

# Chapter 7

## Data Science Institutes and Data Centers



Rae Earnshaw

**Abstract** Optimum ways of addressing large data volumes across a variety of disciplines have led to the formation of national and institutional Data Science Institutes and Centers. The objectives and functions of such institutes and centers are summarized. In reflecting the driver of national priority, they are able to attract academic support within their institutions to bring together interdisciplinary expertise to address a wide variety of datasets from disciplines such as astronomy, bioinformatics, engineering, science, medicine, social science, and the humanities. All are generating increasing volumes of data, often in real time, and require efficient and effective solutions. The opportunities and challenges of data science are presented. The processes of knowledge discovery in data science often require new methods and software, new organizational arrangements, and new skills in order to be effective. Data science centers and institutes provide a focus for the development and implementation of such new structures and arrangements for the development of appropriate facilities, with academic leadership and professional support. These are summarized and reviewed.

**Keywords** Infrastructure for data science • Interdisciplinary research collaboration • Artificial intelligence • Machine learning • Data mining • Statistical inference • Data management • Data visualization • Cultural change • Reward models

### 7.1 Introduction

Increasing volumes of data in a wide variety of disciplines are forcing academic, industry, and government to address optimum ways to analyze such data effectively. To assist with this, institutions are seeking to provide collaborative research environments in order to increase critical mass and be more competitive for research funding and graduate students. Supercomputers have been used for a number of years to process data from the physical sciences relatively successfully, but the data

science revolution caused by the rapid increase in data from a variety of sources could not be satisfactorily addressed solely within the existing disciplines and structures. Therefore, it has been necessary to initiate a transformation of both the processes of knowledge discovery, and also the institutional environments in which this discovery takes place.

## 7.2 Sources of Data

The Moore-Sloan Data Science Environments comprising New York University, University of California at Berkeley, and the University of Washington (2018) outlined the key elements of a data science environment and also included a summary of current data source as follows:

*Consider the scale and complexity of data sources coming online: simulations of scale and resolution unimaginable only a few years ago (e.g., global climate models, universe-scale n-body simulations), networks of tiny but powerful sensors (e.g., on the seafloor; in the forest canopy; in living organisms; in buildings, roads and bridges), high-bandwidth remote-sensing platforms (e.g., satellites like Terra and Aqua with Moderate Resolution Imaging Spectroradiometers, telescopes used for survey astronomy projects like the Sloan Digital Sky Survey and the Large Synoptic Survey Telescope), high-throughput laboratory instruments (e.g., gene sequencers, micro-and macro-scopic imaging equipment, flow cytometers, mass spectrometers), city-wide urban sensing platforms (e.g., connected vehicles, environmental sensors, ubiquitous cameras), repositories of open government data driven by a new culture of transparency, and social science data created in digital form (e.g., global economic indicators; social network data; consumer activities, including purchasing, mobile phone usage, and internet clickstreams). These advances share a common trait: they produce data with relentlessly increasing volume, velocity, and variety—data that must be captured, transported, stored, organized, curated, accessed, mined, visualized, and interpreted. Data-intensive discovery, or data science, is a cornerstone of 21st-century discovery [1].*

## 7.3 Rationale for Data Science Institutes and Centers

The drivers for the creation of centers to provide a focus for these initiatives include many of the following:

- Alignment with national and international priorities.
- Provision of a top-slice of the institution's budget to support the center.
- Ability to attract significant grant support from funding agencies with priorities in the area of big data.
- Ability to attract advanced level skills and expertise.
- Infrastructure to provide a secure environment to deal with large datasets.
- Opportunities to support research collaboration on big data across an institution.
- Provide research expertise in machine learning, data mining, data visualization, data management, and statistics.

- Provision of interdisciplinary expertise to address a wide variety of application areas.
- Provide an appropriate environment to give support for major national initiatives.
- Gain local and national industry support.
- Gain civic interest and support.
- Opportunities for graduate students to work on a variety of research problems.
- Attractive environment to deliver Masters courses in Data Analytics.

