INFORMATION AND SOCIETY

MICHAEL BUCKLAND



INFORMATION AND SOCIETY



The MIT Press Essential Knowledge Series

Robots, John Jordan Waves, Frederic Raichlen

Auctions, Timothy P. Hubbard and Harry J. Paarsch

Understanding Beliefs, Nils J. Nilsson Cloud Computing, Navan Ruparelia Computing: A Concise History, Paul E. Ceruzzi The Conscious Mind, Zoltan L. Torey Crowdsourcing, Daren C. Brabham Free Will, Mark Balaguer Information and Society, Michael Buckland Information and the Modern Corporation, James W. Cortada Intellectual Property Strategy, John Palfrey The Internet of Things, Samuel Greengard Memes in Digital Culture, Limor Shifman Metadata, Jeffrey Pomerantz The Mind-Body Problem, Jonathan Westphal MOOCs, Jonathan Haber Neuroplasticity, Moheb Costandi Open Access, Peter Suber Paradox, Margaret Cuonzo Self-Tracking, Gina Neff and Dawn Nafus

INFORMATION AND SOCIETY

MICHAEL BUCKLAND

© 2017 Massachusetts Institute of Technology All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher. This book was set in Chaparral Pro by Toppan Best-set Premedia Limited. Printed and bound in the United States of America. Library of Congress Cataloging-in-Publication Data is available.

ISBN: 978-0-262-53338-6

10 9 8 7 6 5 4 3 2 1

CONTENTS

Series Foreword vii Foreword by David Bawden ix Preface xiii

1 Introduction

Information – My passport – The division of labor and the need to know – Agendas of others – Information society – Truth, trust, and belief – The structure of this book

2 Document and Evidence 21

Information as thing – Documents and document anatomy – The history of information technology – The rise of data sets – Some practical initiatives – Problems of later use – Bibliography reconsidered – World brain and other imagery – Summary

3 Individual and Community 51

What individuals do – What communities know – Culture – Documents as the activity of others – The social and the individual – Physical, mental, and social dimensions of information – Summary

4 Organizing: Arrangement and Description 71

Collections – Arrangement and lists – Description – Not so easy! – Generating descriptions – The basic mechanism – Summary

5 Naming 89

 $Topic\ descriptions-Documentary\ languages\ for\ naming\ topics-Time\ and\ naming-Mention\ and\ meaning-Naming\ is\ cultural-Summary$

6 Metadata 111

The first purpose of metadata: Description – Creating indexes – Index terms – The second use of metadata: Search – What, who, where, and when – Relationships among index terms – Facets and context – Summary

7 Discovery and Selection 135

Retrieval and selection – The anatomy of selection machinery – Searching text – A library catalog – Searching the web – Other examples – Summary

8 Evaluation of Selection Methods 153

Relevance, recall, and precision – Recall with random, perfect, and realistic retrieval – Precision with random, perfect, and realistic retrieval – Trade-off between recall and precision – Some problems with relevance – Why relevance is difficult – Summary

9 Summary and Reflections 167

Summary – The past and the future – Coping: Orality, literacy, and documentality – What kind of a field?

Appendix A: Anatomy of Selection 183
Appendix B: Retrieval Evaluation Measures 189
Further Reading 199
Glossary 203
References 207
Index 215

SERIES FOREWORD

The MIT Press Essential Knowledge series offers accessible, concise, beautifully produced pocket-size books on topics of current interest. Written by leading thinkers, the books in this series deliver expert overviews of subjects that range from the cultural and the historical to the scientific and the technical.

In today's era of instant information gratification, we have ready access to opinions, rationalizations, and superficial descriptions. Much harder to come by is the foundational knowledge that informs a principled understanding of the world. Essential Knowledge books fill that need. Synthesizing specialized subject matter for nonspecialists and engaging critical topics through fundamentals, each of these compact volumes offers readers a point of access to complex ideas.

Bruce Tidor
Professor of Biological Engineering and Computer Science
Massachusetts Institute of Technology

It is a truism to say that we live in an "information age" or an "information society," but it is nonetheless impossible to deny that information (along with data and knowledge, if we wish to make to make the distinction) is now central to the functioning of all developed societies.

