Contents

[Data science institutes information on definitions/goals 2](#_Toc437608532)

[Diagrams 7](#_Toc437608533)

[People 12](#_Toc437608534)

[Notes Data Science/Big Data literature 17](#_Toc437608535)

[Notes Big Data & Society Publications 20](#_Toc437608536)

[Notes Theory, Culture, Society Publications 22](#_Toc437608537)

# Data science institutes information on definitions/goals

**University College London, Data Science Institute**

The following projects are currently underway at the UCL Big Data Institute:  
1) Investigation of the role and impact of researchers in academic networks  
2) Investigation of relationships between researchers on academic networks  
3) Models of citations and downloads as heterogeneous count time series.  
4) Prediction of functionality of novel proteins.  
5) Adaptive User Modelling for Personalized Experience.

**Eindhoven**

The center is run as a doctoral school with a scientific research program consisting of seven major research programs:

Process analytics   
Customer journey  
Smart maintenance & diagnostics  
Quantified self  
Data value and privacy  
Smart cities  
Smart grids

The Data Science Center Eindhoven (DSC/e) is TU/e’s response to the growing volume and importance of data. 90% of the data in the world today has been created in the last two years alone and the world's data will grow by 50 times in the next 10 years. Moreover, human and organizational activities are intertwined with the digital universe. Therefore, data science is growing in importance and becoming an integral part of different types of engineering and scientific research.

**Delft**

Delft Data Science is TU Delft’s framework initiative for research activities in data science. Delft Data Science, or DDS, is front-end for research, as catalyser for research foci and scientific impact and visibility, and front-end for education and valorisation, for societal impact and visibility.  
DDS synergises different fields of expertise, encompassing various disciplinary research areas and selected application domains. DDS is characterised by the focus on the engineering aspects of data science, in line with TU Delft’s key strengths.

In terms of the scientific disciplines involved, DDS combines interdisciplinary expertise and skills covering all the key dimensions of the data science field, with scientists involved in research around ‘hardware for big data’, ‘software for big data’, ‘data management for big data’, ‘data interaction and visualisation’, and ‘social and human centered data analysis’. This way, TU Delft is able to address the needs related to Big Data sense-making in a large variety of problems faced by its societal, industrial partners, and stakeholders over a wide range of domains. TU Delft and Delft Data Science have successful collaborations with leading industries active in the data science field and a solid embedding in the EU and national R&D context.

**Amsterdam**

We live in a world in which data are generated in ever increasing quantity and information derived from it has become a main driving force for scientific discovery and innovation.

We can only fully employ these vast amounts of data if we have the methodologies to store and process it and turn it into valuable and accessible information by analysis and modeling, leading to understanding and a basis for informed decisions. Society is becoming increasingly reliant on data and the tools and methods to acquire and analyze it. Important sources of data that are only starting to be explored come from social media, on-line full text science literature, on-line video material, on-line click and interaction patterns, financial transactions, customer behavior, sensors and scientific instrumentation. Being able to employ such data will improve the efficiency of our government, the quality of our health care, steer innovation in business, the effectiveness of our military and last but not least, catalyze new discoveries in science.

**Warwick**

WDSI brings together not only the huge strengths of Warwick's Departments of Computer Science, Mathematics and Statistics, but also the wider community both at Warwick (researchers from many departments, with interests ranging widely from genomics to business analytics) and internationally through the extensive WDSI programme of research workshops and visitors.

The primary focus of WDSI is world-leading research, and especially the development of novel mathematical, statistical and computational approaches to the acquisition, management and analysis of "big data". Warwick is a founding partner (with Cambridge, Edinburgh, Oxford and UCL) in the newly established[Alan Turing Institute](https://turing.ac.uk/) for data science.

**Manchester**

Manchester's Data Science Institute acts as an access point to the University’s expertise in data science, facilitates interactions between data science researchers and problem holders, owns the University’s data science strategy, and will deliver sustainable support for the community.

Manchester has an engaged data science community of almost 250 investigators, with methodologists embedded in Schools across the University addressing problems in extracting meaning from data, managing data volume, the variety of data used in analyses, the velocity with which it is produced and the veracity of those data.

Data science has a home in all four of the University's faculties (Engineering and Physical Sciences, Humanities, Life Sciences and Medical and Human Sciences) supported across the whole data life cycle by work in the schools of Computer Science and Mathematics. From information management, through analytics, to practical applications. This creates a virtuous circle, where challenging real-world problems drive the methodology research agenda, whilst providing a natural driver for building new algorithms and methods.

**Edinburgh**

Edinburgh Data Science acts as a hub for national and international data across the arts and sciences; developing networks of expertise in managing and accessing all forms of data.

* Within the University of Edinburgh, data science is viewed as a major strength and strategic opportunity across disciplines.
* Our immersion in data makes the distance between observation and conclusion shorter and more immediate. This fundamentally changes our view of academic study; it allows experiments that were impossible to become practical and it allows radically new systems to be engineered. We shape these new forms of study and engineer these systems.
* Social sharing of data creates new and unexpected insights across traditional academic areas and we are building the expertise and social structure to develop this potential.

**Dundee**

There has been a recent upsurge of commercial interest in both Business Intelligence (BI) and Data Science (DS).

