FAKULTÄT
FÜR !NFORMATIK

Faculty of Informatics

Conquering Data: The State of Play in Intelligent Data Analytics

Allan Hanbury
Institut für Softwaretechnik und interaktive Systeme
TU Wien

Conquering Data

Identifying the data challenges of the future

Conquering Data:

The State of Play in Intelligent Data Analytics

Ralf Bierig, Florina Piroi, Mihai Lupu, Allan Hanbury
Institute of Software Technology and Interactive Systems, Vienna University of Technology, Austria
Helmut Berger, Michael Dittenbach, Marita Haas
max.recall information systems GmbH, Vienna, Austria

12 September 2013

This paper is the basis for the creation of a research roadmap for *Conquering Data* in Austria. It summarises techniques for Intelligent Data Analytics, presents applications in which the analysis of data has the potential to generate significant value, discusses research challenges in the Intelligent Data Analytics area, and presents an Intelligent Data Analytics Competence Landscape of Austria. Further information on this work is available online at `http://conqueringdata.at/`

Contents

1 Introduction 2

2 Conquering Data with Intelligent Data Analytics 2
2.1 Search and Analysis 3
2.2 Semantic Processing 3
2.3 Cognitive Systems and Prediction . . 4
2.4 Visualisation and Representation . . 5
2.5 Evaluation and Benchmarking 5

3 Intelligent Data Analytics Applications 5
3.1 Commerce and Retail 5
3.2 Manufacturing and Logistics 6
3.3 Transportation and Travel 6
3.4 Tourism and Hospitality 7
3.5 Healthcare 7
3.6 Finance and Insurance 8
3.7 Public Sector and Government . . . 8
3.8 Education 9
3.9 Telecommunications 9
3.10 Energy and Utilities 9
3.11 Law and Law Enforcement 9
3.12 eScience 10
3.13 Further Application Areas 10

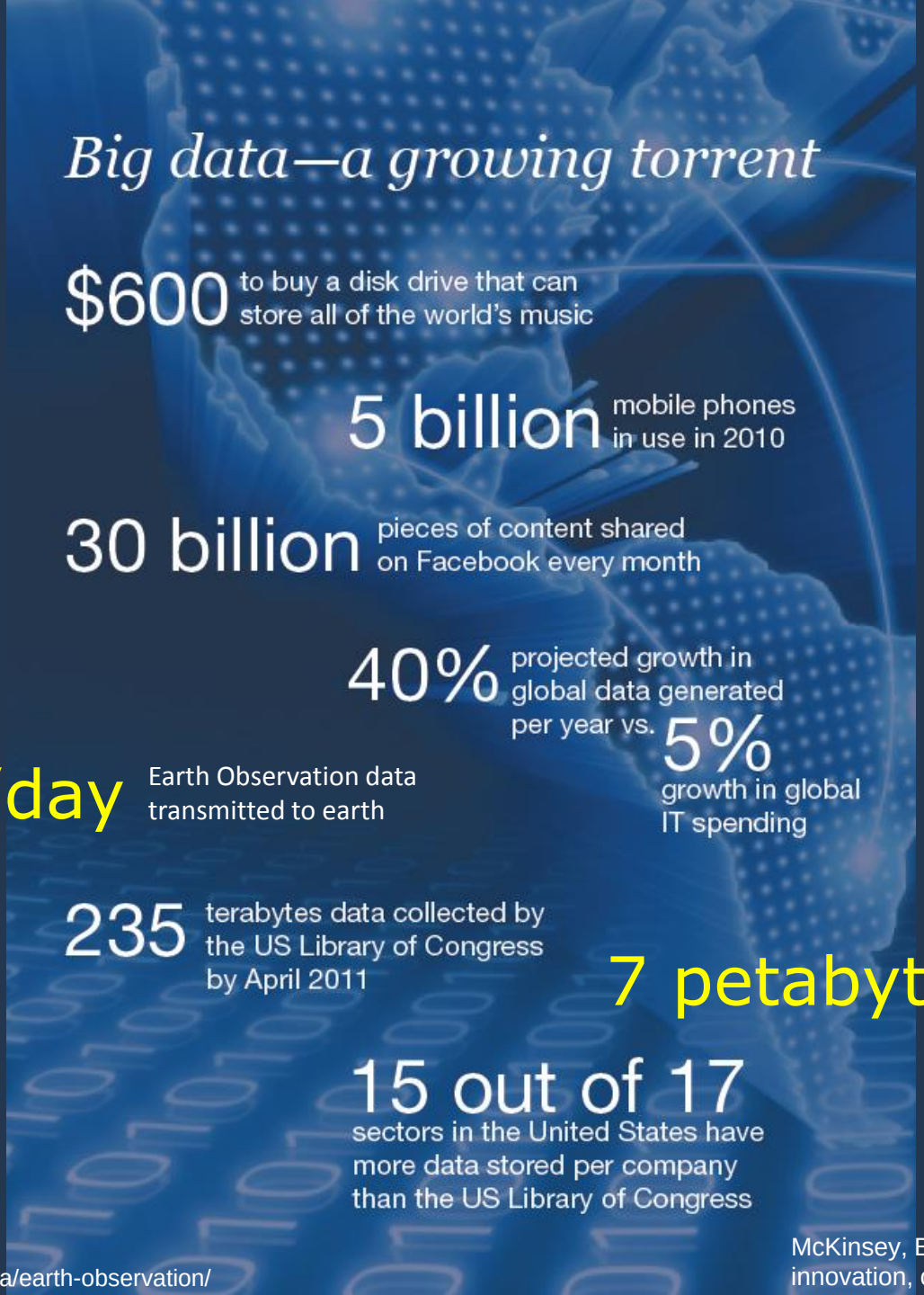
4 Challenges and Open Issues 11
4.1 Privacy and Security 12
4.2 Algorithm and Data Issues 13
4.3 Evaluation 14
4.4 Data Ownership and Open Data . . 14
4.5 Data Curation and Preservation . . 15
4.6 Qualified Personnel 15