It is conventional to suggest that we came to this situation through a series of "information revolutions," by which a new technology, using the word in its broadest sense, drastically changed the way information is recorded and communicated. The number and nature of these revolutions varies between commentators, but typically they include the introduction of writing, printing, mass communications, the digital computer, and the Internet.

A cogent analysis by Luciano Floridi argues that we are living in an age of "hyperhistory," in which the well-being of individuals and societies is entirely dependent on information and communication technologies. Floridi's contention is that we are seeing an "informational turn" or "fourth revolution," following the scientific revolutions of Copernicus, Darwin and Freud (Floridi 2014). We should regard ourselves as informationally embodied organisms, "inforgs," embedded in an informational environment, the "infosphere," in which the boundaries between our online and offline environments merge.

Given this embedded centrality of information in modern society, it is not surprising that it is studied, from various points of view, by a number of disciplines, including computer science, media studies, psychology, sociology, mathematics, education, economics, and philosophy. These are only the disciplines interested in information in the sense of meaningful, communicable information. The list lengthens if we include conceptions of information in other domains, such as physics and biology (Robinson and Bawden 2013).

The one discipline that has information as its sole object of interest is information science. This grew during the twentieth century from the concerns of the "documentation movement," which sought to understand the nature of documents of all kinds, and hence to provide access to them in a much more sophisticated way than conventional catalogs and indexes could provide (Wright 2014). The advent of the digital computer gave an impetus to the new discipline, which has overlapped with, while remaining distinct from, computer science. Information science concerns itself with all aspects of the organization and communication of recorded information, with the information and digital literacies needed to make use of it, and with associated ethical issues. The insights of the discipline are crucially relevant in developing the dramatically changing infosphere.

There are a number of good texts setting out the basics of information science; I am co-author of one such (Bawden and Robinson 2012). But these are typically aimed internally: at faculty, students, and practitioners within the subject. If we believe, as I do, that information science has many insights to offer to a much wider context, then we need books that specifically address a wider audience. Michael Buckland's book is the first to attempt this.

An impressive feature of the book is the way in which such a breadth of material is brought together clearly and concisely. It is pleasing to see how Buckland integrates the "traditional concerns" of information science—in particular, how information resources are described, organized and retrieved, and the ways in which people and groups behave with information—with thoughts about the nature of the documents through which recorded information is communicated. The new forms of document which have emerged in networked digital environments have led to a renewed interest in the nature of documents, and this kind of conceptual analysis, though lower key than new technological developments, is just as likely to be valuable in ensuring effective communication, and good use, of information. I applaud Buckland's vision of information science as a broad and inclusive field of study; only such a holistic approach can do justice to the issues that emerge when information takes center stage in society.

Michael Buckland modestly writes that his interpretation draws on the work of many people and little of it is original. While this may be true in the sense that the material presented in his book has mostly been published before in some form, I think there is great originality in the way it has been selected, organized and presented for a nonspecialist audience. This is highly commendable, and the book deserves to be widely read. Its success will, perhaps, be measured by whether it is the first of a succession of publications, bringing the insights of information science and the documentation movements to a wider audience. We must hope that proves to be the case, as these perspectives are sorely needed if the infosphere is to develop and flourish.

David Bawden
Centre for Information Science, City University London

References

Bawden, David, and Lyn Robinson. 2012. Introduction to information science.

Floridi, Luciano. 2014. The fourth revolution: How the infosphere is reshaping human reality. Oxford: Oxford University Press.

Robinson, Lyn, and David Bawden. 2013. Mind the gap: Transitions between concepts of information in varied domains. In *Theories of information, communication and knowledge: A multidisciplinary approach*, ed. Fidelia Ibekwe-SanJuan and Thomas Dousa, 121–141. Berlin: Springer.

Wright, Alex. 2014. Cataloging the world: Paul Otlet and the birth of the information age. New York: Oxford University Press.

We now live, we are told, in an *information society* and this is considered to be a very important development. But what does it mean? And what are the consequences? In answer, we offer a short informal introduction to ways in which *information* and *society* are related and to our everincreasing dependence on a vast increase in documents and data of many kinds.