Many institutions with significant investments in research and development have already set up such centers and others have plans to do so. Because the senior management of the university has decided such an institute/center is a high priority for the institution, it often provides funding for the setting up and operation of the facility. It is therefore not a direct charge on existing faculty budgets. Many such centers do not take faculty from existing faculty structures but are an additional resource center which is more centralized within the institution and where faculty and researchers may come together to access and share computational resources, expertise, and ideas.

## 7.4 Objectives and Functions of Data Science Institutes and Centers

Tables 7.1, 7.2, 7.3, and 7.4 show a sample of institutes and centers in a variety of institutions and countries. These are not listed in any order of priority or importance. The tables are only indicative and are not intended to be representative of particular countries, nor are all countries included.

A common driver in all the above organizations is the continually increasing sizes of the data to process in many disciplines.

There is a high degree of commonality in the objectives and functions of the data institutes and centers as illustrated in Tables 7.1, 7.2, 7.3, and 7.4. These include the following:

- To advance R&D in data analytics in all disciplines.
- Enable all fields, professions, and sectors to develop through the application of data science.
- Connect government, industry, and academia.
- Accelerate research, innovation, and training in data-intensive science.
- Support interdisciplinary research.
- Address current problems in society and the environment.
- Develop new mathematical and statistical theory, and quantitative and computational methods.
- Support graduate studies in data science (e.g., Masters degrees in Data Analytics and research).
- Support postdoctoral researchers.

**Table 7.1** A sample of Institutes and Centers in Data Science in the USA

USA	Institute/Center	Objectives and functions
University of California at Berkeley	Institute for Data Science <a href="https://bids.berkeley.edu/">https://bids.berkeley.edu/</a>	Facilitate and nurture data-intensive science Advancing scientific discovery through collaboration across research domains
University of California at San Diego	Halicioğlu Data Science Institute <a href="https://datascience.ucsd.edu/">https://datascience.ucsd.edu/</a> <a href="https://datascience.ucsd.edu/about/index.html">https://datascience.ucsd.edu/about/index.html</a> <a href="https://youtu.be/dHuICSADOZQ">https://youtu.be/dHuICSADOZQ</a> (video 5 min)	By developing new tools and educating the next generation of data scientists, our goal is to improve our quality of life as digital data continues its integration into the infrastructure of commerce, health care, government, and education
Columbia	Data Science Institute <a href="https://datascience.columbia.edu/">https://datascience.columbia.edu/</a>	To advance the state of the art in data science To transform all fields, professions, and sectors through the application of data science To ensure the responsible use of data to benefit society
Georgia Tech	Institute for Data Engineering and Science <a href="http://ideas.gatech.edu/">http://ideas.gatech.edu/</a>	Data science foundations, and data-driven discovery Connect government, industry, and academia to advance foundational research, and accelerate the adoption of big data technology
MIT	Institute for Data, Systems, and Society <a href="https://idss.mit.edu/">https://idss.mit.edu/</a>	Addressing complex societal challenges by advancing education and research at the intersection of statistics, data science, information and decision systems, and social sciences
New York University	Center for Data Science <a href="https://cds.nyu.edu/">https://cds.nyu.edu/</a>	Tools to harness the power of big data

(continued)

**Table 7.1** (continued)

USA	Institute/Center	Objectives and functions
Purdue University	Integrative Data Science Institute <a href="https://www.purdue.edu/data-science/">https://www.purdue.edu/data-science/</a>	The initiative focuses on advancing the frontiers of research and the application of data science to pressing, socially relevant issues, as well as a new campus-wide, transformational data science education initiative
Vanderbilt	Data Science Institute <a href="https://news.vanderbilt.edu/2018/08/13/vanderbilt-data-science-institute-launched/">https://news.vanderbilt.edu/2018/08/13/vanderbilt-data-science-institute-launched/</a>	Promotes and facilitates data-driven research Advance foundational research and data science skills across campus and to leverage the university’s collaborative culture
Washington	eScience Institute <a href="https://escience.washington.edu/">https://escience.washington.edu/</a> <a href="http://data.washington.edu/">http://data.washington.edu/</a>	Advancing data-intensive discovery in all fields Answer fundamental questions through the use of large, complex, and noisy data

