

Chapter Title: Introduction

Book Title: Open Data and the Knowledge Society

Book Author(s): Bridgette Wessels, Rachel L. Finn, Kush Wadhwa, Thordis Sveinsdottir, Lorenzo Bigagli, Stefano Nativi and Merel Noorman

Published by: Amsterdam University Press

Stable URL: <https://www.jstor.org/stable/j.ctt1pk3jhq.5>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



This content is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0). To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>.



Amsterdam University Press is collaborating with JSTOR to digitize, preserve and extend access to *Open Data and the Knowledge Society*

JSTOR

1. Introduction

The idea of a knowledge society has been advanced over the last two decades, but the transition to such a society has not yet been realised in reality. Discussions around a knowledge society have largely focused on a knowledge economy and information society, rather than a mobilisation to a knowledge society. These debates have, however, taken place prior to the rise of open data and big data and the ensuing development of an open data movement. This book considers the role of the open data movement in fostering transformation to a knowledge society. The characteristics of the open data movement include the strong conviction of the value of open data for society, attention to the institutional aspects of making data open in an inclusive way and a practical focus on the technological infrastructure that is key to enabling a knowledge society. At the heart of any mobilisation is an emerging open data ecosystem and new ways of producing and using data – whether ‘born digital’ data, digitised data or big data – and how that data, when made openly available, can be used in a well-informed and beneficial way by societal actors.

The book examines how the idea of open data has been taken up by civil society actors and the policymaking community. It considers whether these actors’ activities constitute a social movement that is seeking to mobilise open data and, significantly, whether that work is fostering a transition from an information society based on a knowledge economy into a knowledge society. In order to assess this broad question, it is necessary to explore some key areas of work that are needed to facilitate open data. These include changing institutional frameworks around data, generating data formats that can be made open, generating technical infrastructure and governance models, and addressing research practices and legal and ethical concerns in making data open. To mobilise each of these areas, change is required in the way that each works, along with the creation of new processes and practices. Further, and beyond change in each area, each aspect of change has to interact and link with the other, so that an holistic open data environment is developed. Even though these aspects are important in the mobilisation of open data, social participation in the mobilisation of knowledge society is also needed for such a transformation to occur. Considering participation in the transformation to a knowledge society and participation in a knowledge society raises questions about the position of science in society and the way in which citizens, businesses and civil society actors can participate by using open data. Only then, when the aspects of an open data environment

come together and societal actors can use open data in a socially-defined way, can we say that there is a transformation to a knowledge society.

To assess the role of open data in society and in any transformations to a knowledge society, it is necessary to define what the information society, knowledge economy and knowledge society. In this book, we use *information society* to refer to societies in which information is a central feature in production, innovation and consumption, and which is organised via digital networks. This type of society often has a strong service sector and its economy is driven by knowledge garnered from flows of information. A *knowledge economy* is the economic structure of an information society, because the economy is driven by knowledge that is created from information. Further, this economy is characterised by rapid and continuous innovation, and is global in scope. It draws on an educated workforce within commercial and university research centres that specialise in handling data and information to remain globally competitive in a dynamic and fast-paced global economy. The idea of an information society and its attendant knowledge economy is based on a model where information is not an open commodity and, hence, innovation and growth are managed through private investment and outsourced university spin-outs. This differs from the notion of a *knowledge society*, although there is debate about the precise definition of this. In general terms, a knowledge society distinguishes itself from an information society and knowledge economy because it sees information and knowledge as open to all. Its central value is openness, which means that data, information and knowledge are seen as a 'commons' or shared asset in society. This has the potential to allow any member of society to use data to engage and participate in economic, social, political and cultural projects. Thus, a transformation to a knowledge society is radical in that it seeks to foster open social relations amongst people.

The main argument of this book is that the combination of a proliferation of data and the open data movement are significant features in the possibility of generating and mobilising a knowledge society. A key aspect of mobilising data within a knowledge society framework is the actions of a network of actors who together generate an open data movement, which also interacts with a range of public and private institutions and high levels of digital and digitised data. One factor in the mobilisation of knowledge society is how data can be made openly available and then utilised within wider society. It is envisaged that open data has the potential to foster economic growth and social well-being. However, for this potential to be realised, the data will have to be of high quality and able to be reused and shared across society. The book focuses on how the open data movement is

interacting with three features that are shaping a new data environment: (1) the emergent characteristics of data; (2) a new socio-technical data ecosystem; and (3) a new configuration of institutions that are shaping and mobilising data across a data ecosystem and wider society, along with the development of interpretive communities.