The word information is used here in an ordinary, everyday way. Other, specialized uses in law, statistics, thermodynamics, cryptography, and elsewhere we leave to others. We are concerned with information as influencing what we know, with the role of communication and, especially, recorded information, in our daily lives, and with how information is found. We are, therefore, concerned with beliefs, social agendas, and changing technologies. These are complex areas that resist simple, formal explanation. There are conceptual and theoretical difficulties that we will mention but do not claim to solve. Our purpose is to provide a descriptive introduction which draws on the work of many people and little of it is original. Since it is an elementary introduction, the detailed citing of sources ordinarily found in academic writing is not provided. However, much of the text is adapted from earlier publications which are identified in the Further Reading section at the end where some additional sources are also suggested.

I thank Wayne de Fremery for his insight on Information Theory, Wayne Heiser for help with Zublin's letter, and Lisa Börjesson, Colin B. Burke, Vivien Petras, and Lin "David" Wang for their helpful comments on early versions of the text.

INTRODUCTION

The word *information* has been used with several different meanings. In this book we are concerned with information in society, in everyday human experience. So we may call this *realistic* information science and distinguish it from studies associated with statistical and other technical analyses in cryptanalysis, signaling systems, thermodynamics, and other areas (*formal* information science). Specialized, technical uses of the word *information* unrelated to human knowing and everyday experience are outside our scope. Our scope can be illustrated by considering a passport, a rather complex document that plays a powerful role in social control.

We are concerned with human perception, social behavior, changing technologies, and issues of trust. So what we examine will often be complex, untidy, or unclear, and this prevents simple, satisfactory scientific analyses. The growing importance of information derives from the progressive division of labor, which characterizes our transition from hunters and gatherers to an increasingly complex society. We depend more and more on others, which requires coordination and communication and means, in practice, dependence on information. This is not a simple or neutral situation, because many others seek to advance their agendas by using different forms of information to shape our beliefs and behaviors in ways they want.

Information

Use of the word *information* increased greatly during the twentieth century and has developed many different meanings. The lack of settled terminology and figurative use, as in *information society*, make discussion difficult. The lack of agreement on its meaning makes it suitable for slogans and for enthusiastic metaphors. Anyone wishing to be precise and clear should either declare the particular meaning being used or, better, use some other, more precise word or phrase (such as *data*, *record*, *document*, or *knowledge imparted*), since for each of the principal meanings of the word *information* there is also another suitable, more specific word.

We depend more and more on others, which requires coordination and communication and means, in practice, dependence on information. Words get used in creative ways, and there are differences in the way they are used in different contexts. Multiple words may be used to refer to an individual thing, while individual words are commonly ambiguous, used to refer to multiple different things. English words ending in "-ation" are usually ambiguous, referring variously to a process, an event, an object, or an outcome. (Consider *communication* and *regulation*.) The character string "i n f o r m a t i o n" has been used to refer to many different things. So any simple assertion in the form "Information is ..." has little meaning and encourages confusion unless it is made clear which of the meanings is intended.

In the middle of the twentieth century, the word information was adopted as a technical term (notably as information theory) in engineers' calculations of reliability in telephony and similar signaling systems. This use related the word information to a series of developments in logic, probability, and computation that proved very fertile in some important fields, notably cryptanalysis, electrical engineering, and thermodynamics. In these developments, it is ordinarily assumed that information is true, knowledge is true, and differences can be rendered as bits. But these important technical developments and assumptions about the truth value of information have little overlap with everyday human experience, so it is important to recognize that there are two fields of

study, both of which have used the name *information* science, but they have little in common beyond using the same name. Each seems to have limited interest in, or relevance to, the other.

We can make an analogy with philology, the field in the humanities concerned with the examination of texts. Here a distinction is made between examination of the text itself, known as the lower criticism, and the examination of a text in its material and sociohistorical context, the higher criticism. We might, comparably, refer to the lower information science and the higher information science, or, more diplomatically, make a distinction between formal and realistic approaches to the study of information.

In this book, we are concerned with information in relation to everyday human experience, and thus with the complex multiplicity of messages, records, documents, and perceptions in our lives; the difficulties associated with meaning and finding the most relevant information; and the need to trust sources and documents. This, then, is an introduction to "higher," or realistic, information science in its cultural context rather than the *formal* information science described above and discussed elsewhere.

