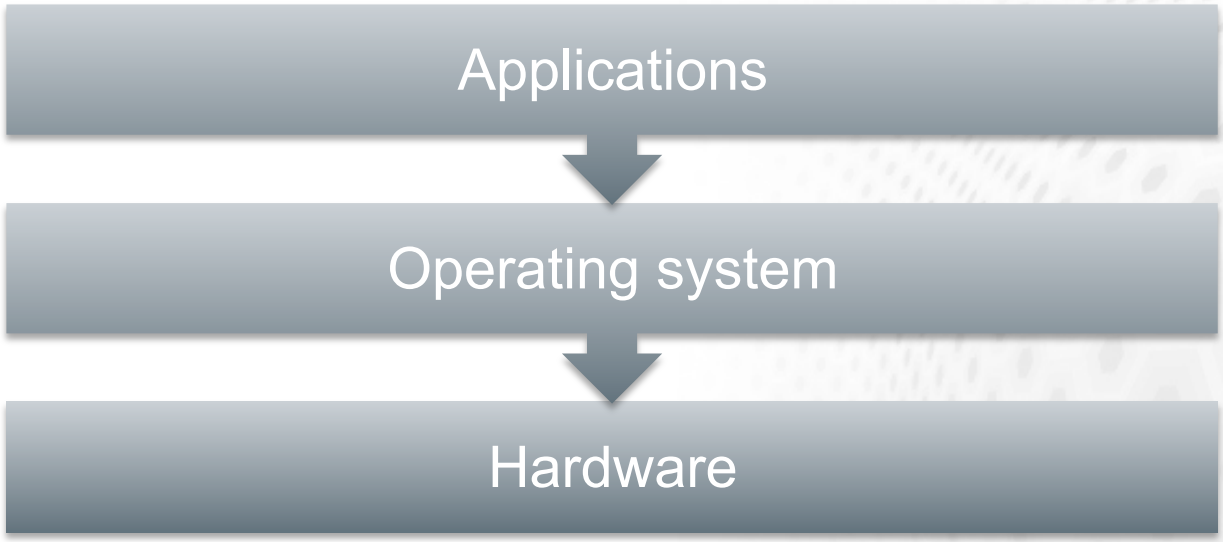
- IT services for universities, polytechnics, governmental institutes and companies
- Services are in general free of charge for academic users in Finland
- Participates in many international projects (such as EUDAT and RDA)