5 Austrian Intelligent Data Analytics Competence Landscape 16
5.1 Research Landscape 16
5.2 Service Providers 18
5.3 End Users 19

6 Conclusion 19

http://conqueringdata.at

FAKULTÄT FÜR !NFORMATIK

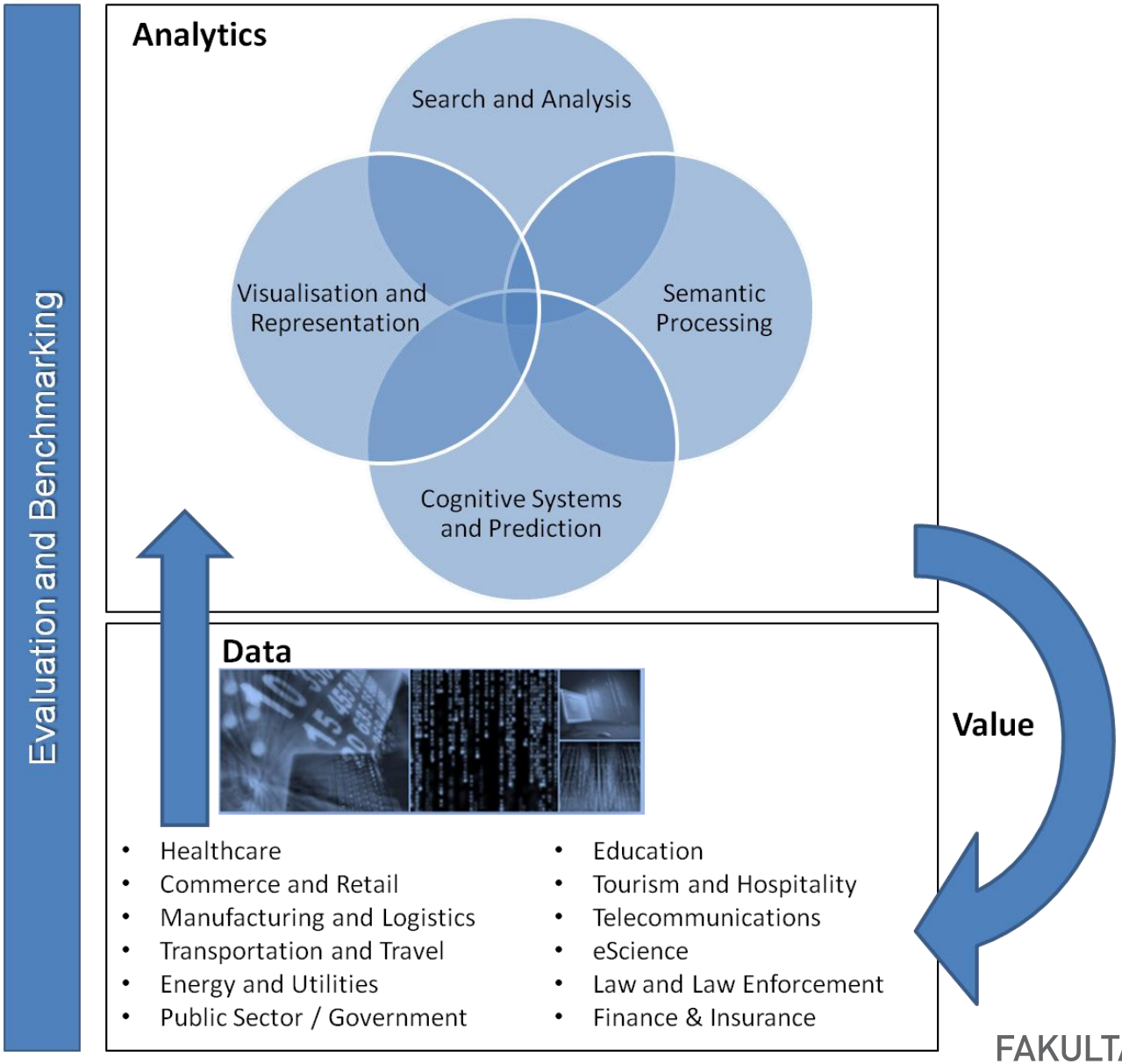


Terabytes/day Earth Observation data transmitted to earth

7 petabytes Earth Observation data in ESA archives (2012)

http://www.scidip-es.eu/science-data/earth-observation/

McKinsey, Big data: The next frontier for innovation, competition, and productivity, 2011


Evaluation and Benchmarking
Analytics
Search and Analysis
Visualisation and Representation
Semantic Processing
Cognitive Systems and Prediction
Data
Value
Healthcare
Commerce and Retail
Manufacturing and Logistics
Transportation and Travel
Energy and Utilities
Public Sector / Government
Education
Tourism and Hospitality
Telecommunications
eScience
Law and Law Enforcement
Finance & Insurance

Intelligent Data Analytics Research Fields

1 Search and Analysis

- Information Retrieval
- Computer Vision
- Speech Processing
- Process Analysis
- Network Science
- Ubiquitous Computing
- Information Integration
- Statistics
- Data Mining
- Bioinformatics
- Digital Preservation
- Algorithmic Efficiency

2 Semantic Processing

- Information Extraction
- Knowledge Engineering
- Semantic Web
- Natural Language Processing

3 Cognitive Systems and Prediction

- Machine Learning
- Computational Neuroscience
- Reasoning
- Recommender Systems
- Decision-Support System
- Simulation
- Crowdsourcing
- Brain-Computer-Interface