We need to include more than the study of what are traditionally called documents, because we communicate through gesture, language, and the use of material objects. Our central theme is that in modern society, interpersonal relationships are increasingly *indirect*, through messages,

records, and other forms of document. In what follows, *information* is used in an ordinary, everyday sense with two related meanings: (1) what we infer from gestures, language, texts, and other objects; and (2) material forms of communication—bits, books, and other kinds of physical messages and records. We start with an example.

My Passport

My passport is more powerful than I am, because I cannot cross frontiers without it, but it could cross them without me. This small, printed booklet seems, at first sight, to be a good example of the static nature of traditional media. But the inside is more complex. It has multiple and often changing components. There is a photograph of me and my handwritten signature. There are marks for optical character recognition inside the front cover and a bar code inside the back cover, each of which make it a digital document capable of being read into computers. The pages became filled up with marks stamped by frontier officials that record my travels, and extra pages were added to accommodate more. Additional documents have been inserted: elaborate visas issued by the Chinese, Russian, and Vietnamese governments that provided income for them and permission for me. A biometric security code was inserted at Heathrow Airport, and some small security stickers have been attached to the back cover. A more recently issued passport would also include a chip capable of transmitting my name, nationality, gender, birthday, birthplace, and portrait.

My passport is going to expire, as I will also. The passport's expiration date, unlike mine, is exactly known and, unlike mine, easily extended with a small fee for renewal. So although parts of my passport are carefully designed to prevent alteration, it is physically dynamic, as are many other types of information, especially those in electronic form. Even the most conventional document, writing on paper, is technology. Technology is now increasingly electronic—so-called information technology—as if pen and paper were not information technologies.

The social aspect of my passport is clear when we remember that it is not really the passport itself that allows me to cross frontiers or board airplanes, but guards enforcing regulations. In remote areas where there is no physical barrier, I could cross a frontier with or without a passport; illegally, perhaps, but it could be done. If the frontier is not well marked, I might even cross it unintentionally. So the power of my passport does not arise simply from the document itself, but from more or less enforced social regulations within which passports are used as an evidentiary device within a system of controls embedded in complex bureaucratic systems.

Strictly speaking, a government can only control (or try to control) its own borders, not those of other countries, but acceptance of the validity of my passport extends internationally through requests and agreements. I used to carry a British passport that had an impressive printed page elegantly inscribed with a statement reminiscent of a nineteenth-century imperialist power: "Her Britannic Majesty's Secretary of State requests and requires in the name of Her Majesty all those whom it may concern to allow the bearer to pass freely without let or hindrance and to afford the bearer such assistance and protection as may be necessary."

There is a cognitive aspect as well. A guard needs to examine the passport in order to be satisfied that the document is in order; that—judging from the description and portrait in it—it is, in fact, *my* passport and not somebody else's, and that it has not expired. If a fake passport appeared to be in order and seemed to belong to the bearer, then it would be accepted and the traveler would be allowed to pass. Passports work on trust, not on truth.

Because a false or altered passport would be trusted if it appeared correct, forged and stolen passports have value for individuals not eligible for a valid passport of the kind they would like or who prefer for some reason to travel using a false identity. In 1994, my passport was stolen when abroad, and the local US embassy issued me a replacement passport clearly marked as valid for one year

only. It was later renewed for another nine years, but the renewal statement was hidden inside the back. So for nine years this renewed passport appeared at first sight to have expired. Most guards noticed the original expiration date and then looked for evidence of renewal, but a significant number did not notice the expiration because they expected travelers to have current passports and did not examine it carefully enough to see that mine had apparently expired.

Frontier guards now usually run the passport's machine-readable code through a reader, and so delegate verification to some remote machine that compares the passport's codes with records already stored there. In other words, the guardian role is partly delegated to machine-readable codes, reading devices, and, somewhere, a machine programmed to respond to the encoded evidence. The human guard only needs to see that I resemble sufficiently the photographic portrait in the passport. Biometric technology has been developed to which that visual task could be delegated, so it is not hard to imagine a passport control station operating without any direct action by a human guard, much as grocery stores and libraries have experimented with self-service checkout.