**Table 7.2** A sample of Institutes and Centers in Data Science in Canada

Canada	Institute/Center	Objectives and functions
University of British Columbia	Data Science Institute <a href="https://dsi.ubc.ca/about">https://dsi.ubc.ca/about</a>	Catalyze data science innovation Incubate and accelerate research, innovation, and training in data-intensive science
Carleton	Institute for Data Science <a href="https://carleton.ca/cuids/">https://carleton.ca/cuids/</a>	Interdisciplinary research and graduate studies in data science
Montreal	Data Science Institute <a href="https://datascienceinstitute.org/">https://datascienceinstitute.org/</a>	Data science solutions and methodologies to find innovative solutions to the complex problems of our customers
Simon Fraser University	Key Big Data initiative <a href="https://www.sfu.ca/key/">https://www.sfu.ca/key/</a> Research Data Center (RDC) <a href="http://www.sfu.ca/rdc.html">http://www.sfu.ca/rdc.html</a>	Tools, training, and expertise to unlock the potential of big data RDC is part of the British Columbia Inter-university Research Data Centre, supported by partner universities, SSHRC/CIHR, and Statistics Canada to facilitate access to data for crucial social research

**Table 7.3** A sample of Institutes and Centers in Data Science in the UK

UK	Institute/Center	Objectives and functions
Alan Turing Institute	<a href="https://www.turing.ac.uk/">https://www.turing.ac.uk/</a>	National Institute for Data Science and Artificial Intelligence is to make great leaps in research in order to change the world for the better
Ada Lovelace Institute	<a href="https://www.turing.ac.uk/news/turing-partners-nuffield-foundation-announce-new-ps5-million-ada-lovelace-institute">https://www.turing.ac.uk/news/turing-partners-nuffield-foundation-announce-new-ps5-million-ada-lovelace-institute</a>	Ensure data and Artificial Intelligence (AI) work for people and society
Cambridge	Big Data Strategic Research Initiative <a href="https://www.bigdata.cam.ac.uk/">https://www.bigdata.cam.ac.uk/</a>	Applications ranging from astronomy and bioinformatics to medicine, social science, and the humanities Research also addresses important issues around law, ethics, and economics in order to apply big data to solve challenging problems for society
Edinburgh	Edinburgh Data Science <a href="https://www.ed.ac.uk/data-science">https://www.ed.ac.uk/data-science</a>	Bring the university community together with an emphasis on communication, sharing of best practice, and future development
Exeter	Institute for Data Science and Artificial Intelligence <a href="https://www.exeter.ac.uk/idsai/">https://www.exeter.ac.uk/idsai/</a>	Innovate new means of interrogating and understanding data and then to innovate and apply cutting-edge data analytical methodologies to diverse questions
Imperial	Data Science Institute <a href="https://www.imperial.ac.uk/data-science/">https://www.imperial.ac.uk/data-science/</a>	Collecting, preparing, managing, analyzing, interpreting, and visualizing large and complex datasets
Lancaster	Data Science Institute <a href="http://www.lancaster.ac.uk/dsi/">http://www.lancaster.ac.uk/dsi/</a>	Interdisciplinary approach to contemporary data-driven research challenges
Leeds	Leeds Institute for Data Analytics <a href="https://lida.leeds.ac.uk/">https://lida.leeds.ac.uk/</a>	Brings together applied research groups and data scientists from all disciplines, opening up new opportunities to understand health and human behavior and casting light on the action required to tackle a wide range of social and environmental problems

(continued)

Table 7.3 (continued)

