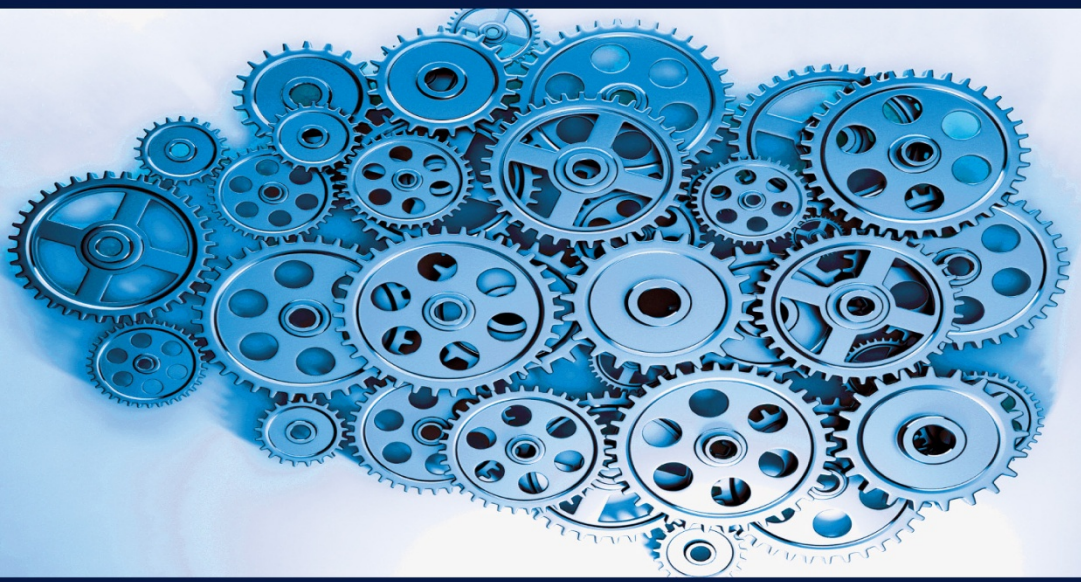
INFORMATION SYSTEMS, WEB AND PERVASIVE COMPUTING SERIES

INTELLECTUAL TECHNOLOGIES SET



Volume 4

Digital Humanities

History and Development

Olivier Le Deuff

iSTE

WILEY

Digital Humanities

Thanks to Michael Buckland for having opened up this path.

To all past, present and current players in digital humanities.

To all those whom I have not cited but whom I have nevertheless not forgotten.

Intellectual Technologies Set

coordinated by
Jean-Max Noyer and Maryse Carmes

Volume 4

Digital Humanities

History and Development

Olivier Le Deuff

ISTE

WILEY

First published 2018 in Great Britain and the United States by ISTE Ltd and John Wiley & Sons, Inc.

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms and licenses issued by the CLA. Enquiries concerning reproduction outside these terms should be sent to the publishers at the undermentioned address:

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

© ISTE Ltd 2018
The rights of Olivier Le Deuff to be identified as the author of this work have been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

Library of Congress Control Number: 2018931940

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN 978-1-78630-016-4

Contents

Introduction ix

Chapter 1. The Republics of Letters: the Need to Communicate and Exchange 1

1.1. Republic of Letters 2
1.2. The role of journals and the beginning of scientific information. 6

Chapter 2. The Science of Writings and Documentation 13

2.1. The importance of media and technology. 15
2.2. A science of writing? 16
2.3. From bibliotheconomy to bibliology 18
2.4. Between documentation and documentality 19

Chapter 3. From Lists to Tables, the Question of Indexing 23

3.1. In the beginning was the index 24
3.2. The need to handle information 28
3.3. Index and hypertext 29
3.4. Indexing as a design 30
3.5. Indexing knowledge versus indexing existences. 33

Chapter 4. The Need to Find Information 37

4.1. Information overabundance 37
4.2. The review process 40
4.3. Retrieving information 41
4.4. Between editorialization and information architecture 43