The passport is evidence offered as a substitute for firsthand knowledge of a person's identity and citizenship. Its use depends on social regulations backed by military force, and also on cognitive activity: the guard has to read

it and believe that the passport is valid and that it is being used by the proper bearer. Finally, the machine-readable codes make it into a piece of machinery engaged in complex systems.

This small, printed booklet is a complex, dynamic, multimedia device with print, manuscript, and machine-readable scripts. It is carefully designed both to resist improper alteration and also to be changed in permitted ways. The passport plays a significant social role as a device used to control personal travel, and it is also widely used to serve other purposes when establishing one's identity is needed, for example, when boarding an airplane or dealing with a bank. This combination of varied physical features, cognitive perceptions of it as trustworthy, and use as a tool for social control makes it a rich example. Modern passports came into use a century ago, and the role, complexity, and powerful affordances of my passport make it a suitable emblem of contemporary society.

The Division of Labor and the Need to Know

The shift to dependence on documents has a long history. Cultures have developed from hunting and gathering to agriculture, industry, and sophisticated services. Common themes in these developments are the division of labor and increased interdependence of people and institutions.

As individuals, very few of us grow our own crops, kill the animals we eat, milk cows ourselves, or grow our own coffee beans. Similarly, we do not, by ourselves, make the technology we use, construct the buildings we live in, or generate the energy resources we depend on. Instead, we depend on others.

The division of labor allows us a higher standard of living through the development of specialized skills and greater efficiency from economies of scale, but, in consequence, we have become much more dependent on each other in many ways. We rely increasingly on other people, on technology, and on the infrastructure of transportation, financial services, regulations, and other developments that make this interdependence possible. Others, in turn, depend on us.

The exchange of goods and services requires markets, and markets depend on knowing what choices are available and on what terms. Markets are information systems. The better-informed buyers and sellers are, the more perfect the market is said to be. Less often stated is that markets and this interdependence also require an increase in communication and documentation. If we are to buy goods and services instead of providing our own, then we need to know who provides what we want, how much we shall need to pay, and whether what is offered is what we think it is. We can ask, of course, but mostly we depend (and anybody we ask will depend) on documents: price lists,

content descriptions, warranties, availability, limitations, and so on. Since we cannot ascertain entirely by ourselves everything we would like to know, we have no choice but to depend on what others tell us, and so we must also decide whom and what to trust. In this situation, it is unrealistic to make a distinction between believing and knowing. What you know is what you believe. The more confident you are in a belief, the more likely you are to consider it knowledge.

The increasing dependence on the knowledge of others—on "secondhand knowledge"—has two aspects: the ever-increasing division of labor, which makes us more dependent on others; and the ever-increasing reliance on communication (mostly documents) for the coordination that interdependence requires.

Culture and society develop through communication and collaboration. But, increasingly, we cannot communicate directly person to person. The best we can do is to use documents that record what that other person said, wrote, or did. The work of others and their ideas are incorporated in documents, both by them and about them, much as technology incorporates the labor of past inventors. Viewed this way, documents have become the connecting tissue that enables complex societies to function. Documents have increasingly become the means for monitoring, influencing, and negotiating relationships with others.

Documents have increasingly become the means for monitoring, influencing, and negotiating relationships with others.

Agendas of Others

The differing forms of documents and varied motivations in their use are reflected in the case of a letter written in May 1856 by Llewellyn Zublin to his son in Iowa about the dramatic murder of a newspaperman on a San Francisco street. The letter was later acquired by a dealer, who sold it to Berkeley's Bancroft Library for its collections supporting research on the history of California. In 2000, the letter was photocopied, keyed into a computer, and a printout produced, which was then used as copy by students in a class on hand-press printing. The students reproduced the letter as a pamphlet. Some extra copies of this elegant booklet were retained by the Bancroft library for possible use in its fund-raising. In 2001, one of the student printers reconstructed the history of this text in its successive formats for an assignment in a seminar on document theory—and, now, we use it as an example that shows the distinction between a text and text-bearing documents and involves quite varied forms of document (handwritten letter, photocopy, digital file, computer printout, and hand-press printing). It also reflects very varied motivations: family friendship, commerce, support for historical research, technical training, a library's fund-raising, obtaining educational credit, and academic theorizing.