RECODE

The conclusions of and information within this book are based largely on the empirical work conducted within a European Commission funded project called Policy Recommendations on Open Access to Research Data in Europe (RECODE). The primary objective of RECODE was to reduce fragmentation within the open access to research data ecosystem by providing evidence-based and overarching policy recommendations based on good practice. In order to achieve this, RECODE was based around four grand challenges and five disciplinary case studies. The grand challenges included an in-depth, empirical investigation of (1) stakeholder values and inter-relationships; (2) technological barriers; (3) legal and ethical issues; and (4) institutional and policy issues. The five case studies were comprised of:

- **Particle physics and particle astrophysics** – Experiments associated with particle physics often produce extremely large volumes of data. For example, the Large Hadron Collider (LHC) at CERN produces about 15 petabytes of data each year, and requires a custom-made computing grid to collect, analyse and store all of the data produced. As part of the RECODE project, the physics department at the University of Sheffield was used to identify and examine any legal and ethical issues involved in collecting, disseminating, storing and processing large quantities of numerical data from experiments related to particle physics. Central to this examination was the fact that the expertise and resources necessary for storing and processing the data are only available to established experts in the field and/or very large consortia.

- **Health and clinical research** – The case study focused on the ethical issues surrounding data sharing and open data as well as data security of highly sensitive human data. RECODE used the European Commission-funded Markers for emphysema versus airway disease in COPD (EVA)¹ project as

1 See http://cordis.europa.eu/project/rcn/87739_en.html for more information.

a starting point to examine the legal and ethical issues associated with giving open access to health research data. This meant interviewing project experts as well as associated experts in clinical, health and biological data.

- **Bioengineering** – The bioengineering case study focused on the relationship between data sets and complex computational models, with a specific focus on the use or processing of data from human subjects. The starting point for the case study was the University of Auckland's Bioengineering Institute (<http://www.abi.auckland.ac.nz/en.html>), and it included additional experts in the Virtual Physiological Human (VPH) community involved in ontology development, standards for model description and curation of model repositories.

- **Environmental sciences** – The environmental sciences case study focused on the Group on Earth Observation System of Systems (GEOSS) (Group on Earth Observations, 2014) and the RECODE investigation, which is primarily situated within the European Commission's Joint Research Centre (JRC). The GEOSS group uses existing systems and applications for geographic monitoring, including taking observations around drought, forestry, biodiversity and other earth science domains, which include contributions from multiple actors in many countries around the world. In addition to providing interoperable access to data, GEOSS also seeks to develop an advanced operating capacity that provides access to analytical models, which scientists from different disciplines can use to make the data more understandable.

- **Archaeology** – The archaeology case study focused on 'Open Context' (opencontext.org/), a digital repository administered by the Alexandria Archive Institute (<https://alexandriaarchive.org>), a not-for-profit organisation² in the US. Open Context is a free, open access resource that enables the electronic publication of diverse types of research data sets from archaeology and related disciplines. In addition to providing data, it also offers useful information regarding attitudes, practices and policies within the archaeology research ecosystem, along with technical information about depositing, accessing and preserving archaeological data.

Despite the specificity of these case study descriptions, as the RECODE project developed, it became clear that the issues identified in each case

2 Open context is financially supported by The William and Flora Hewlett Foundation, The National Endowment for the Humanities and The Institute of Museum and Library Services.

study should be more broadly considered, extending the research to stakeholders within and outside of each specific study. Therefore, for example, some of the archaeology case study quotes presented in the chapters that follow are not only from participants directly involved in the specific case study organisation, but also from people involved in a range of other organisations within the open access to an archaeological data ecosystem. Thus, they should be read as academic disciplinary case studies, rather than organisational case studies. Nevertheless, the grand challenges and case studies created a matrix where each grand challenge was examined in each discipline. These case studies provided an interdisciplinary grounding, helping to maintain an awareness of discipline-specific issues and practices as well as providing insight into the grand challenges.