Chapter 5. The Researcher's Workstation and the History of Hypertexts 47

5.1. A hypertextual history 49
5.2. Paul Otlet and proto-digital humanities 53
5.3. The success of the Web 59

Chapter 6. The Quantitative Leap: Social Sciences and Statistics 65

6.1. Statistical reasoning 65
6.2. A dispute over methods? 70
6.3. François Furet and history 72
6.4. Between totalitarian science and the end of science 75
6.5. Digital literacy or overcoming opposition 78

Chapter 7. Automatic Processing: Concordances, Occurrences and Other Interpretation and Visualization Matrices 81

7.1. Ostranenie 81
7.2. Concordances 83
7.3. Corpus creation 88
7.4. The word association method 90

Chapter 8. Metadata Systems 93

8.1. Cataloging 94
8.2. Markup language 96
8.3. Folksonomies and the path of open digital humanities 98
8.4. The Web of data, from tree to graph 99

Chapter 9. The New Metrics: From Scientometrics to Webometrics 101

9.1. Bibliometrics and scientometrics 101
9.2. The reduction of science 103
9.3. The statistical expression of science 105
9.4. Nalimov's works 106
9.5. Scientometrics and traditional tools 108

Chapter 10. The Map: More than the Territory 113

10.1. Cybergeography – a forerunner of digital studies? 114
10.2. The map is not (only) the territory 117
10.3. Social networks and sociometry 118
10.4. A cartographic esthetic 120
10.5. Modeling and schematics 122

Conclusion 123

Bibliography 131

Index 147

Introduction

> "As the past has ceased to throw its light upon the future,
> the mind of man wanders in obscurity"
>
> Alexis de Tocqueville,
> *Democracy in America*, Volume II, 1840

This book does not aim to cover the complete history of digital humanities. Its aim is to propose a specific trajectory in history, in an attempt to highlight several trends that place current digital humanities in older and more complex genealogies.

The objective of this book is to examine the long history of digital humanities by highlighting important projects in the history of science and the organization of information from antiquity up to the present day, with an emphasis on indexation and tools for automatic corpus processing.

We identify several genealogies that show that digital humanities are not as recent as we might imagine and that renewal is the consequence of approaches already in use for a very long time. By closely following the themes that have been amplified since the early 20th Century by the fusion of social science and mathematics (and linguistics, statistics, sociology and information science) and by studying the consequences of thoughts on profound documentation problems, this work attempts to show the strong trends in "humanities" today – trends that show in the form of new "empirical

data": the increasingly significant role of algorithms, transforming "text" and classificatory thinking.

This work tries to better understand the current trends in digital humanities by showing the relationship between them and the reasons behind this keen interest through the desire for transdisciplinary reform. Digital humanities are not completely new, nor did they emerge spontaneously; they are based more on a convergence [JEN 06] than on a revolution. Their origins can be found in the previous centuries. By going all the way back to the first index creators like Jean Hautfuney, we intend to show that the relationship between social sciences and human sciences with technical tools is not new and that several interesting attempts were made towards developing efficient documentary instruments in the 20th Century.

The works of Belgian visionary Paul Otlet as well as that of Emmanuel Goldberg are part of this history, even though they were forgotten for a while due to unknown reasons [BUC 06].

What is new might sometimes be old movements that are simply rewritten using apparently more evolved technology. The frequent use of the word 'revolution' is but a veil that covers complex historic structures not only within technical and organizational infrastructure, but also within disciplines, scientific creations and minds.

Consequently, digital humanities surpass computer vision based on human and social sciences. To reduce this history down to the emergence of computing tools and their use by human and social science researchers does neither help us understand the relationships nor the numerous interdisciplinary and transdisciplinary attempts that have made it possible to introduce new methodologies for understanding and analyzing. To reduce the role of digital humanities to the introduction of computer technologies is to reduce humanities and the humanist movement to the use of Latin as a common language during the days of Erasmus (1466–1536).

