Social Big Data

Content and applications

Greater Copenhagen displays a special strength in Social Big Data and data science. This area employs methods from data science, social sciences and humanities to identify patterns in structured and unstructured data, such as texts from social media. Social Big Data provides entirely new opportunities for understanding how behaviours and preferences are shaped and for understanding new market trends among groups of consumers, etc.

Key environments and star researchers

Social Big Data involves several different areas of research. Copenhagen Business School (CBS), the University of Copenhagen and the Technical University of Denmark (DTU) are all heavily involved in the field. At the University of Copenhagen, departments at the Faculty of Social Sciences, some departments at the Faculty of Humanities and the Department of Computer Science host researchers who are active in this field. At CBS, various individuals and areas of research work on Social Big Data, including Ravi Vatrapu at the Department of IT Management, and Thomas Ritter, who heads the CBS Competitiveness Platform. At DTU, various researchers from DTU Compute are heavily engaged in this field. From University of Copenhagen the key star researcher in the area is Prof. David Dreyer Lassen, who received the prestigious Elite Research Award in 2016. He leads a major new research centre focusing on Social Big Data.

Potential for attracting investment

Social Big Data is still in its infancy but it has very considerable potential, both as a new field of research and as an area with great commercial potential. For companies, the perspectives entail better predictions of new trends, deeper understanding of consumer preferences and in gaining new insights into their consumers. Greater Copenhagen is the only region in the Nordic countries to have made a major research effort in this field. Particularly interesting from an international point of view is that the new insights are often discovered by combining unstructured data from social media for example with register-based data on income, education, age, etc.

Denmark is one of the countries in the world with the best register data.

Characteristics of the research area

Social Big Data is a new field of research which combines a broad range of research disciplines in the humanities, data science and social sciences. It involves developing methods to analyse unstructured data such as Tweets, Facebook updates, data from apps and sensors, etc. Furthermore, the research is about developing algorithms for pattern recognition and interpreting and understanding the patterns and relationships that are revealed. The research also holds significant prospects for developing new data-driven corporate business models and patient procedures in hospitals.

This demands a high level of interdisciplinarity with linguists, data scientists, social scientists, anthropologists and economists working closely together to develop new approaches and methods that can translate Social Big Data into new insights for research, the business sector and society in general.

International top quality niches

Researchers in Greater Copenhagen have previously had an interest in Social Big Data, which has already secured a solid basis. Greater Copenhagen's strength lies in a combination of several features, including excellent, close collaboration between researchers in the social sciences and bright researchers in computer science who have extensive experience of Big Data science and in developing specific algorithms, etc.

The University of Copenhagen and DTU Compute have jointly established the Copenhagen Center for Social Big Data Science, which is a unique environment in the world, working on combining unstructured behavioural data with more than 40 years of register data for the entire population.

At the Center for Communication and Computing, researchers from the humanities and natural sciences also work on digital cultures and the interaction between societal phenomena, ICT technology and big data. Among many other things, the centre investigates interactive visualisation of social data and the use of Big Data in healthcare and welfare solutions.

There is also good collaboration between researchers at the Department of Computer Science and Nordic Philology at the Centre for Natural Language Processing, which researches ways to comprehend for example tweets and search patterns on the internet.

CBS has various activities relating to Big Data, such as a three year project on Big Social Data Analytics, which has received DKK 6.2m in funding from the Danish Industry Foundation. Researchers and businesses are collaborating on developing new methods and approaches to the use of Social Big Data Analyses as part of companies' business development processes, enabling them to make better decisions on product development, marketing, sales, strategy and to gain greater insights into user behaviours.

The Danish Industry Foundation has also supported the CBS research project - Big Data, Big Business - anchored at the CBS Competitiveness Platform. The project focuses on transforming Big Data into profitable business models by way of taking novel approaches to existing business models. The Center for Business Data Analytics at the Department of IT Management is participating in these projects and working on Social Big Data as well as an extensive range of other areas, such as health services.

An interdisciplinary centre - DABAI - has been established with funding from Innovation Fund Denmark. It focuses among other things on using Social Big Data to create new innovative solutions to support better forms of teaching, better prediction of flooding, better food traceability and a better overview of patients in the healthcare sector

Bibliometric key figures

The bibliometric key figures do not reveal the strength of Greater Copenhagen's Social Big Data research. This can be ascribed to the fact that this research cuts across various different disciplines. It is especially difficult to identify strengths in the humanities and social sciences in the bibliometry.

The table below displays the bibliometric key indicators for the more “hard”

research areas - General Mathematics and Theoretical Computer Science, which play an important role in Social Big Data in general. Output in Greater Copenhagen is generally somewhat lower than in the regions of comparison. But in terms of quality research as the proportion of articles among the 10% most cited in the field, Greater Copenhagen takes 2nd place among the European regions of comparison.

Key bibliometric indicators

	Specialisation	Output ranking (No. articles)	Highly cited article ranking (%).	Co-publication ranking (%)
General mathematics	0.7	6 (464)	2 (21.1%)	5 (1.7%)
Theoretical Computer Science	0.68	8 (1122)	2 (18.0%)	5 (6.3%)

Period: 2005 -2015. Regions of comparison: Amsterdam, Berlin, Dublin, Geneva-Lausanne, Hamburg, Helsinki, Munich, Oslo and Stockholm/Uppsala. Specialisation is an expression of the size of a field of research compared to all research production at University of Copenhagen, DTU and CBS compared with its size in the regions of comparison. A specialisation level 1 indicates that Greater Copenhagen is on a par with the regions of comparison. Specialisation of >1 indicates that Greater Copenhagen is more specialised in the field of research than the regions of comparison. Output ranking measures Greater Copenhagen's position in the field concerned among the regions of comparison in terms of article production (with the absolute numbers of articles in brackets). Highly cited article ranking indicates Greater Copenhagen's placing in the regions of comparison for the proportion of articles in the field of research in Copenhagen that are among the 10% most cited worldwide (percentage in brackets). Finally, co-publication ranking indicates Greater Copenhagen's rank among the regions of comparison for the proportion of articles in the field published jointly with the business sector (percentage of overall article production in Greater Copenhagen in brackets).