UK	Institute/Center	Objectives and functions
Manchester	Data Science Institute <a href="http://www.datascience.manchester.ac.uk/">http://www.datascience.manchester.ac.uk/</a>	Basic research in advanced data analytics and address data-driven research in Health and Biology, Social and Policy, Environment, Urban, Business and Management and the Physical Sciences
Oxford	Big Data Institute <a href="https://www.bdi.ox.ac.uk/">https://www.bdi.ox.ac.uk/</a>	Analysis of large, complex, heterogeneous datasets for research into the causes and consequences, prevention, and treatment of disease
Oxford	Oxford Internet Institute <a href="https://www.oii.ox.ac.uk/study/courses/big-data-analytics/">https://www.oii.ox.ac.uk/study/courses/big-data-analytics/</a>	Collect, clean, analyze, model, and interpret social science data
University College, London	Big Data Institute <a href="https://www.ucl.ac.uk/big-data/bdi">https://www.ucl.ac.uk/big-data/bdi</a>	Investigation of new technologies and analytics as applied to scholarly content and data
University College, London	Centre for Data Science <a href="http://www.ucl.ac.uk/big-data/">http://www.ucl.ac.uk/big-data/</a>	Develop new mathematical and statistical theory, and quantitative and computational methods, to help make sense of large, complex datasets

**Table 7.4** A sample of Institutes and Centers in Data Science in France, Switzerland, Singapore, and Australia

	Institute/Center	Objectives and functions
<i>France</i>		
Paris/Nice	Data Science Tech Institute <a href="https://www.datasciencetech.institute/">https://www.datasciencetech.institute/</a>	Graduate School in France for Applied Data Science, Big Data and Data Engineering
<i>Switzerland</i>		
	Swiss Data Science Center <a href="https://datascience.ch/">https://datascience.ch/</a>	Accompany the academic community and the industrial sector in their data science journey, putting to work AI and ML and facilitating the multidisciplinary exchange of data and knowledge
<i>Singapore</i>		
National University of Singapore	Institute of Data Science <a href="http://ids.nus.edu.sg/">http://ids.nus.edu.sg/</a> <a href="http://datascience.sg/">http://datascience.sg/</a> <a href="http://sdsc.sg/">http://sdsc.sg/</a>	Leverage and strengthen data science expertise for transdisciplinary and translational research into important real-life problems and education of the next generation of data scientists
<i>Australia</i>		
Monash University	Centre for Data Science <a href="https://www.monash.edu/it/our-research/research-centres-and-labs/centre-for-data-science">https://www.monash.edu/it/our-research/research-centres-and-labs/centre-for-data-science</a>	Work with industry, government, and community partners to develop effective data-driven solutions with impact and research expertise. Areas of research expertise include data governance, data management and data archiving, data analytics and visualization in the digital humanities, community, cultural heritage, and not-for-profit sectors
Swinburne	Data Science Research Institute <a href="https://www.swinburne.edu.au/research-institutes/data-science/">https://www.swinburne.edu.au/research-institutes/data-science/</a>	Lead the emerging academic inter-discipline that marries the physical sciences with statistics and computer science. The core business is developing cutting-edge methodologies for handling and analyzing large and complex datasets



## 7.5 Graduate Education and Research

Most of the institutes and centers in the tables also provide a focus for graduate students to study for a Masters degree in Data Analytics and/or to do graduate research for a Ph.D. Many have access to faculty expertise for research and teaching by affiliated appointments with their home department/faculty, or by other linking arrangements such as becoming an Academic Fellow in the institute [2]. Many employ data scientists at Ph.D. level for the research programs within the institute, although attention has to be paid to the issue of the status of such staff within an academic institution to ensure that their career paths and prospects are supported.

## 7.6 National Centers in Data Science

Many countries have identified big data as an area requiring significant development and support. Their governments established working groups to evaluate the current trends in applications data and the needs and requirements of organizations and businesses, in order to evaluate what resources and skills are needed to meet current and future requirements. Some countries have established National Centers to lead and coordinate R&D in the field (for example, [3–5]).

### 7.6.1 USA

In the USA, the National Consortium for Data Science (NCDS) was established in 2013 as a collaboration of leaders in academia, industry, and government to address the data challenges and opportunities of the twenty-first century. The NCDS helps members take advantage of data in ways that result in new jobs and transformative discoveries. Its objective is to enable research, innovation, economic development, and *build a data science community for the economy of the future* [6].