BI is a relatively broad topic, so a full description would be verbose but in general terms, BI is about extracting useful information from a mass of raw data.  A large number of systems, techniques and processes can all be involved in doing this – data warehousing, dimensional modeling and so on.  In addition the inevitable host of abbreviations and acronyms are involved – ETL, OLAP, MDX.

DS has much in common with BI but Data Scientists tends to work more with ‘big data’ and have a greater focus on developing custom algorithms and visualizations.  One good definition of a Data Scientist is that they know more statistics than a programmer and more about programming than a statistician.  The term data scientist doesn’t imply that scientific data is involved, although it certainly can be; most data scientists work on commercial data.

Why Dundee so early? > Because Mark Whitehorn (worked also with IBM)

**University of Essex**

We conduct research into issues and challenges in data exploration and big data from a multitude of perspectives. We drive breakthroughs and innovation in a range of areas, including:

* technologies for management and transfer of big data
* methodological and analytical methods for different types of applications (from financial and business to biomedical)
* socio-economic aspects of data
* ethical, legal and human rights aspects of data

We're currently engaged in a range of projects on data exploration and big data, including:

* innovative tools to enable exploration of complex and specialised data sets
* data analytics driven by ontologies
* pattern discovery in big data research streams
* modelling the high frequency foreign exchange market

**Cambridge**

The Cambridge Big Data Strategic Research Initiative brings together researchers from across the University to address challenges presented by our access to unprecedented volumes of data. Our research spans all six Schools of the University, from the underlying fundamentals in mathematics and computer science, to applications ranging from astronomy and bioinformatics, to medicine, social science and the humanities

In parallel, our research addresses important issues around law, ethics and economics, in order to apply Big Data to solve challenging problems for society.

**Bournemouth**

The main objectives of the DSI are:

**Interdisciplinary training** of highly skilled and internationally excellent researchers and leaders in the Data Science area covering big data, advanced/predictive analytics, data intensive computing and their innovative business, engineering and science applications, in a cross-disciplinary environment.

Delivery of **internationally leading and excellent research contributions** to methodological, algorithmic, technological and applied research in Data Science with a focus on scalability, automation of typical tasks, adaptivity and learning capability of developed methods, tools and products.

Provision of a **unique highly stimulating environment** and opportunities for making substantial academic, societal and commercial impact as well as developing awareness and breadth of knowledge for both students and staff by assembling a multidisciplinary and cross-sectoral team of academic, commercial and public institutions.

As shown below we have structured the DSI into four overlapping themes and highlight the key challenges within each of them together with the names of the major contributing groups, centres and institutes from the University whose data intensive activities are co-ordinated and represented within DSI:

* **Core Data Science Technologies** covering the underpinning mathematical, statistical and computer science expertise and research concerned with new scalable approaches for capturing, storing, managing, analysing and visualising large-scale, complex and diverse data from multiple sources ([Smart Technology Research Centre](https://research.bournemouth.ac.uk/centre/smart-technology-research-centre/); [National Centre for Computer Animation](https://research.bournemouth.ac.uk/centre/ncca/); Computing and Informatics Research Centre; [Statistical and Mathematical Support Centre](https://microsites.bournemouth.ac.uk/mathssupport/));
* **Data Intensive Sciences** focusing on the fundamentally changing nature of research and scientific enquiry, driven and informed by the deluge of data and emerging data-driven approaches ([Conservation Ecology and Environmental Sciences Centre](https://research.bournemouth.ac.uk/centre/conservation-ecology-and-environmental-sciences-group/); [Psychology Research Centre](https://research.bournemouth.ac.uk/centre/psychology-research-group/); [Centre for Archaeology and Anthropology](https://research.bournemouth.ac.uk/centre/archaeology-and-anthropology-group/); [Institute for Studies in Landscape and Human Evolution](https://research.bournemouth.ac.uk/centre/institute-for-studies-in-landscape-and-human-evolution/));
* **Data-driven Real-World Applications** focusing on challenging business, engineering and public sector applications in data rich and changing environments ([International Centre for Tourism and Hospitality Research](https://research.bournemouth.ac.uk/centre/international-centre-for-tourism-and-hospitality-research/); [Smart Technology Research Centre](https://research.bournemouth.ac.uk/centre/smart-technology-research-centre/); [Creative Technology Research Centre](https://research.bournemouth.ac.uk/centre/creative-technology-research-centre/));
* **Data-driven Digital Entrepreneurship** exploring new business models and “data” products and services ([Centre for Entrepreneurship](http://bucfe.com/); [Centre for Intellectual Property and Policy Management](https://research.bournemouth.ac.uk/centre/centre-for-intellectual-property-policy-management/)).

**Columbia**

The Data Science Institute at Columbia University is training the next generation of data scientists and developing innovative technology to serve society. With faculty from a wide range of disciplines, the Institute seeks to foster collaboration to advance techniques for gathering and interpreting data, and to solve the pressing problems that face society. The Institute works closely with industry to bring promising ideas to market.