1. COMPUTING AT CSC?

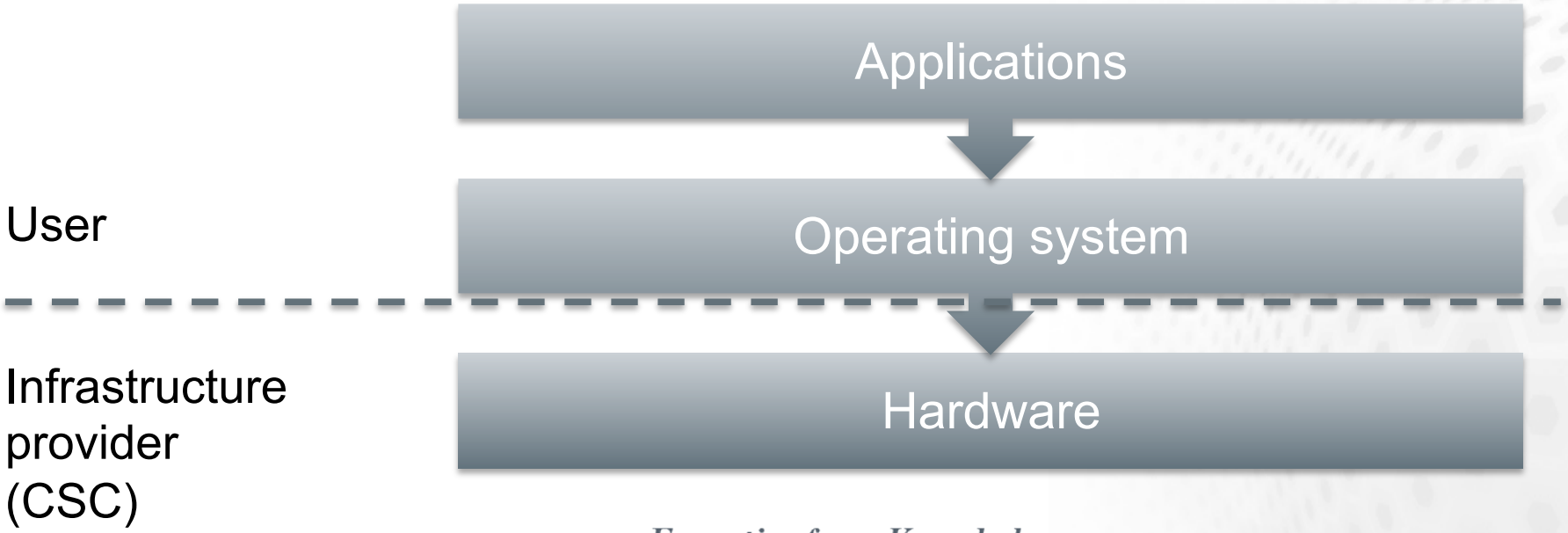
Traditional computing services

User

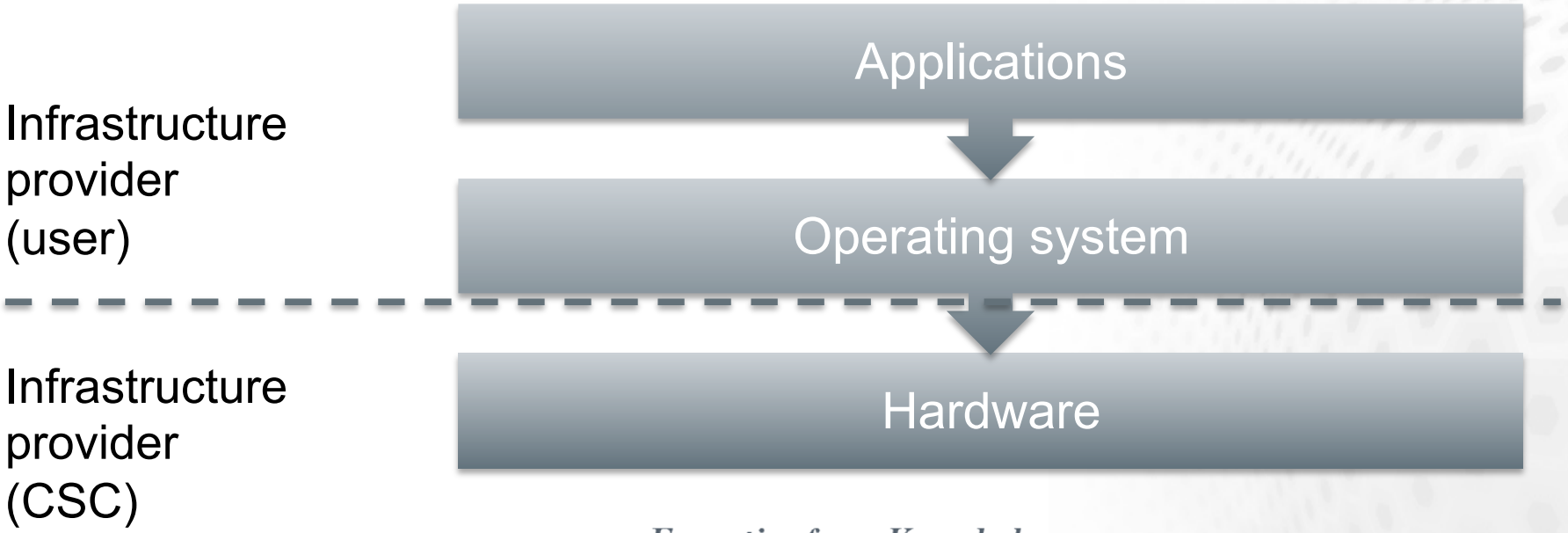
Infrastructure provider (CSC)