The National Science Foundation (NSF) made a number of strategic recommendations with regard to realizing the potential of data science [7]. This included the following:

- Create Data Science Research Centers.
- Invest in research into Data Science Infrastructure that furthers effective data sharing, data use, and life cycle management.
- Support research into effective reproducibility.

- Fund research into models that underlie evidence-based data policy and decision-making.
- Expand funding into deep learning, smart environments, and other artificial intelligence-empowered areas and their use in data-driven applications.

It also made number of recommendations with regard to education and training, creating collections of datasets, addressing aspects of the Internet of Things (IoT) in areas such as security and data privacy, and addressing architectural issues to support emerging data-intensive tasks.

In February 2018, the NSF announced \$30 million in funding through its critical techniques, technologies, and methodologies for Advancing Foundations and Applications of Big Data Sciences and Engineering (BIGDATA) program. The grants were linked with support from Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure, which have each committed up to \$3 million in cloud resources for relevant BIGDATA projects over a 3-year period. A key objective of this collaboration is to encourage research projects to focus on large-scale experimentation and scalability studies [8].

Data.gov [9] is a U.S. government website launched in late May 2009 by the then Federal Chief Information Officer (CIO) of the United States, Vivek Kundra. Data.gov aims to improve public access to high value, machine-readable datasets generated by the Executive Branch of the Federal Government. The site is a repository for federal, state, local, and tribal government information, made available to the public.

## 7.6.2 UK

In the UK, the Alan Turing Institute was established in 2015 to be the national institute for data science and artificial intelligence. It was created by five founding universities—Cambridge, Edinburgh, Oxford, UCL, and Warwick—and the UK Engineering and Physical Sciences Research Council. Eight further universities—Leeds, Manchester, Newcastle, Queen Mary University of London, Birmingham, Exeter, Bristol, and Southampton—joined the institute in 2018. This linking into a UK Research Council and support by leading universities in the UK effectively established it as an important national initiative and resource.

In the UK, the Diamond Report in 2015 [10] made a number of recommendations for improved education in skills with regard to data analytics, and also proposed the following:

*Many breakthroughs in the development of analytical methods and tools have happened at the intersection between different disciplines. An implication is that we need to support interdisciplinary, innovative research projects involving advanced data analytics, statistics and quantitative skills, and that calls for cross-research council collaboration and funding. Our recommendation is for a top slice of the RCUK budget to establish a strategic fund through which interdisciplinary research is funded. RCUK could itself take a strategic and convening role in this space*

and

*There are currently many agencies in the UK exercising leadership to address the skills shortages arising in industry from the data revolution. However, no single body has all the answers to what are system-wide challenges. Collaboration is needed to address the national challenges identified in our research. We call on relevant stakeholders, including the Tech Partnership, the Royal Statistical Society, the UK Commission for Employment and Skills, the E-Infrastructure Leadership Council, the Digital Economy Council, techUK, the ODI, HEFCE and the research and sector skills councils to set up a cross-cutting taskforce around data analytics to identify good practices for education and skills provision and spur collaboration across industry.*

and

*...grassroots activities could be usefully complemented by a higher visibility network following the example of an organisation like the US National Consortium for Data Science, or Scotland's Data Lab, a £11.3 million initiative supported by the Scottish Funding Council, Highlands and Islands Enterprise and Scottish Enterprise, which 'enables new collaborations between industry, public sector and universities driven by common interests in the exploitation of data science, provides resources and funding to kick start projects, deliver skills and training, and helps to develop the local ecosystem by building a cohesive data science community.' We believe that a data science network along these lines should be developed with involvement from existing communities of analytical practice, the Alan Turing Institute, the Data Lab, the Catapults, and major Data Science institutes at universities like Imperial College, UCL, Manchester and Warwick.*

A report from Nesta in 2015 [11] also proposed the following:

*The 'big data explosion' requires new analytics skills to transform big datasets into good decisions and innovative products.*