4 Visualisation and Interaction

- Visualisation
- Visual Analytics
- Rendering
- Sonification

5 Evaluation and Benchmarking

Challenges

Data Fusion and Data Integration

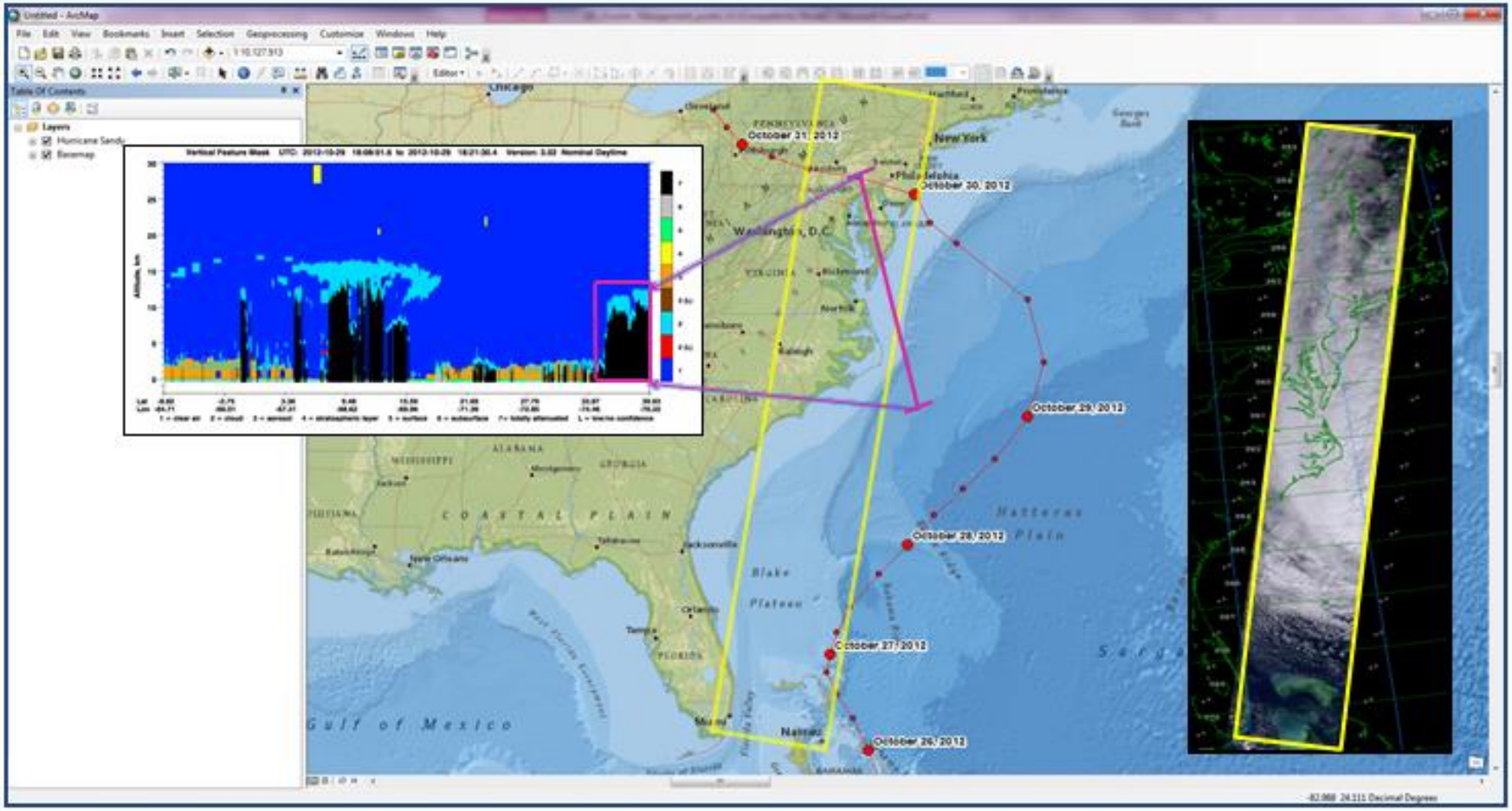


Figure 2. Produced by ASDC data, this map shows the track of Hurricane Sandy in 2012 with MISR and CALIPSO data products being highlighted in yellow and purple, respectively. Image Credit: Emily Northup & Matthew Tisdale.

http://www.earthzine.org/2013/10/18/using-big-data-in-geographic-information-systems-for-observing-earths-climate/