Key arguments for the research area's potential to attract investment

The increasing use and spread of social media, apps and digital devices leaves extensive digital traces and data behind us that can be analysed and provide valuable insights into modern preferences and attitudes and how these change over time. At the same time, the widespread digitalization of the business sector (transition to Industry 4.0) provides entirely new opportunities for developing digital services and new digital business model.

Many companies are already in the process of harvesting social data and identifying patterns as well as interpreting the different types of data. The aim is to be better at decoding new trends in the market and to use Big Data as the basis for developing better products, efficient processes and new types of digital service.

Denmark has a special strength in this area due to its highly technologically literate population and unique opportunities for combining unstructured data and register data. Greater Copenhagen has, as the only region in the Nordic countries, already undertaken a range of major projects to research the use of Social Big Data.

Star researchers and major scientific breakthroughs

Recent years have seen several interesting research projects focusing on Social Big Data.

At the Department of Computer Science at the University of Copenhagen, Prof. Ingemar Cox works on gathering and analysing data from the internet that can be used to predict epidemics and reveal other social phenomena. Prof. Cox, who is the most cited computer scientist in Denmark, collaborates with Google. Prof. Anders Søgaard, also at the Department of Computer Science (and with a past at the School of Authors and in linguistics), does research into artificial intelligence, language technology and machine learn-

ing. His research has been crucial for developing and improving machine translation systems such as Google Translate.

The Carlsberg Foundation provides funding for "The Peoples' Internet" project headed by Prof. Klaus Bruhn Jensen, which combines ethnographic and Big Data-based methods to study how the internet's potential is exploited differently in the EU, China and the US.

The Centre for Information and Bubble Studies under Prof. Vincent Hendricks of the Department of Media, Cognition and Communication at the University of Copenhagen is also supported by the Carlsberg Foundation. The centre uses interdisciplinary perspectives to investigate social bubbles, for example bubbles in financial systems, political opinion forming in the social media or appropriate/inappropriate 'group thinking' in corporate boardrooms. The centre also studies how companies and investors for example navigate their way through Big Data, including the risks to which companies may be exposed when they try to extract value from digital data.

At CBS, Prof. Ravi Vatrapu and Prof. Thomas Ritter are both responsible for the two Big Data projects being funded by the Danish Industry Foundation. Their research involves companies and researchers investigating the challenges and potentials in the use of social media data in business development and actual data-driven business models. Another exciting research project is Social Fabric, which focuses on young people's friendships and their use of mobile phones.

Researchers at the University of Copenhagen and DTU gathered data over a three year period on how 1000 young students use their mobile phones. Their ambition is to use this data to identify friendships and networks amongst young people and to illuminate what friendships mean for students' chances of completing or dropping out of their education.

The research project is headed by Prof. David Dreyer Lassen, one of the leading researchers in Social Big Data, who is also the Director of the Copenhagen Centre for Social Data Science.'

Large talent pool

Exploiting Social Big Data requires close interaction between skilled sociologists, economists and linguists, etc. on the one hand and smart data processors and computer scientists with skills in statistics, algorithms and data science on the other.

Greater Copenhagen already has many bright researchers, PhDs and graduates who have built up considerable knowledge and experience in the field as a result of the many projects implemented in recent years such as Social Fabric, DABAI and the Center for Business Data Analytics and the Big Data - Big Business project at Copenhagen Business School.

One of Greater Copenhagen's major strengths is its strong tradition of data scientists being trained in algorithm design and machine learning. They typically have long experience in interdisciplinary collaborations with researchers from other fields of research.

In addition to the Department of Computer Science, talented PhDs, postdocs and Master's students specializing in the field also come from DTU Compute, CBS and environments at the Faculty of Humanities and Social Science at the University of Copenhagen The overall numbers of talented individuals in Big Data Science will get a major boost in the coming years when the data centre for ESS is anchored at the Niels Bohr Institute.

Unique research facilities

Greater Copenhagen has several strong cards when it comes to attracting foreign investment to the region in the field of Social Big Data.

One of the strongest cards is the unique opportunity for combining data from social media with register data, an area in which Denmark is among the world leaders. This makes it possible to do analyses in which behaviour is correlated with differences in financial circumstances, geography, etc. Danes

are also some of the most digitally literate people with the most widespread use of computers, internet, etc. This means that there are good opportunities in Denmark to undertake natural experiments from which interesting new data can emerge.

One specific example of how unique new data sets can be built up comes for example from data from the Social Fabric project that includes data on 1000 young people's use of mobile phones over a period of three years. This type of data constitutes an important foundation for an entirely new type of research and insights based on analysis of Social Big Data.

Extensive corporate collaborations

Commercial exploitation of Social Big Data is still at an early stage and interaction with the corporate sector in this field is still relatively limited.

However, there are some specific examples of enhanced commercial interaction within the framework of the Danish Center for Data Analytic Driven Innovation, a major project supported by Innovation Fund Denmark and with the participation of the University of Copenhagen and DTU. This allows companies to get a "Big Data" consultation, in which two experts in the field work with the company to exploit Big Data Analytics in the company's innovation processes.

Another example is at CBS, where they have set up a major project entitled Big Data – Big Business. With funding from the Danish Industry Foundation, the project aims to provide advisory services for companies on opportunities for using Big Data in their business development and business model innovation.