### Key findings

- *There isn't a one-size fit all to creating value from data. Our research reveals three types of 'Data active' businesses: Datavores who base their decisions on data and analysis, Data Builders working with big datasets, and Data Mixers who combine data from different sources. We also find 30% of 'Dataphobe' businesses who seem to have given the data revolution a pass.*
- *Data-active companies (especially Datavores and Data Builders) perform better than the Dataphobes. Our econometric analysis reveals that they are over 10% more productive than the Dataphobes after controlling for other factors.*
- *Data-active companies are recruiting more analysts, and combining more disciplines to build a data science capability. This isn't proving easy: For example, two thirds of Datavores struggled to fill at least one vacancy. 80% of them identified problems in at least one skills area. Data-active companies are particularly concerned about the lack of domain knowledge in analysts, the lack of people with the right mix of skills and the lack of experienced analysts.*
- *Technology is changing fast in the data space, and employers are keeping the skills of their data analysts fresh through a variety of approaches. 80% do internal training. Significant proportions (between a third and two thirds) are using innovative training methods like data competitions, online courses and meetups. Only a fifth use universities to train their staff.*

In addition, the Ada Lovelace Institute was set up in the UK in 2018 to coordinate R&D in Artificial Intelligence (AI) and ensure that AI is able to work for people and society [12].

## 7.7 Opportunities for Data Science Research

One of the principal opportunities of a Data Science Institute is to incorporate major research initiatives. An example of this is at the Leeds Institute for Data Analytics. This hosts the following:

- MRC Medical Bioinformatics Centre (£7 million funding) [13] and
- ESRC Consumer Data Research Centre (£11 million funding) [14].

The MRC Medical Bioinformatics Centre aims to create and sustain the infrastructure, facilities, understanding, and culture changes required to enable groundbreaking and productive bioinformatics research at the interface between the clinic, health records, and high volume molecular and phenotypic datasets. It also collaborates with a wide range of business organizations and healthcare providers. In addition, it has been able to leverage an additional £14 million of funding from other sources.

The Consumer Data Research Centre creates, supplies, and maintains data for a wide range of users [15]. It works with private and public data suppliers to ensure efficient, effective, and safe use of data in social science. It is led by the University of Leeds and University College, London, with partners at the Universities of Liverpool and Oxford. One of its functions is to provide a national service using the data that can be accessed via the data store, such as point of sale receipts, travel records, and market research data. It works with private and public data suppliers to ensure efficient, effective, and safe use of data in the social sciences.

Being able to host national data science research initiatives has the following advantages for a Data Science Institute:

- Provides infrastructure for data storage, access, and analysis.
- Provides a center and focus for data science expertise.
- Opportunities for interdisciplinary collaborations.
- Opportunities for cross-fertilization of ideas across disciplines.
- Able to leverage further research grant funding in associated disciplines.
- Opportunities to attract external sponsorship, endowments, and donations.
- Attract interest within the institution in data science.
- Attractor externally for data scientists and graduate students.
- Attract national and international interest in the research.
- Seminar programs with national and international researchers.

A number of these advantages also apply to a Data Science Institute that does not have an associated national facility under its aegis, but they may not be evident to the same extent. In some cases, the setting up of a Data Science Institute has attracted a national facility; in other cases, the institute has focused initially on a pre-existing national facility within the institution, and then broadened the infrastructure to support other disciplines.

## 7.8 Challenges in Data Science Research

Data science research faces a number of challenges. These include the following:

- Discipline-based faculties and budgets within academic institutions do not generally support wider ways of working, because they are based on historic structures which can be difficult to change.
- Interdisciplinary collaborations can be difficult to initiate and operate.
- Grant awarding mechanisms can be biased toward single disciplines because of the way the funding agencies are compartmentalized along disciplinary lines.
- Promotion criteria for faculty can be difficult to formulate and agree (particularly for tenure).
- Career advancement within academia can be difficult in interdisciplinary areas.

Harvey [16] lists the current big data challenges as follows:

- Dealing with data growth.
- Generating insights in a timely manner.
- Recruiting and retaining big data talent.
- Integrating disparate data sources.
- Validating data.
- Securing big data.
- Organizational resistance.

Because of the complexity of the analysis of large amounts of data, questions can arise with regard to the verifiability of the results that are obtained. In order to provide confidence in the publication of the results, it is increasingly required that it clearly specifies how they have been obtained, and give access to the data in order that the results can be verified. This can be facilitated by the use of open-source tools and by supporting reuse and open science [17].