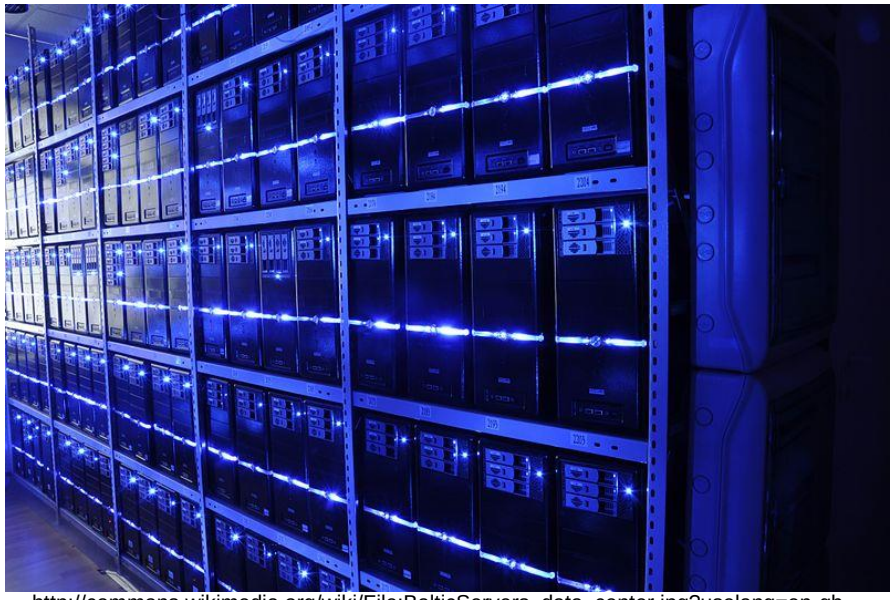
Experimental Infrastructure

- Which algorithm / method is best for a specific problem?
- Sharing data / algorithms / workflows
- Computational power for large datasets

HELIX NEBULA THE SCIENCE CLOUD

GEO GROUP ON EARTH OBSERVATIONS

Geohazard Supersites & Natural Laboratories



http://commons.wikimedia.org/wiki/File:BalticServers_data_center.jpg?uselang=en-gb