More than 150 faculty members from nine schools across Columbia are affiliated with the Data Science Institute:

* [The Fu Foundation School of Engineering and Applied Science](http://engineering.columbia.edu/)
* [Graduate School of Arts and Sciences](http://gsas.columbia.edu/)
* [Columbia Journalism School](http://www.journalism.columbia.edu/)
* [Columbia Business School](http://www8.gsb.columbia.edu/)
* [Graduate School of Architecture, Planning and Preservation](http://www.arch.columbia.edu/)
* [School of International and Public Affairs](http://sipa.columbia.edu/)
* [Columbia University Medical Center](http://cumc.columbia.edu/)
* [Mailman School of Public Health](http://www.mailman.columbia.edu/)
* [Columbia Law School](http://www.law.columbia.edu/)

**Chicago**

A collaboration between the Harris School of Public Policy and the [Computation Institute](http://www.ci.uchicago.edu/), the Center for Data Science and Public Policy brings together data science and public policy experts. Its mission is to conduct research and create computational and data-driven solutions to large-scale social problems in areas such as healthcare, education, sustainability, and community development.

The Center's goals include:

* Creating Educational and Training Programs for students as well as professionals who are in government and non-profit organizations
* Conducting collaborative research with Government and Non-Profit partners
* Developing reusable software tools that can be used and extended by Government and Non-Profit organizations.

**Virginia**

The Data Science Institute is an institute for large scale, complex data analysis. It is a unique confluence of computation, science, engineering, mathematics, statistics, commerce, social science, humanities, law, & more.

Mission:

* To achieve recognized excellence in research and education in the interdisciplinary field of data science.
* Our competitive advantage: An institute that crosses disciplines, departments, schools and colleges to leverage UVA’s combined capabilities in data science.

**Boston**

Everywhere we look, from the way we deliver healthcare to our strategies for sustaining cities, managing supply chains and assessing how students learn, Big Data is transforming our ability like never before to think critically, plan effectively, and quickly make informed decisions on millions of everyday challenges.

Data Science — the methodical extraction of knowledge from data — ties together hundreds of existing disciplines to help scientists, engineers, physicians, professionals and researchers of virtually every background create new efficiencies and design smart solutions in real time.

As a globally recognized leader in interdisciplinary research, Boston University is committed and uniquely positioned to be a forerunner in this rapidly expanding and evolving field.

**Berkeley**

Who we are

Founded in 2013, the Berkeley Institute for Data Science (BIDS) is a central hub of research and education at UC Berkeley designed to facilitate and nurture data-intensive science. People are at the heart of BIDS. We are building a community centered on a cohort of talented data science fellows and senior fellows who are representative of the world-class researchers from across campus and are leading the data science revolution within their disciplines.

How We Work

Our initiatives are designed to bring together broad constituents of the data science community, including domain experts from the life, social, and physical sciences and methodological experts from computer science, statistics, and applied mathematics. While many of these individuals rarely cross professional paths, BIDS actively seeks new and creative ways to engage and foster collaboration across these different research fields.

Six themes, see website.

**Michigan**

The Michigan Institute for Data Science (MIDAS) is the focal point for the new multidisciplinary area of data science at the University of Michigan. This area covers a wide spectrum of scientific pursuits (development of concepts, methods, and technology) for data collection, management, analysis, and interpretation as well as their innovative use to address important problems in science, engineering, business, and other areas.

Data science is now widely accepted as the fourth mode of scientific discovery, on par with theory, physical experimentation and computational analysis. Techniques based on Big Data are showing promise not only in scientific research, but also in education, health, policy, and business.

**University of Massachusetts**

Data Science develops methods to collect and analyze large-scale data, and uses them to make discoveries and decisions > definition

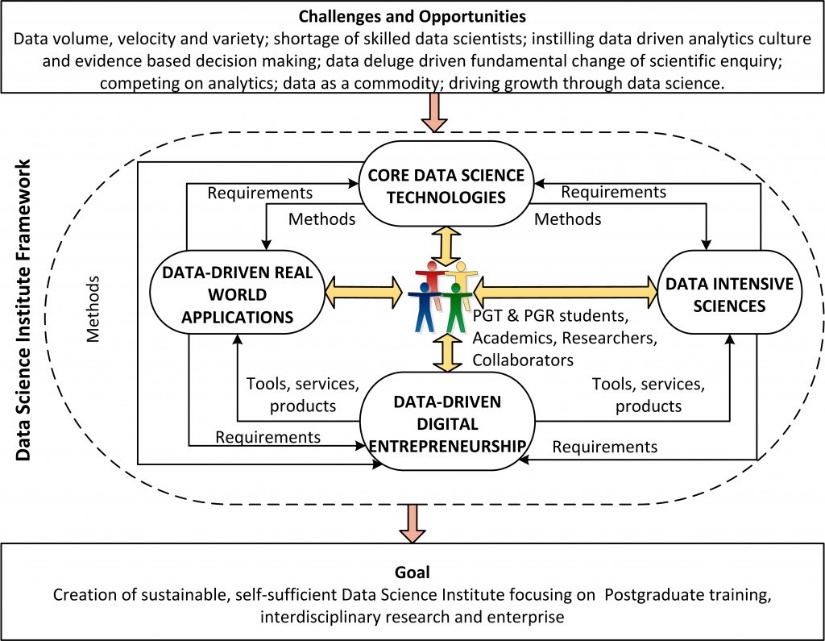
<https://ds.cs.umass.edu/>

**Rochester**

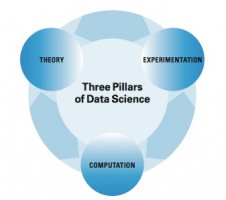
Also has a definition: “Data science is the creation and application of powerful new methods to collect, curate, analyze, and make discoveries from large-scale data”