The Moore-Sloan Data Science Environments: New York University, UC Berkeley, and the University of Washington (2018) [1] detail the aspects of institutional change that are required in order to implement data science successfully. These include the following factors:

- Career paths and alternative metrics.
- Education and training.
- Software tools, environments, and support.
- Reproducibility and open science.
- Working spaces and culture.
- Ethnography and evaluation.

Earnshaw [18] details the issues associated with interdisciplinary research and development. The following aspects were considered to be potential advantages:

- Further the growth at the boundaries of existing disciplines.
- Potential for knowledge transfer to industry and society.
- Freedom and opportunity in new areas of research and development.
- Publication of significant results.
- Lines of reporting can be more flexible.

The following were considered to be areas where significant challenges could arise:

- Lack of high-ranking interdisciplinary journals and conferences.
- Lack of general respect in the peer community for interdisciplinary research publications—they can be perceived to be “less pure” than those in a single discipline.
- Blue sky research is still perceived to be of higher kudos than anything that may be more applied.
- Applications to grant awarding bodies and agencies—where reviewers from the different disciplines may have different views.
- Reconciling different cultures and working practices in different disciplines.
- Relationship to senior faculty and the university.
- Tenure committee considerations, where criteria are well established for single disciplines but are often not so clear for interdisciplinary research. This can discourage junior faculty from working in this area as it can be perceived as high risk from a career point of view.
- Employment of faculty on part-time or zero-hours contracts.

As interdisciplinary research gains traction, momentum, and international acceptance through initiatives such as data science, it is possible that some of the challenges listed above may be addressed more effectively than has been the case to date.

## 7.9 Conclusions

The growth in the significance and extent of data from many disciplines has been outlined. This presents a major challenge as it is not possible to analyze this effectively using traditional data processing methods. New approaches are needed, and the opportunities and advantages of data science and Data Science Institutes have been reviewed. Such solutions often require interdisciplinary methods, and therefore effective collaborations are necessary and important in these new environments. These represent a significant shift in the way traditional academic research has been performed within disciplines, and therefore major efforts are needed to address these challenges. However, the rewards are great as the majority of future research and development is likely to be in the context of large datasets which traverse disciplines in the way they arise, and in the methods used to extract verifiable information and knowledge from them.

### Further Reading

Grus, J. *Data Science from Scratch: First Principles with Python*, O'Reilly Media, Sebastopol, CA (2015). <http://shop.oreilly.com/product/0636920033400.do>

Graduate Programs in Computer Science. <https://www.datasciencegraduateprograms.com/>

External Funding Opportunities. <https://ics.psu.edu/resources-for-researchers/external-funding-opportunities/>

NSF Critical Techniques, Technologies and Methodologies for Advancing Foundations and Applications of Big Data Sciences and Engineering (BIGDATA), USA. [https://www.nsf.gov/funding/pgm\\_summ.jsp?pims\\_id=504767](https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504767)

National Institutes for Health (NIH) Computational Genomics and Data Science Program, USA. <https://www.genome.gov/10001735/computational-genomics-and-data-science-program/>

Examples of Big Data Initiatives and Funding Projects. <https://www.icpsr.umich.edu/icpsrweb/content/DSDR/bigdata-examples.html>

Big Data, Arts and Humanities Research Council (AHRC), UK. <https://ahrc.ukri.org/research/fundedthemesandprogrammes/themes/digitaltransformations/bigdata/>

Engineering and Physical Sciences Research Council (EPSRC), UK

New Approaches to Data Science, 2016. <https://epsrc.ukri.org/files/funding/calls/2016/newapproachestodatascienceinfodayslides/>, <https://epsrc.ukri.org/funding/calls/newapproachestodatascience/>, <https://www.ubdc.ac.uk/news-media/2018/january/big-data-project-gets-3m-funding-boost/>

Biotechnology and Biological Sciences Research Council (BBSRC), UK

Data Driven Biology. <https://bbsrc.ukri.org/funding/grants/priorities/data-driven-biology/>