The specific RECODE methodology consisted of a three-step process: an extensive review of the literature, interviews with stakeholders within the disciplinary case studies and stakeholder validation workshops. The literature reviewed consisted of policy literature (e.g. national, European and international policy), practice literature (e.g. publication, data management and ethical protocols), grey literature (e.g. manifestos, white papers and blog posts from open access organisations/visionaries), and academic literature. In total, the project conducted 65 semi-structured interviews with academics, researchers, policymakers, data centre staff, legal experts, scientific publishers and other experts working within the field of open access to research data. Findings from the literature review and case studies were then further validated in five stakeholder workshops, which, in total, counted 168 workshop participants from 35 countries. The workshop attendees came from different stakeholder groups, e.g. policymakers, data managers, researchers, academics, librarians and publishers. The validation events offered the opportunity to discuss and debate the findings of RECODE and test their relevance and applicability in a broader context.

RECODE was an original opportunity to examine the interrelationship between a particular sphere of the larger open access movement and the impact of that movement on both stakeholders within it and the larger scientific and research culture. It is the first study to take the open access movement itself as an object of investigation. Thus, it provides empirical evidence about the ways in which data is being shaped by and shaping research culture, the institutional, technical and social ecosystems that are emerging within this open data movement and the ways in which institutions have adapted to this emerging landscape.

The book

This book is a jointly authored book that is based on the research in the RECODE project. Bridgette Wessels is the lead author and she wrote Chapters One, Two, Three, Four and Ten as well as overseeing and commenting on the work of the other authors. Rachel Finn wrote Chapter Eight, co-wrote Chapter Nine and read and commented on the whole book. Merel Noorman wrote Chapter Five, Thordis Sveinsdottir wrote Chapter Six, Lorenzo Bigagli and Stefano Naviti wrote Chapter Seven, and Kush Wadhwa co-wrote Chapter Nine.

In order to explore the issues of open data discussed above, the book is structured in the following way. Chapter Two, 'Defining a Knowledge Society', reviews the debates about the meaning of a knowledge society in relation to discussions around an information society. It argues that the way in which the open data movement is driving for data to be openly available is a key feature of the emergence of a knowledge society. There are, however, barriers and risks in making data open, which are discussed in the chapters that follow. The term 'knowledge society' was first coined by Peter Drucker in 1969, but it was not developed further until the mid- to late 1990s, when scholars such as Robin Mansell (1998) and Nico Stehr (1994) explored the idea further. Work by Mansell and Stehr points out that debates about an information society cannot be separated from considerations about a knowledge society. This is because the notions of an information society rest, to some degree, on commercial and economic networks of society that are technologically supported, whereas the concept of a knowledge society encompasses other dimensions, such as ethical and political concerns within social life. Stehr (1994, 2004, 2012) argues that the development of a knowledge society is a gradual process that is not deliberately triggered by human design but, instead, is shaped by the way that new technologies, new data fields, new needs, and new imaginations interact and configure to produce new possibilities and innovations. Stehr (1994) defines knowledge as a capacity for action, which, he argues, has multifaceted implications. In general terms, knowledge is different from information, in that it requires cultural interpretation to create it. In summary, UNESCO makes some ethical and normative suggestions within a development paradigm by asserting that, to 'remain human and liveable, knowledge societies will have to be societies of shared knowledge' (UNESCO 2005). It is thus arguing for an open approach to knowledge, and data is very much part of this dynamic. We conclude this chapter by noting that, although there are various definitions of knowledge society

or societies, there is a strong call from regional to global actors for policy for 'activating knowledge' (Soete 1997).

The main argument of Chapter Three, 'Visions of Open Data', is that the value of data is realised through its own characteristics as well as by the way it can be used in specific contexts. It is difficult to define data in a way that is sufficiently precise, yet broad enough to capture the richness and diversity of data. 'Data' as a concept can be viewed as the lowest level of abstraction, from which information and then knowledge are derived. In general terms, data is a set of values of qualitative or quantitative variables. The presence of data, as something already found, or something created through research processes, or as a by-product of social media is seen to have value, either in itself or in its reuse. The chapter considers the way that definitions of open data have been developed by actors in civil society. In particular, it notes how some civil society actors' visions of open data link with ideas about open knowledge and an open knowledge society. In the context of civil society organisations that act as advocates for open data, the focus is on the characteristics of open knowledge and what its context of use could be. Another vision of open data can be found at the government level. In this context, there are a set of principles and guidelines about open data as well as visions for its use. Another area that is developing visions of the use of open data is academia, where the focus is on open access to research data and how that might benefit the academic and scientific communities. In the area of big data, the proliferation of open data carries significant potential for the construction of big data sets; however, the integration of these data resources has yet to be adequately realised to enable this transition.