The history of digital humanities that we are proposing here is more than a simple relationship of human and social sciences to computer technologies. This history begins, in the gradual constitution of current science, where there is a need to distinguish objects, words and things, in keeping with the title of Michel Foucault's work [FOU 66]. Moreover, it is just as much an archaeology of digital humanities that must be undertaken, as our undertaking relies on traces that allow for us to better understand the historical, scientific and technical evolution of what we now call digital humanities.

I.1. A history of milieus of knowledge

"Learning circles" refer to classical places of knowledge like libraries, and also take devices for intellectual work, both individual and in groups, into consideration. In fact, the concept of a circle allows for easily integrating aspects of mediation and communication, notably documentary mediation, and also the role of media that renew interaction modes between professionals and users. Thus, the circle concept extends the concept of the "place of learning" [JAC 14] that puts forward a visible space, while the circle concept also refers to the concept of an "associated milieu", which was defined by Simondon [SIM 58] and made popular by Bernard Stiegler [STI 08], and which best articulates technical objects and social individuals. Associated milieus allow for the individualization of people and objects of which they are composed in an innovative and non-rigid manner, as users have an understanding of the device and there are possibilities to modify and improve tools and its users.

"Learning circles" are also mediation spaces that allow for a better understanding of the relationships that exist between mobilized devices and individuals. In these milieus, there are relations between people and knowledge, so much so that this history is based on the evolution of communication and information processes that will be interesting to scholars, researchers, and librarians, too [LIT 11].

This work is intended to be, above all, a starting point for more in-depth studies relating to the history of pre-digital humanities, that is periods, players, and devices that precede current digital humanities, and which cannot be ignored, as they constitute the founding of the current evolution.

There are no revolutions and radical rifts in digital humanities, but instead a boom linked to a gradual metamorphosis.

Digital humanities are based on an approach that goes beyond the simple computation of textual elements, without which we would consider automatic language processing as the main part of this domain. Before such tools were developed, a need to quickly and easily access information emerged in the previous centuries. A rise in the overabundance of information during the Renaissance is a continuation of the discovery of the New World, the rediscovery of ancient texts and the availability of more printed books. But finally, it is not only these aspects that led to an overabundance of information. In fact, the need to search for and compile information explains the growing infobesity of the "classical" period. During this period, researchers, scholars and compilers wished to preserve information. The loss of works dating from antiquity was decisive in their minds. The desire to gather information and knowledge drove people like Theodor Zwinger (1533–1588) to megalomania, who edited an encyclopedia after accumulating information. We can note a need to accumulate points and, more importantly, not lose them; such was the effect of having lost a major part of ancient texts that it became crucial to not face the same situation. Thus, compilers felt they contributed to the common good by undertaking this task. But it was also an essential intellectual step, as Francis Bacon (1561–1626) noted in *Sylva Sylvarum* [BAC 70], as there was a strong need to accumulate new information before elaborating it. This practice of accumulating coupled with note taking draws upon an ancient encyclopedic model with Pliny the Elder (23–79), who had started developing tools for classification, like tables of contents, to organize the wealth of accumulated information.

Consequently, summarizing and note taking were important for when the original information was lost. Moreover, Erasmus acknowledged his debt to these authors, whom he quotes extensively, but the originals were lost. The role of humanism in the accumulated knowledge of humanity and different civilizations is to develop knowledge. The role of digital humanities is to follow in this path without creating any rifts with classic methodologies and more specifically by undertaking archaeological work:

> "From the limit-experience of the Other to the constituent forms of medical knowledge, and from the latter to the order of things and the conceptions of the Same, what is available to archaeological analysis is the whole of Classical knowledge, or rather the threshold that separates us from Classical thought and constitutes our modernity. It was upon this threshold that the strange figure of knowledge called man first appeared and revealed a space proper to the human sciences. In attempting to uncover the deepest strata of Western culture, I am restoring to our silent and apparently immobile soil its rifts, its instability, its flaws; and it is the same ground that is once more stirring under our feet" [FOU 66, p. 15].