## References

1. Creating Institutional Change in Data Science, The Moore-Sloan Data Science Environments: New York University, UC Berkeley, and the University of Washington (2018). <https://bids.berkeley.edu/news/how-create-institutional-change-data-science-and-encourage-data-driven-discovery>, [http://msdse.org/files/Creating\\_Institutional\\_Change.pdf](http://msdse.org/files/Creating_Institutional_Change.pdf)
2. Imperial: <https://www.imperial.ac.uk/data-science/get-involved/dsi-academic-fellowship/> (2018)
3. UK Government Transformation Strategy, 2017–20. <https://www.gov.uk/government/publications/government-transformation-strategy-2017-to-2020/government-transformation-strategy>
4. UK Government Digital Service: <https://www.gov.uk/government/organisations/government-digital-service>
5. Turing Institute: <https://www.turing.ac.uk/about-us> (2015)
6. Benefits of the public-private partnership for NCDS. <https://datascienceconsortium.org/>, <https://www.hpcuserforum.com/presentations/virginia-april2015/KnowlesNCDS-overview.pdf>
7. Berman, F., Rutenbar, R., Christenson, H., Davidson, S., Estrin, D., Franklin, M., Hailpern, B., Martonosi, M., Raghavan, P., Stodden, V., Szalay, A.: Realizing the Potential of Data Science, Final Report from the NSF Computer and Information Science and Engineering Advisory Committee Data Science Working Group (2017). <https://www.nsf.gov/cise/ac-data-science-report/CISEACDataScienceReport1.19.17.pdf>, [https://www.nsf.gov/events/event\\_group.jsp?group\\_id=20023&org=NSF](https://www.nsf.gov/events/event_group.jsp?group_id=20023&org=NSF)
8. NSF: <https://www.hpcwire.com/2018/02/07/nsf-adds-30m-bigdata-program-aws-google-azure-participate/>, [https://www.nsf.gov/news/news\\_summ.jsp?org=NSF&cntn\\_id=244450&preview=false](https://www.nsf.gov/news/news_summ.jsp?org=NSF&cntn_id=244450&preview=false) (2018)
9. <https://en.wikipedia.org/wiki/Data.gov>
10. Diamond, Sir I.: Making the most of Data: Data skills training in English Universities (2015). <https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2015/making-the-most-of-data-training-skills-english-universities.pdf>, <https://www.universitiesuk.ac.uk/news/Pages/report-looks-at-whether-universities-produce-graduates-with-data-skills.aspx>
11. Nesta: Talent and the Data Revolution (2015). <https://www.nesta.org.uk/report/skills-of-the-datavores-talent-and-the-data-revolution/>, [https://media.nesta.org.uk/documents/skills\\_of\\_the\\_datavores.pdf](https://media.nesta.org.uk/documents/skills_of_the_datavores.pdf)
12. Ada Lovelace Institute: <https://www.turing.ac.uk/news/turing-partners-nuffield-foundation-announce-new-ps5-million-ada-lovelace-institute> (2018)
13. MRC Medical Bioinformatics Centre: <https://lida.leeds.ac.uk/research/mbc/>, <https://gtr.ukri.org/projects?ref=MR%2FL01629X%2F1>
14. ESRC Consumer Data Research Centre: <https://lida.leeds.ac.uk/research/consumer-data-research-centre/>
15. Consumer Data Research Centre Data Service: <https://data.cdrc.ac.uk/>
16. Harvey, C.: Big Data Challenges, Datamation (2017). <https://www.datamation.com/big-data/big-data-challenges.html>
17. Kitzes, J., Turek, D., Deniz, F. (eds.): The Practice of Reproducible Research: Case Studies and Lessons from the Data-Intensive Sciences. Oakland, CA: University of California Press, CA (2018). <https://www.practicereproducibleresearch.org/#>
18. Earnshaw, R.A.: Chapter 3 in State of the Art in Digital Media and Applications, pp. 21–28. Springer, Cham, Switzerland (2017). <https://www.springer.com/gb/book/9783319614083>