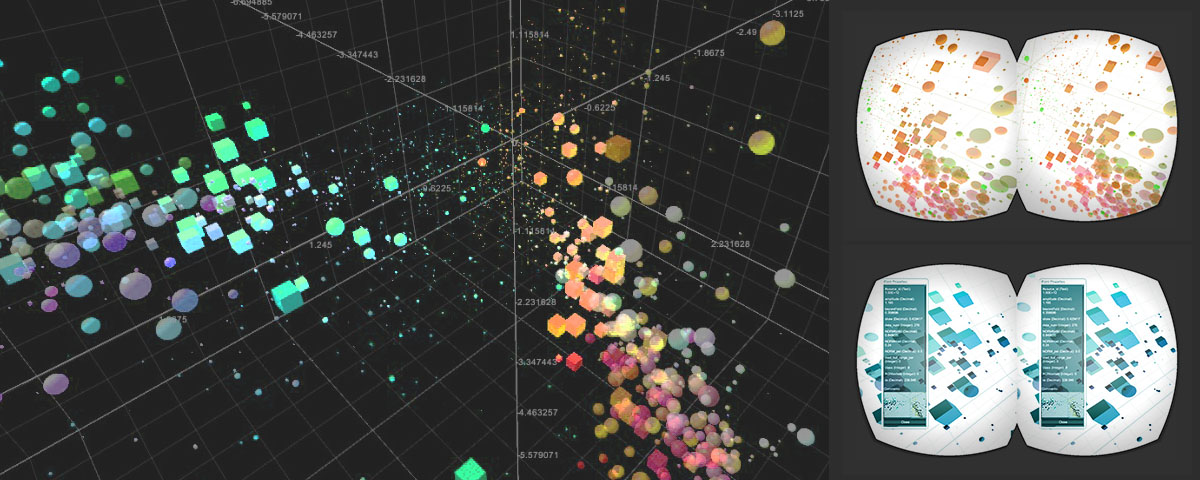
<https://www.rochester.edu/data-science/about/data-science.html>

# Diagrams

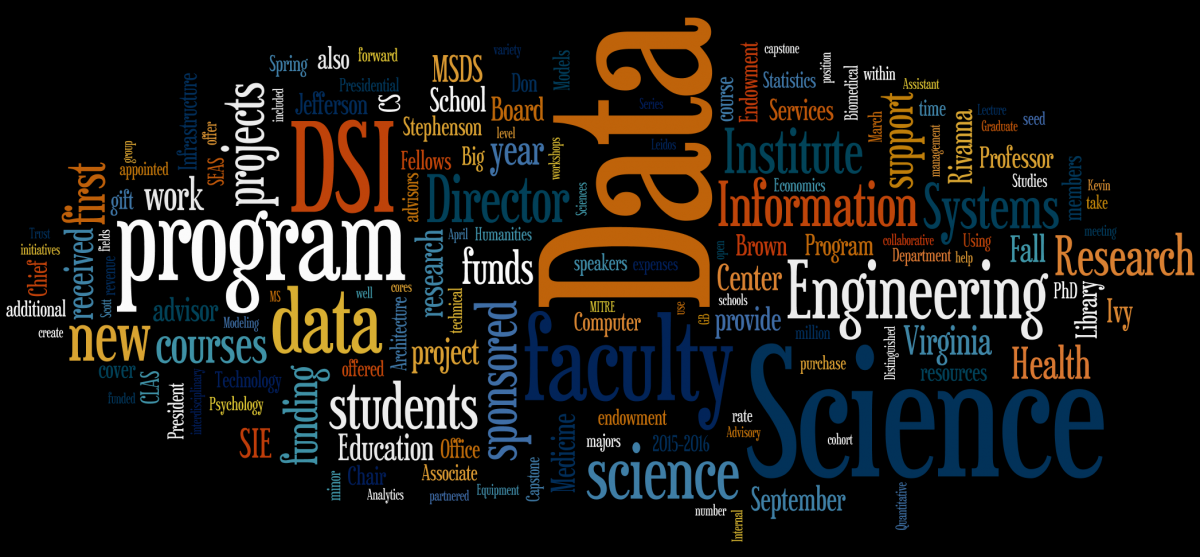
[](https://research.bournemouth.ac.uk/wp-content/uploads/2015/03/DSI_Framework.jpg)Bournemouth:

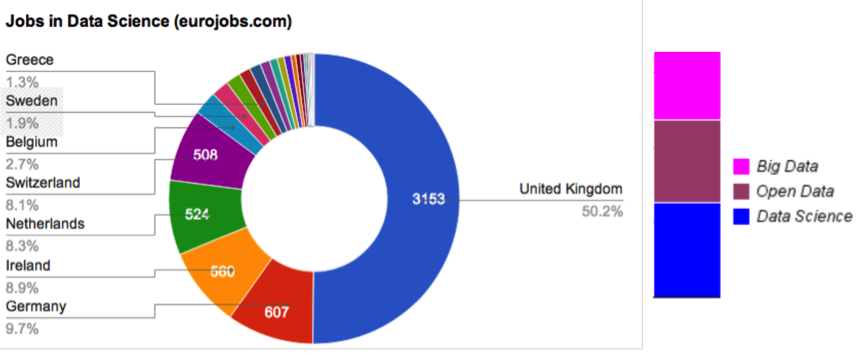
Northwestern University:



Caltech:

Virginia:

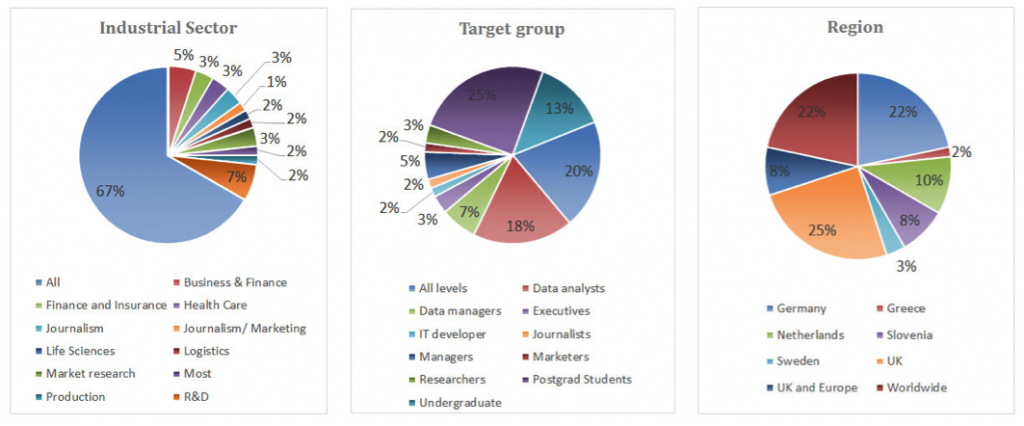
European Data Science Academy:

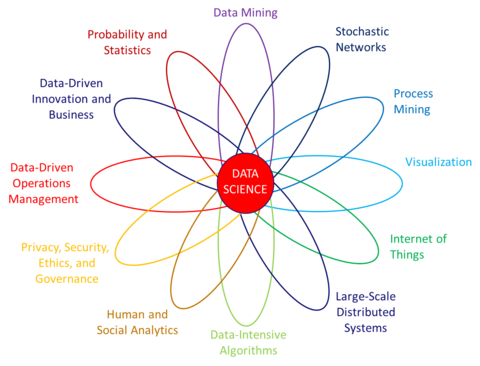


European 2:

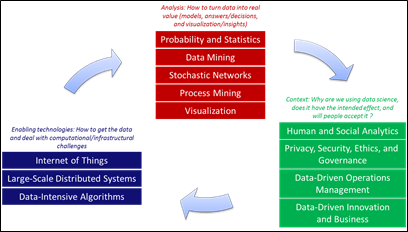


European 3:

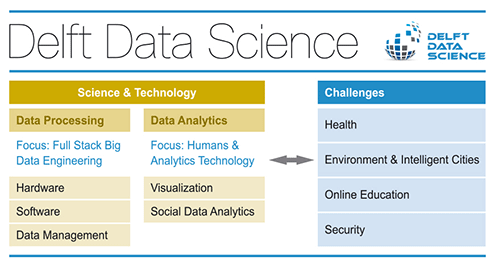


Eindhoven:

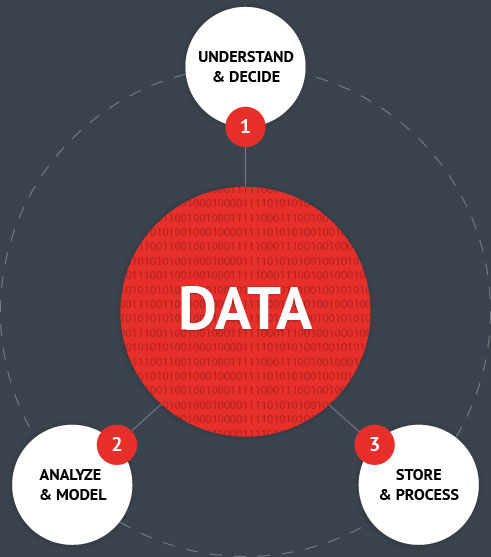
Eindhoven 2:



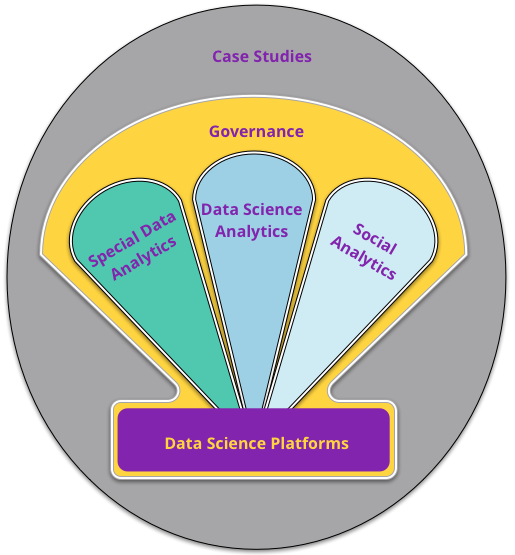
Delft:



Amsterdam:



Manchester:



UCL:



# People

|  |
| --- |
| Amsterdam, VU University & Uni of Amsterdam  Directors   * Maarten de Rijke, computer science * Marcel Worring, data science, business analytics * Henri Bal, computer science * Sanne Abeln, computer science, bioinformatics |
| Ann Arbor, University of Michigan  Directors   * Alfred Hero, engineering * Brian Athey, computational medicine * H.V. Jagadish, engineering/computer science * Vijay Nair, statistics |
| Berkeley, University of California  Staff   * Charlotte Cabasse, geography * Fernando Perez, computer science * Kevin Koy, environmental science * Matthias Bussonnier, biophysics * Nathaniel Smith, computer science * R.S. Geiger, computer science, STS * Saul Perlmutter, physics * Stéfan van der Walt, computer science |
| Boston, Boston University  Too many to list |
| Bournemouth, Bournemouth University  Core members   * Bogdan Gabrys, computational intelligence * Ross Hill, geoinformatics * Richard Stafford, computational ecology * Katarzyna Musial-Gabrys, computer science * Marcin Budka, data science * Venky Dubey, robotics * Paul Yoo, data science * Duncan Golicher, geoinformatics * Damien Fay, computational intelligence * Dean Patton, entrepreneurship * Neil Vaughan, engineering |
| Cambridge, University of Cambridge  Too many to list |
| Charlottesville, University of Virginia  Too many to list |
| Chicago, University of Chicago   * Rayid Ghani, computer science * Matt Gee, computer science * Eduardo Blancas Reyes, engineering * Benedict Kuester, statistics * Jane Zanzig, statistics |
| Colchester, University of Essex  X |
| Delft, Technische Universiteit Delft  Too many to list, different disciplines: computer engineering, software engineering, computer graphics and visualisation, web information systems, network architectures and services, parallel and distributed systems, multimedia computing, pattern recognition and bioinformatics, applied mathematical statistics, interactive intelligence. |
| Dundee, University of Dundee   * Mark Whitehorn, analytics * Andy Cobley, computer science * Yasmeen Ahmad, data scientist * Jonathan Black, software engineer * Lucie Salwiczek, business intelligence |
| Eindhoven, Technische Universiteit Eindhoven  Too many to list |
| European Data Science Academy   * Theo-Jan Renkema, manager of advanced data analytics at Rabobank * Ernistina Menasalvas, computer science * Anders Arpteg, analytics machine learning manager at Spotify * Dave Clarke, chief data scientist Asystec   Many others |
| Halifax, Dalhouse University   * Dr. Michael Bliemel, Faculty of Management (representing the Dean of Management) * Dr. Tom Marrie, Dean of Medicine * Dr. Evangelos Milios, Vice-Dean, Research, Faculty of Computer Science * Dr. Andrew Rau-Chaplin, Professor, Faculty of Computer Science * Dr. Nur Zincir-Heywood, Faculty of Computer Science * Dr. Stan Matwin, Director, (ex officio), Faculty of Computer Science |
| Lancaster University  Many people. |
| Leiden, Universiteit Leiden  Too many to list |
| Manchester, University of Manchester  Too many to list:  <http://www.datascience.manchester.ac.uk/people/?section=Data%20Science%20Institute&sortfield=posttitle&sortorder=ascending&sorttype=text> |
| Warwick, Warwick University   * Director David Firth, statistics. |
| London, Imperial College   * Yi-Ke Guo, computer science * David Hand, mathematics * Jeremy Nicholson, medicine * Jonathan Haskel, economics |
| London, University College   * Patrick Wolfe, statistics + computer science * Sofia Olhede, statistics * Sylvain Robbiano, communication * Swati Chandna, statistics |
| Alan Turing institute  No list |
| Rochester, University of Rochester   * Henry Kautz, computer science * Scott Steele, public health sciences * Michelle Vogl, communication |
| Sydney, University of Technology  Too many to list |
| Edinburgh, University of Edinburgh   * Professor Andrew Morris, medicine * Professor Dave Robertson, informatics * Dr Elizabeth Elliot, English |
| Montreal, Paris, Brussels, Geneve  X |
| New York, New York University   * Foster Provost, information systems * Juliana Freire, computer science * Roy Lowrance, computer science |
| New York, Columbia University   * Kathleen McKeown, computer science * Patricia Culligan, civil engineering * Garud Iyengaer, industrial engineering * David Madigan, statistics * Paul Sajda, biomedical engineering * David Blei, statistics |

University of North Carolina  
No list

New York, Simons foundation

* Tarmo Aijo, systems biology
* Richard Bonneau, systems biology
* Nick Carriero, software developer
* Dmitri Chklovskii, neuroscience
* Ian Fisk, computer science

University of Pennsylvania

* Michael Kearns, computer science
* Rakesh Vohra, economics
* Erol Akcay, biology
* Tom Baker, law & health sciences

Also innovation, genetics, statistics, communication, mathematics

University of Rotterdam

* Jan van Dalen, economics
* Dennis Fok, economics
* Marcel van Oosterhout, management

Loughborough  
No people yet

Northwestern University

* Justin Starren, medicine
* Ramana Davuluri, medicine

Monash University

* Geoff Webb, computer science

Others in computer science, information technology, philosophy,

Caltech

* Julian Bunn, computer science
* Ciro Donalek, computer science
* Matthew Graham, statistics

University College San Diego

* Ilkay Altintas, director, computer science (and a woman!)
* Daniel Crawl, computer science
* Charles Cowart, computer science
* Shweta Purawat, computer science

University of Massachusetts

* Andrew McCallum, machine learning, computer science
* James Allan, computer science
* Yanlei Diao, engineering
* David Jensen, engineering
* Benjaming Martin, computer science

Université Paris-Saclay  
No list

University of Washington

Too many to list, but: computer science & engineering,

RTI International

Collaborates with different universities and disciplines.