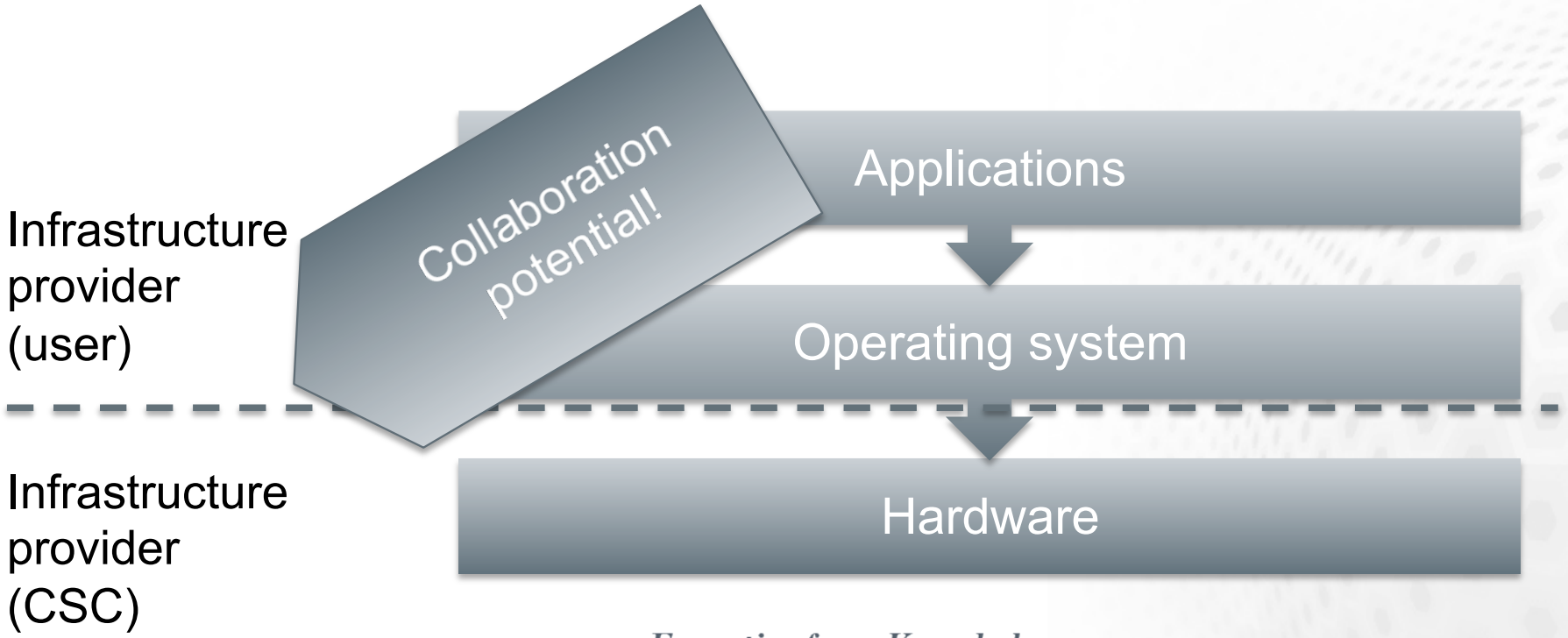
Cloud computing services (Pouta)



Cloud computing services (Pouta)