I.2. A critical perspective

Digital humanities developed a strong critical approach [LIU 08, BER 12b], concerning the methodologies that they mobilize. This book tries to extend this work by showing that current interrogations are placed in a more complex perspective; one which is broader than we think it is. Here, we must think about the organization of knowledge, the need to work on voluminous corpora, the concurrence of misinformation, and pseudo-scientific discourses that come from interrogating the role of human and social sciences in our current societies. We find that dreams and desires which determine the temptation to accumulate and conserve as much heterogeneous data as possible in order to study them do not ultimately differ so much from that of modern desires. If the accumulation creates a necessity to

categorize and eliminate doubtful information, practices are sometimes different. Natural history, in its beginnings, did not manage to do this work, and confused elements from field observations with elements from second-hand work that were of a phantasmagoric character. Furthermore, Bacon's *Sylva Sylvarum* can be translated as "forests of materials" [HEG 85], which means a very heterogeneous assemblage, whose scientificity is still not established. A shortcoming that was found with the compilation of Conrad Gesner (1516–1565) is the description of an existing plant being next to that of an imaginary creature like the monocerote, a kind of unicorn. This is what Georges-Louis Leclerc, comte de Buffon (1707–1788) denounced and deplored in the attempt by naturalist Ulisse Aldrovandi (1522–1605) to complete a scientific natural history, which Foucault rightly spoke of in *Les mots et les choses (The Order of Things)* [FOU 66].

Currently, the wish to accumulate data can also pose a problem with the feeling that accumulation can compensate for real work in selection and interpretation. This is the criticism against the heterogeneity of data and sources that succeeds the accumulation period.

Thus, this book belongs to an archeology of knowledge and methodologies, going beyond the current debates and successes of the digital humanities movement to try to better uncover the forms and norms that have led, little by little, to its development and emergence. This history becomes gradually difficult to write, facing the risk of being reduced down to a short history – that of the players and, notably, the pioneers – a timely but reductive one, as it often ignores the lines of action, and the continuities and discontinuities. A conceptual history seems to constitute an interesting alternative.

It is insufficient to observe some exemplary projects from the last five years in order to understand digital humanities. It is important to go back farther into the past and dwell longer in the study of the emergence of modern science, the source of these humanities that constitute half of the expression that we aim to analyze here:

> "If we question ourselves about the significance, at the dawn of modern times, of the completely unexpected triumph of the new science, we must first take off the lenses of Newtonian synthesis and try to start again from a series of interrogations that, towards the end of the 16th Century, were at the heart of the European debate. How does one come out of the crisis of the Aristotelian-Ptolemaic cosmology? How was the earth really made? What are its governing laws? In which language can they be translated?" [BUC 09, p. 19]

This bringing together of humanities, henceforth commonly called social and human sciences, and the digital that refers to computation and use of tools to improve observation and analysis, requires that we interrogate the phenomenon of division of knowledge domains that developed gradually during the previous centuries and particularly since the end of the 19th Century. This is also a re-interrogation of the role of researchers, those whom were then called savants, a word that we no longer dare to use, as there is so much information that one person cannot claim to know everything; it is due to this that one must specialize. The savant refers more to personalities like Albert Einstein, who belong to the popular imagination. Others like Vannevar Bush can also be given this title. Still in the popular imagination, the savant refers henceforth to the image of someone who is on the verge of the forbidden, so much so that he is nearly crazy, taking insane risks with the present and the future, like doctor Frankenstein. If the new man thus envisaged by science fiction presents to us a science that has lost all logic, what role can be given to digital humanities vis-à-vis theories and methodologies of post-humanism [DAV 10], even transhumanism? Milad Doueihi [DOU 11] put forth the idea of digital humanism, reviving the principles of classical humanism to better understand current transformations.

However, this humanism is possible only if we shine light on the gradual constitution of humanities and if we highlight the contexts, methodologies and ideas involved in its gradual emergence. And this development was not possible without the evolution of means of information and communication, which have made the idea of the