# Notes Data Science/Big Data literature

**Boyd & Crawford:**

Big data, benefits and costs, data analytics, big data as socio-technical phenomenon, social media, philosophy of science.

“The era of Big Data is underway. Computer scientists, physicists, economists, mathematicians, political scientists, bio-informaticists, sociologists, and other scholars are clamouring for access to the massive quantities of information produced” (663). “Big data is less about data that is big than it is about a capacity to search, aggregate, and cross-reference large data sets” (663). Define BD as “a cultural, technological, and scholarly phenomenon that rests on the interplay of technology, analysis and mythology” (663).

Ask critical questions of BD, like > which systems are driving BD practices and which are regulating these?

BD not just about big sets of data, also about the computational turn in research. > BD has created a system of knowledge that changes objects of knowledge and can inform how we understand human (and non-human) networks and communities (665).

It’s also very much about automation (of research).

BD looks like it can produces objective knowledge, but does not as data of course still needs to be analysed.

Interpretation will always be at the centre of data analysis.

Nature of large data sets of course influences how it can be analysed and the questions which can be asked so is not necessarily better.

**Imperial:**

“The institute is distinctive because of its multifaceted and collaborative approach, which encompasses work across a broad spectrum of disciplines” (p. 2).

Data Science is “an essential element of all modern interdisciplinary scientific activities” (3).

“It acts as the glue to facilitating collaborative scientific discovery” (3).

“Data Science is not only concerned with the tools and methods to obtain, manage and analyse data, it is also about extracting value from data and translating it from asset to insight” (3).

“The Data Science Institute has been established to conduct research on the foundations of data science and to foster the development of advanced theory, technology and systems that contribute to the state-of-the-art in data science and big data” (3).

**Warwick:**

“The phrase Big Data seems to have mysteriously emerged from nowhere over the last couple of years and now it commonly appears in newspaper headlines, the pronouncements of politicians, and the promises made by marketing groups” (1).

“There is no doubt that we are in an era where technology makes it feasible to gather data at unprecedented levels of volume, velocity and variety, which brings with it both incredible potential and immediate problems” (1).

Mody:

Attention to what scientists do since the late 1970s.

Practice oriented laboratory ethnographies, controversy studies, historical portraits of science.

“Today, STS scholars take for granted that an understanding of scientific *knowledge* requires a thick engagement with scientific *practice*” (1).

**Data science and Prediction:**

“Data” part of data science is new in that it is about data that is heterogeneous and unstructured, “emanating from networks with complex relationships among its entities” (p. 2).

“From an engineering standpoint, it turns out that scale matters in that it has rendered the traditional database models somewhat inadequate for knowledge discovery” (3) > Exactly and this is why there are so many DSI's all over the place. To deal with the fact that new opportunities of data collection require new models and methods.

**Big data, big problems:**

“Big data, as a single technique, is big news” (38).

“Much of social science ethics focuses on rights and responsibilities toward the individual human participant. Big data as a technique does not accommodate this well” (39) > Interesting point, how in this age of Big Data is research ethics seen?

**Bigger sociological imaginations:**

“Over the last several years, the term ‘big data’ has found its way into the discourse of a number of fields, including the social sciences. There seem to be more people with opinions about big data than there are studies utilizing large social data sets” (583).

Big Data is a challenge > but not a new one.

“yielding special issues of journals, new funding lines, new degree programs, and job titles” (584).

There’s people embracing this but also people critical > for instance it is not even clear what big data is > “practices and technologies that are only very loosely related” (584).

Why big data is interesting > “it has become entrenched in the formal institutions that support scholarly work in the social sciences” (585)> tracks in big data or data science, journals etc.

Also “the greatest promise of big data is the opportunity to connect the very large scale of social interactions with the micro-interactions of everyday relationships” (585).

**Editorial**

“It is difficult, nay, impossible to open a popular publication today […] and not run into a reference to data science, analytics, big data, or some combination thereof” (443).

Question is whether big data/data science is really new, what the advantages are, new areas of enquiry etc.

“There are significant new questions and opportunities created by the availability of big data and major advancements in machine intelligence” (444).

“perhaps it is time to set our sights higher, beyond our traditional journals, to communicate with the larger community of scientists and businesses” (445) > which is precisely what is happening with the creation of all these data science institutes.

“It is a time of opportunity for social scientists that have heretofore been hamstrung by the lack of data. For the first time we are able to observe and measure human behaviour on a global scale” (445).

**Data Science, Predictive Analysis**

“Big data carries with it the opportunity to change business model design and day-to-day decision making that accompany emerging data analysis” (77).

“The growth in the quantity and diversity of data has led to data sets larger than is manageable by the conventional, hands-on management tools. To manage these new and potentially invaluable data sets, new methods of data science and new applications in the form of predictive analytics, have been developed” (77). >>> and in academia new infrastructures, centres, institutes are surfacing to deal with this.

Definition of data science > “the application of quantitative and qualitative methods to solve relevant problems and predict outcomes” (78).

Perhaps this is also partly why they find collaboration between different disciplines so worthwhile. Because every discipline has their own domain knowledge but data science and big data practices can have influence on many fields. > because data scientists need deep domain knowledge.

The problem with being a data scientist. Need collaboration because of this. > “typically there is no single individual that can possibly have all of what is needed by a data scientist” (78).