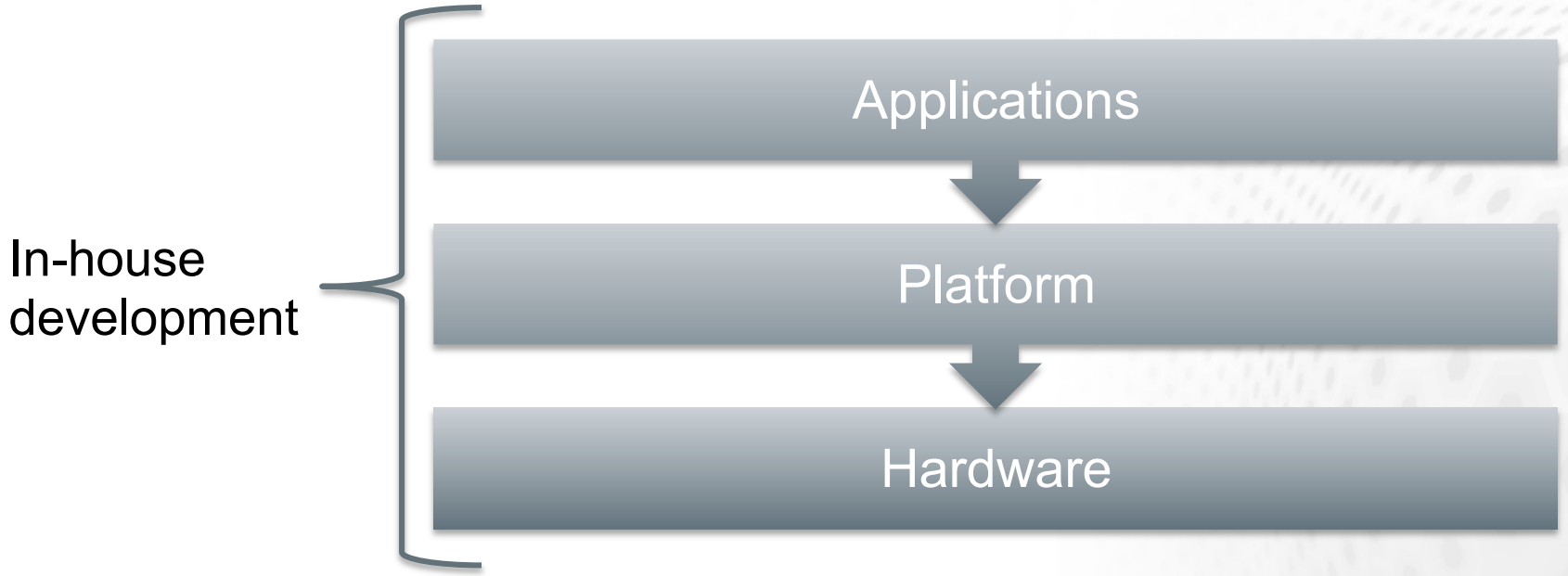
Cloud computing services



2. DATA SCIENCE AT CSC?

WE ARE RAMPING UP SERVICES

Development at all levels



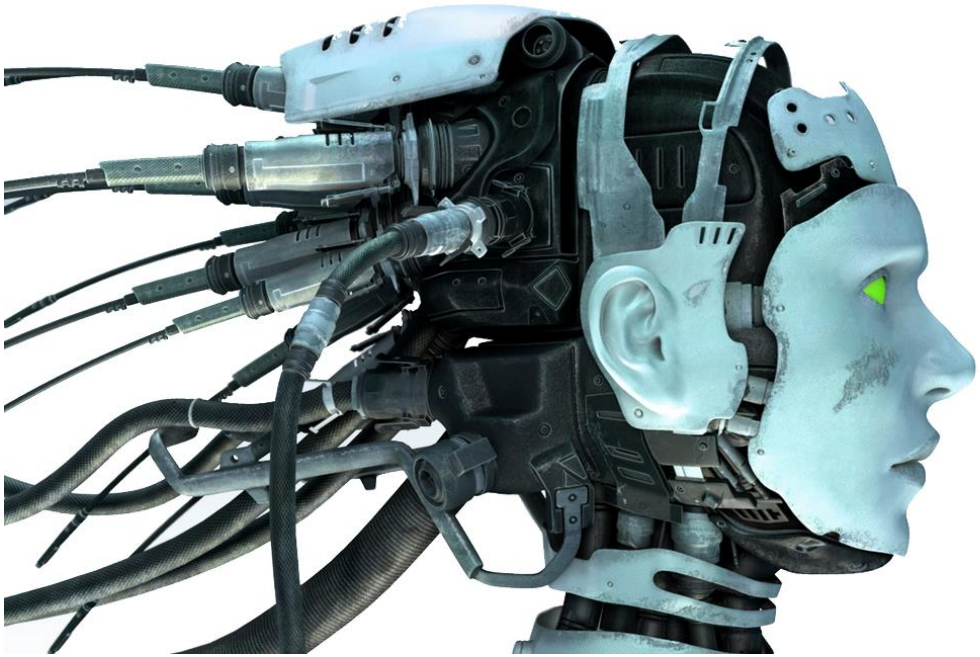
Collaboration opportunities

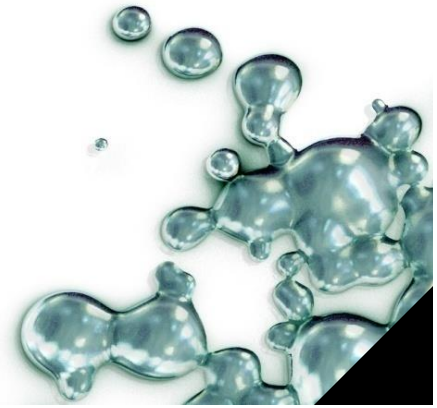
- Active in-house development on the area
- Looking for pilot users, co-projects etc.
- Does not need to be big or heavy (“fork our Hadoop installer at Github”)