Chapter Four, 'Mobilising Open Data', addresses how open data is being adopted in society. As the previous chapters show, the position of scientific knowledge in wider social and economic life has changed in late modernity. They also show that data and open data are being discussed in civil society and by governments in ways that focus on the possible social benefits of open data. These two areas – the changing role of knowledge in society and the possible benefits of open data – should be viewed in relation to each other, because the aspirations for open data are often couched within understandings of the role that knowledge plays in society. Furthermore, the possibility of open data combined with changing senses of the position of knowledge in society are utilised in various visions of a more knowledgeable society. Within these discussions, there is a recognition of the potential benefits to society, as well as the possible threats and risks of open data. An overarching theme in these discourses is that of 'open' – both in terms of open data and open society. The chapter brings together these discussions

by showing how the main theme of openness in open data is a key driver of change, concluding that there is a strong conviction by those in the open data movement that open data has the potential to be valuable for society, in both general and specific terms. The consensus about the value of open data is applied to a wide range of social and economic areas, such as open government, development and human rights, innovation and commerce. However, the realisation of that ambition is complex and requires technological, institutional, legal and cultural change in a social transformation to a knowledge society.

Chapters Five, Six, Seven and Eight are all based upon findings from RECODE, where empirical information from the project is used as a platform from which to consider the interrelationships between providing open access to research data and progressing towards a knowledge society. Chapter Five, 'Institutions in the Data Ecosystem', discusses the importance of institutions in socio-technical change and in their role as curators of data in the mobilisation of a knowledge society. The chapter discusses the role that these play in the changing data environment, and refers to diverse institutions including research councils and institutes, foundations, policymakers, advocacy groups, Civil Society Organisations (CSOs), universities, scholarly societies, intergovernmental organisations (IGOs), standards organisations, service providers, data centres, information aggregators, libraries and archives, publishers, professional associations, scholarly societies, public sector data collectors, and private sector big data collectors. These institutions form part of an emerging open data ecosystem and are grouped in this chapter by their key functions, which include funding, creating data, creating data repositories, curating data, accessing data, and using and disseminating data. The chapter also addresses some of the challenges that such institutions face. In particular, it explores the challenges involved in navigating between the competing interests of heterogeneous stakeholders, entrenched institutional cultures and wide-ranging and, sometimes, conflicting ideas about open data. It considers the opportunities for institutions to contribute to an open data ecosystem that will benefit the knowledge society.

Chapter Six, 'Scientific Disciplines, Scientific Practice and Making Research Data Open', argues that the move to openness requires a change in research practices and the ways in which data is gathered, stored and analysed. It also argues that scientific disciplines face different barriers in their move towards open access to research data, since the research process requires procedures to ensure the generation of data that can be openly available, accessible, and reusable. The chapter

draws on qualitative research in five scientific disciplines: archaeology, bioengineering, environmental sciences, health and clinical research and particle physics. It demonstrates how their contribution to knowledge generation is based on different scientific practices, and details how they are having to adapt in order to participate in the move towards open access research data. The chapter also highlights the increasing and emergent complexity of contemporary scientific practice with respect to interdisciplinary research, sophisticated technology (e.g. simulations), international collaborations and vast amounts of data, discussing how these may complicate the process of research and, consequently, the production, analysis and storage of data. The chapter considers how the move to openness in science is supported by the notion of the knowledge society, where knowledge is no longer limited to a small, exclusive group, but is seen as a public good. Science plays an important role in society, because the production and advancement of knowledge are among its key aims. Open access to research data is one aspect of transferring knowledge and gaining the ability to bring new knowledge closer to the public. In order for this to happen, disciplines will need to consider the role of data in their research practices and decide how it can be made reusable and accessible. This chapter considers the barriers and opportunities that exist within different scientific disciplines.

Chapter Seven, 'Environmental Data, Technical and Governance Issues', introduces some of the specific features of mobilising open data, by focusing on one context in which this is already happening – the geospatial data sector, including environmental data and earth sciences. This case study shows how the use of open data to support the challenges facing the world in terms of global environmental challenges is supported by scientists, governments, policymakers and activists. This generalised sense of consensus has helped stakeholders to mobilise some levels of open data within its broad community. In addition, efforts to strengthen the political cohesion of geographical regions (e.g. the EU), to digitise public administration, to better understand and mitigate global-scale phenomena (e.g. climate change), or the growing interest in space programmes, are all greatly contributing to the momentum of the open data movement in the geospatial sector. The chapter elaborates on the geospatial data ecosystem and the way that its stakeholders are addressing technological issues such as interoperability at the infrastructural as well as the semantic level. The chapter also considers the issue of governance, which is recognised as one of the most important aspects of developing open access to geospatial data. This requires mutually-agreed policies on the exchange, sharing, access and use of interoperable

data and services across various levels of public authority and different sectors of society, at a global level.