**Cultivating a research agenda for data science**

Focus is on four issues/areas of improvement:

* Integrating algorithms into big data production systems and workflow systems
* Selection of a data movement technology
* Develop effective architectural and software engineering techniques
* Identifying methods and approaches for scientific (and other) collaborations.

# Notes Big Data & Society Publications

**Kitchin (2014)**

Fourth paradigm of science? See table p. 3 > experimental science, theoretical science, computational science and now exploratory science > “Data-intensive; statistical exploration and data mining” (p. 3). Suggestion is that “Big Data ushers in a new era of empiricism, wherein the volume of data, accompanied by techniques that can reveal their inherent truth, enables data to speak for themselves free of theory” (p.3) > But data still has to be analysed! “A new mode of data-driven science is emerging within traditional disciplines in the academy” (p. 3). Does this mean the end of theory and the return of empiricism? “There is a powerful and attractive set of ideas at work in the empiricist epistemology that runs counter to the deductive approach that is hegemonic within modern” (p.4). The suggestion is that with Big Data “a new mode of science is being created, one in which the *modus operandi* is purely inductive in nature” (p.4). But there are four downsides > samples are still samples (no matter how big), big data software still designed for certain purposes, data can never speak for themselves without human framing, and who has the skills *to* analyse (coincides with the argument that data scientists can never know enough when they work alone). Data driven science is born from data instead of theory, so still some advantages. In the end, not necessarily empiricism returns and theory ends, “it is argued by some that data-driven science will become the new paradigm of scientific methods in an age of Big Data because the epistemology favoured is suited to extracting additional, valuable insights that traditional ‘knowledge-driven science’ would fail to generate” (p.6).

Consequences of move towards data-driven science might be changes in understanding of environmental systems (p.6). > Big Data epistemology change approaches towards research in beta sciences, but “their trajectory in the humanities and social sciences is less certain” (p.7). Differences in consequences depend on type of scholars in social sciences (positivist v post-positivist).

Digital humanities, what is going on there? As a field “not been universally welcomed” (p.8), “fostering weak, surface analysis, rather than deep, penetrating insight” (p.8) > is this also a worry for data science institutes????????? According to Kitchin “the same kinds of argument can be levelled at computational social science” (p.8).

Another important issue with Big Data: “much Big Data and analysis seem to be generated with no specific questions in mind, or the focus is driven by the application of a method or the content of the data set rather than a particular question” (p. 9).

“There is little doubt that the development of Big Data and new data analytics offers the possibility of reframing the epistemology of science, social science and humanities, and such a reframing is already actively taking place across disciplines” (Kitchin, 2014, p. 10).

**Leonelli (2014)**

“the novelty of Big Data science lies in (1) the prominence and status acquired by data as scientific commodity and recognised output both within and beyond the sciences and (2) the methods, infrastructures, technologies and skills developed to handle […] data” (p. 2).

Three innovations that Big Data brought > comprehensiveness, messiness (which has to be embraced), triumph of correlations.

One of the problems still with Big Data is a lack of “adequate funding to support and develop online databases” (p. 6) > this is not the case anymore, lots of funding all around the world.

**Diesner (2015)**

“Preparing big social data for analysis and conducting actual analytics involves a plethora of decisions, some of which are already embedded in previously collected data and built tools” (p. 2).

**Mützel (2015)**

“projects using Big Data, from data journalism to computational social science, have little engagement with sociology, although many sociological insights could strengthen analyses; in turn, sociologists could benefit from enhanced computing and visualisation skills” (p. 3).

“Big Data and its methods of analysis challenge the praxis of doing sociology. But, to be sure, sociology has much to contribute to the new arenas of social science research: because of its insights and techniques to study meaning and how the social is structured, sociology makes itself very relevant to data science projects mining large data sets” (p. 3).

# Notes Theory, Culture, Society Publications

**Beer & Burrows (2013)**

“[D]igital data inundation is not just a narrow technical methodological matter for the social sciences; it has been argued that it has far broader implications for disciplinary jurisdiction, the relationship between the academy, commerce and the state, and, indeed, for the very nature of the sociological imagination” (p. 47-48).

“Popular culture is at the centre of the transformations that have facilitated the accumulation of digital data; but it is also at the heart of the issues and debates that face the social sciences” (p. 48).

Scott Lash (2007)

**Ruppert, Law & Savage (2013)**

“Social worlds are thus saturated, being done and materialized by digital devices and what is increasingly being understood as ‘big data’ of various kinds” (p. 23).

“We are concerned with the implications of digital devices and data for reassembling *social science methods* or what we call the *social science apparatus*” (p. 24).

“While digitization is a complex and indeterminate process of intensification whose effects are uncertain, we suggest that it has the potential to reawaken and rework long-established social and political relations” (p. 24).

“Our objective is thus to pose questions about the consequences of digital devices for social scientific ways of knowing” (p. 24).

Make three main arguments:

* Have to trace how material and productive effects of the digital are reconfiguring social science practices
* Have to explore the limits of external critique and the extent to which standard methods and conceptual tools help us to understand information from the outside
* Develop an immanent critique that draws on STS literature on inscriptions

E-science institutes examples, p. 29. Use.

“we do not simply need to rework methods technically, but also to rethink their ontological assumptions including, for instance, their often humanist underpinnings” (p. 30).

The digital “often turn out to instantiate and reconstitute older practices, forms of stabilization and control” (p. 40).