ZENROBOTICS®

Building Brains for Robots

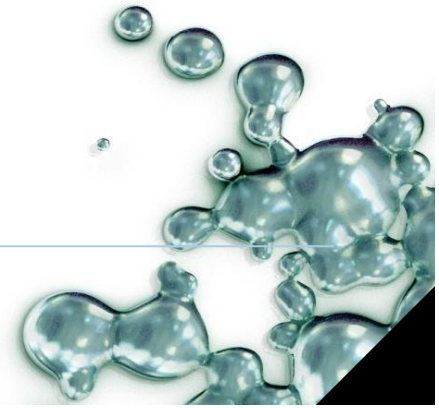
Harri Valpola





ZenRobotics

- Founded in 2007, spin-off of Harri Valpola's neurorobotics group
- Machine learning + robotics = robots that cope in the real world
- First application area: waste sorting
- Currently around 50 employees
- First fully robotic waste sorting plant opened in Viikki, Helsinki, in May 2014



Collaboration with Aalto

- Robots are intelligent because of machine learning (both supervised and reinforcement learning)
- Collaboration starting with Aalto (Jorma Laaksonen's group) regarding object recognition
- **We are looking for a student to work in that project**
- ZenRobotics is also developing new AI
- Currently active research in deep learning, collaboration with Aalto (Tapani Raiko's group) and University of Helsinki

Reaktor

Hannes Heikinheimo

hannes.heikinheimo@reaktor.fi
+358 50 357 4441

19.9.2014

Data Science at Elisa

Aalto digi-breakfast 19 Sep 2014

Markus Virtanen

elisa

siinä on ideaa.

Data Science at Elisa

€ **1.55** billion
revenue
(2012: 1,55)

€ **255** million
profit before tax
(2012: 269)

4.54 million
mobile subscriptions, Finland, Estonia
(2012: 4,45)

565,700
fixed broadband subscriptions
(2012: 505 100)

2.3 million
customers

220,000
shareholders

4,200
Elisa employees in 46 locations

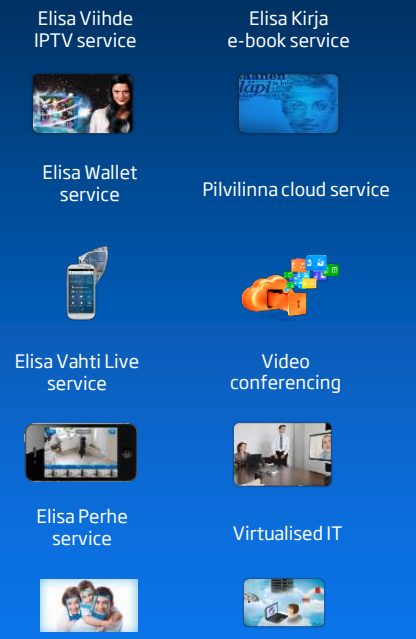
€ **202** million
capital expenditure investments

- Markus?
  - Elisa 2011- (sw, ci, 3gpp), 2013- data analytics
  - Working mainly with spatial and network related data, helping other units with IPTV and customer survey data
- Elisa - big company with lots of real data
  - Network (2G/3G/4G/IP elements, bandwidth allocation, infra invest optimization, call quality drops within frequency/cell changes, closer to real-time SOC alarms)
  - Services (order & delivery processes, IPTV, server logs, predictive maintenance, problem tickets)
  - CRM (business analytics, new product development, marketing impacts/segmentation, churn, recommendation engines, customer surveys and feedback systems)

markus.virtanen@elisa.fi

Collaboration

- Sanitized non-public datasets for partners, e.g. mobile network services
- Thesis writers
- Data visualization competition with smaller datasets
- Help with Elisa-related external data
  - General opinion from social media, news, search engines, web forums, etc.



New services

markus.virtanen@elisa.fi

questions and discussion

what is next?