Chapter Eight, 'Navigating Legal and Ethical Frameworks', address the issues of ethical, legal and regulatory frameworks in the sciences and humanities for mobilising open data. In many contexts, ethical, legal and social issues have been construed as a barrier or challenge to providing open access to data, especially data that raises intellectual property considerations or data relating to people that could infringe people's personal privacy (particularly identifiable individuals). However, alongside the need to meet legal obligations and ethical standards around research and data collection, some stakeholders are being strongly encouraged to enable a realisation of the knowledge society, either through making as much of their data as possible open, or by exploiting their data to enable innovation. This is being advocated, in particular, by policymakers, funders and some civil society organisations within the open data movement. Researchers, data centres and institutions emerge as key stakeholders in relation to these (sometimes) competing demands, and these groups are often leveraging existing infrastructures or devising new solutions to tackle these issues simultaneously. This chapter examines the interplay between ethical, legal and regulatory frameworks in the provision of open access to research data in order to enable the knowledge society. It addresses the intersecting and, sometimes, competing governance structures being navigated by researchers, institutions and data centres. At the governmental level, these may be legal or legislative obligations, such as privacy, data protection and intellectual property, which may be mandated by national or supranational (e.g. European) levels of government. The chapter asserts that understanding legal and ethical obligations as challenges or barriers can bring about that very effect. Strong legal protections and ethical practice will foster trust in data practices, institutions and governance structures, which will encourage stakeholders to provide data and to open and share that data. Constructing solutions to the legal and ethical issues will therefore support efforts to integrate open data into the knowledge society.

Chapter Nine, 'Big Data, Open Data and the Commercial Sector', examines opportunities for innovation through the intersections between big data and open data. It focuses on policymakers' expectations that big data and the increasing availability of external data resources would result in significant opportunities for innovation in Europe, the US and globally. While large technology companies in the US have dominated the big data ecosystem thus far, policymakers expect that small- and medium enterprises (SMEs) should also be key beneficiaries of this potential for

innovation. Using information from the Open Data 500 lists, this section examines the extent to which SMEs have been able to capitalise on this opportunity for innovation. It argues that the fruits of the big data evolution are situated within a complex innovation space that initially produces benefits for large enterprises, which have significant capacity for investment in infrastructure and capabilities. These benefits slowly expand to permeate into small actors directly within the large company's ecosystem, then onto actors outside the bounded ecosystem, in wider society. The chapter also examines the extent to which policymakers' and civil society organisations' advocacy for open data might introduce additional complexity into this innovation space, by prioritising the openness of data in ways that might make it difficult for SMEs to protect the sustainability of their innovations. Finally, the discussion concludes with a consideration of the uneven distribution of big data innovation globally. Rather than expressing concern about a lack of visible innovation in Europe in comparison to the US, the chapter argues that Europe's strong protections for privacy and intellectual property rights themselves provide an opportunity for innovation. Thus, instead of trying to emulate the US, Europe should be investing in creating an innovation space that reflects and reinforces European values, particularly those about responsible innovation, as this will ultimately support some elements of a knowledge society.

Chapter Ten, the conclusion, argues that the vision sought by the open data movement is a key feature of the way that social participation and innovation is being considered by policymakers and industry as well as by the civic society stakeholders and institutions involved in the production of knowledge. The open data movement is interacting with other key features discussed in the book, which are: (1) the emergent characteristics of data; (2) a new socio-technical data ecosystem; and (3) a new configuration of institutions that are shaping and mobilising data across a data ecosystem and wider society, along with the development of interpretive communities.

The development of a new data ecosystem is currently in its infancy, so several issues need to be addressed, such as governance, interoperability, data curation, licensing and ethical issues. Furthermore, institutions and organisations in both the public and private sectors are reconsidering how they value data and how they might share and make that data open in ways that could benefit society. The open data advocates have come together to form a movement as a network of networks and, in this form, constitute a key actor in fostering open data. There will, however, need to be a shift in social imagination in terms of how to use data, as Stehr argues, as well as further development of the data environment and its interpretive

communities in order to mobilise a knowledge society. Nonetheless, the open data movement has moved society some way towards being able to provide an important element in the mobilisation and transformation from an information society to a knowledge society.