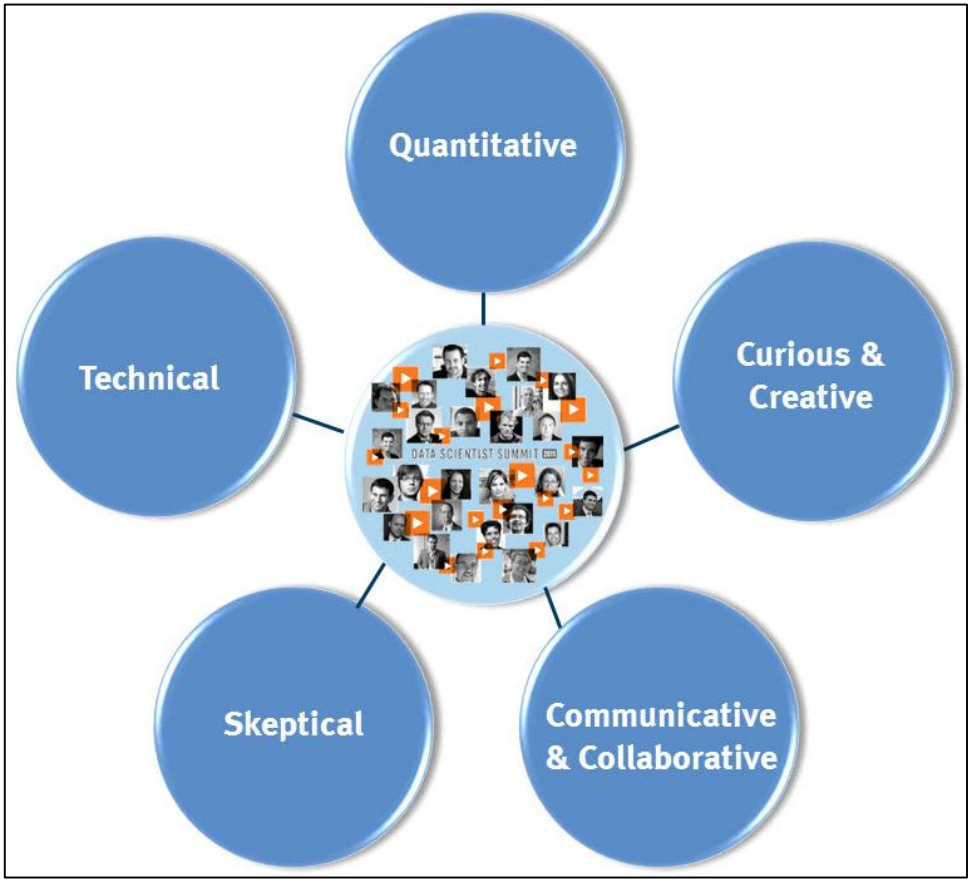
Privacy and Security

Google Streetview

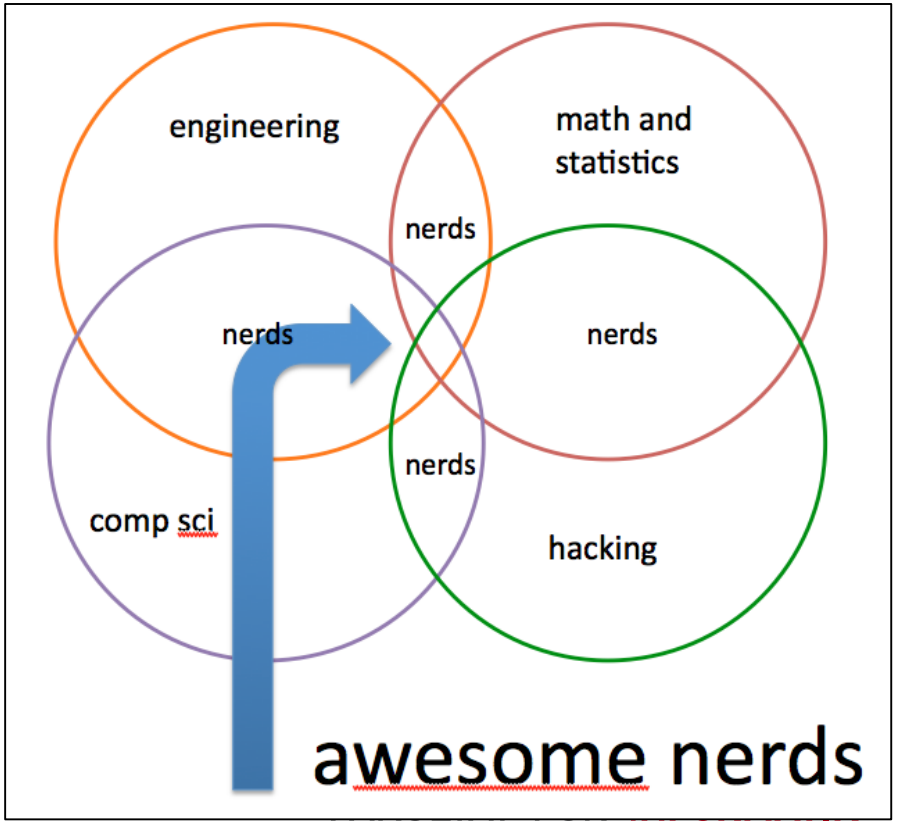
Qualified personnel

- Data scientists

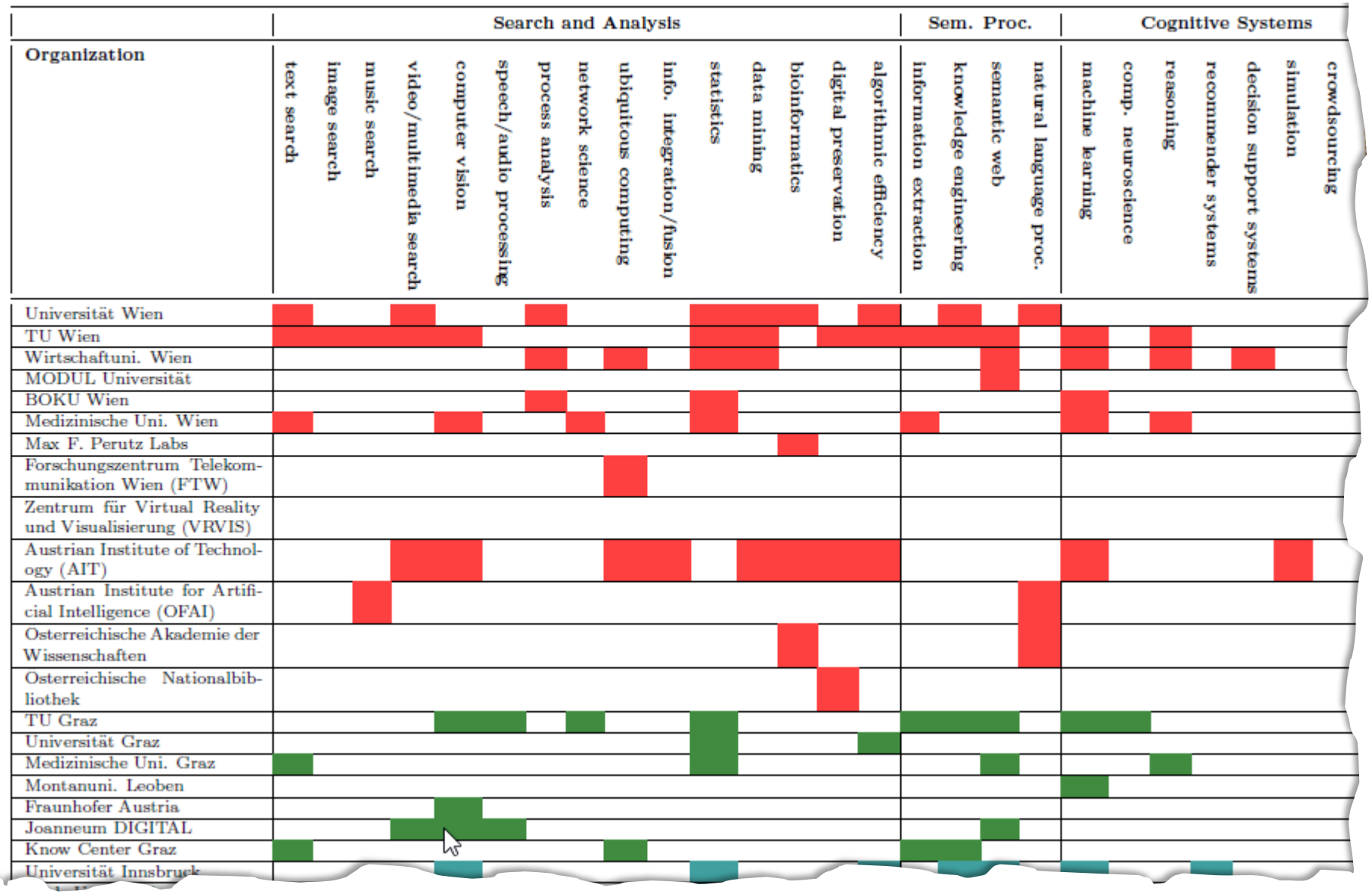
Data Scientist: The Sexiest Job of the 21st Century



https://infocus.emc.com/david_dietrich/a-data-scientist-view-of-the-world-or-the-world-is-your-petri-dish/



http://www.forbes.com/sites/danwoods/2012/03/08/hilary-mason-what-is-a-data-scientist/


Search and Analysis
Sem. Proc.
Cognitive Systems
Organization
text search
image search
music search
video/multimedia search
computer vision
speech/audio processing
process analysis