It is not only our own needs that should concern us in understanding the role of information in society, but also the agendas of others. Examples are easy to find:

- schools use textbooks to guide our learning and to control teachers.
- religions use sacred texts to inspire particular beliefs and obedience.
- artists produce images to please and challenge us.
- merchants invest heavily in advertisements to influence what we buy.
- politicians make statements to seek votes and support for their campaigns.
- entertainers use varied media to amuse us and generate income from us.
- individuals use messages to communicate and attract attention.
- museums use the selective presentation and interpretation of objects to explain the past.
- mass media constantly transmit programs to entertain us, influence us, and satisfy advertisers.
- libraries provide access to selective collections of documents to facilitate our reading.

- social media allow the very rapid dissemination of comments.
- recording devices are used to monitor our purchases, movements, and behavior.

This list could be extended indefinitely. As the list builds up, we see more and more of our lives included. The choice of examples is less important than the cumulative evidence that our lives are permeated by messages, records, and documents used to influence our behavior and to shape our culture.

Information Society

Any claim that our "information society" is special or remarkable implies a contrast with some other "non-information society" that is different in some noteworthy way. But since all groups, and by extension all societies, develop their collective character through shared activity, through collaboration and communication, a "non-information society" would be a contradiction in terms.

It is unlikely that medieval people were less talkative than modern people. The significant difference between the most-developed current societies and less-developed societies cannot be the fact that information itself is more important but, rather, that some aspects of development referred to by the phrase *information society* involve wider use of information. My passport and the examples on the list provide an explanation. The passport is a nonoral form of information: a document. We can see that although the actions themselves are mostly not, in themselves, new, changes in technology have facilitated a greatly increased activity. The real change is in the rise of records. It would be more accurate to speak of an emerging *document society*.

Truth, Trust, and Belief

A traditional, academic view is that knowledge is *justified* true belief, that information, which leads to knowledge, must therefore be true, by definition, and that knowledge is (or could be) composed of a series of propositions. But even in the world of analytical philosophy, this is problematic, except for the understanding that knowledge is belief. In our daily lives, the presumption that all information is, by definition, true has no basis in common sense or personal experience. We may want to know if a statement is true, and we may well be doubtful, but in practice we usually end up trusting the evidence, an expert, a wise person, or a friend. Without trust, we would be paralyzed. Our

relationship with documents is based on trust, which becomes more necessary as well as more problematic as communications become less and less direct.

Referring to the patterns in the physical universe as "information" appears redundant or metaphorical. The characteristics of the universe (shapes, forms, patterns, physical processes, and so on) are what they are, and so issues of truth do not arise. Calling all physical differences information seems more confusing than helpful.

The Structure of This Book

We started by emphasizing our concern with information in everyday life, with a realistic approach to documents and data, rather than with more formalistic analyses in engineering. After some cautionary comments on uses of the word *information*, we then considered how a passport and a handwritten letter reveal the pervasive roles of documents in society, both in enabling the division of labor and in the advancement of many different agendas.

The following chapters note the rising flood of data, documents, and records of many kinds and the way they are used as well as analyses of how we cope with information. Chapter 2, "Document and Evidence," reviews different meanings of the word information, outlines the dramatic long-term growth of documents and data (the

"information explosion"), and notes the rise of techniques and initiatives to handle their organization, discovery, and use. Chapter 3, "Individual and Community," examines what individuals do with information, what communities know, the central role of culture, and how there are always physical, mental, and social aspects to information. Chapter 4, "Organizing: Arrangement and Description," is a summary of how collected documents are arranged and described so that we can identify and find a copy when needed. Chapter 5, "Naming," considers the nature and complexity of describing. Chapter 6, "Metadata," discusses how document descriptions, also known as metadata, are used for two purposes: to characterize documents and, by making indexes, to find the ones we want. Chapter 7, "Discovery and Selection," introduces the matching of queries and metadata for locating known documents and the more difficult task of identifying previously unknown documents likely to be of interest. Chapter 8, "Evaluation of Selection Methods," explains the standard evaluation of selection methods and acknowledges problems with relevance. The final chapter, "Summary and Reflections," repeats key points from earlier chapters and considers some of their implications for how we should understand information in society.