network science
ubiquitous computing
info. integration/fusion
statistics
data mining
bioinformatics
digital preservation
algorithmic efficiency
information extraction
knowledge engineering
semantic web
natural language proc.
machine learning
comp. neuroscience
reasoning
recommender systems
decision support systems
simulation
crowdsourcing
Universität Wien
TU Wien
Wirtschaftuni. Wien
MODUL Universität
BOKU Wien
Medizinische Uni. Wien
Max F. Perutz Labs
Forschungszentrum Telekommunikation Wien (FTW)
Zentrum für Virtual Reality und Visualisierung (VRVIS)
Austrian Institute of Technology (AIT)
Austrian Institute for Artificial Intelligence (OFAI)
Osterreichische Akademie der Wissenschaften
Osterreichische Nationalbibliothek
TU Graz
Universität Graz
Medizinische Uni. Graz
Montanuni. Leoben
Fraunhofer Austria
Joanneum DIGITAL
Know Center Graz
Universität Innsbruck

Die Studie

- Durchgeführt von:
  - max.recall information systems GmbH
  - Institut für Softwaretechnik und Interaktive Systeme, TU Wien
- http://conqueringdata.at
- Die Studie wird Mitte Jänner 2014 veröffentlicht

Dr. Allan Hanbury

Institut für Softwaretechnik und interaktive Systeme
TU Wien
Favoritenstr. 9-11/188
1040 Wien

Telefon: +43 1 58801 188310
Mobil: +43 676 978 0991

e-Mail: hanbury@ifs.tuwien.ac.at