Multimedia Systems and Applications

Series editor

Borko Furht, Florida Atlantic University, Boca Raton, USA

More information about this series at http://www.springer.com/series/6298

Shalin Hai-Jew
Editor

Data Analytics in Digital Humanities

Springer

Editor
Shalin Hai-Jew
Kansas State University
Manhattan, KS, USA

Multimedia Systems and Applications
ISBN 978-3-319-54498-4 ISBN 978-3-319-54499-1 (eBook)
DOI 10.1007/978-3-319-54499-1

Library of Congress Control Number: 2017935738

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This book is for R. Max.

Preface

The digital humanities, putatively the intersection between the humanities disciplines and computation, was popularized in the early 1990s. In the intervening decades, the "digital humanities" has not yet settled on a defined self-identity. One indicator of this is that of the dozens of "DH" manifestos on the Web; they all have differing and competing visions for the field. Another indicator is the rich variety of work being done under these auspices that does not fall into simple summaries and descriptions. The Digital Humanities Manifesto 2.0, which originated from nine seminars co-taught by Jeffrey Schnapp and Todd Presner, and was released by the UCLA Mellon Seminar in Digital Humanities, reads in part:

> **Digital Humanities** is not a unified field but an array of convergent practices that explore a universe in which: (a) print is no longer the exclusive or the normative medium in which knowledge is produced and/or disseminated; instead, print finds itself absorbed into new, multimedia configurations; and (b) digital tools, techniques, and media have altered the production and dissemination of knowledge in the arts, human and social sciences. The Digital Humanities seeks to play an inaugural role with respect to a world in which universities—no longer the sole producers, stewards, and disseminators of knowledge or culture—are called upon to shape natively digital models of scholarly discourse for the newly emergent public spheres of the present era (the www, the blogosphere, digital libraries, etc.), to model excellence and innovation in these domains, and to facilitate the formation of networks of knowledge production, exchange, and dissemination that are, at once, global and local" ("The Digital Humanities Manifesto 2.0," 2009).

DH work is additive and contributes to learning and methods in a range of extant fields. It is about hybrid mash-ups that pull from analog and digital sources and that give back digitally. There are combinations of research and analytical methods, drawing from quantitative and qualitative methods, going Web scale and small scale, employing a range of frameworks, and combining various disciplines, often through collaborative and distributed teams.

Deconstructing "DH"

The "digital" part of the "digital humanities" involves a range of technologies: the Web and Internet, mobile devices, social media, databases, digital curation platforms, geographical mapping tools, social mapping tools, linguistic analysis software programs, data analytics tools, and others. The "humanities" focus points to a combination of fields: the classics, literature, philosophy, religious studies, psychology, modern and ancient languages, culture studies, art, history, music, philosophy, theater, film, political science, geography, anthropology, linguistics, social work, communications, and others. For all the technological focuses, the DH work is about people and for people through the ages and for the ages. Every work, no matter how technological, contains the human touch and is conceptualized for human benefit. The works are human-readable, in general, and are packaged as stories and understandings. The works are often aesthetically dazzling: think interactive webscapes, data visualizations, 3D virtual immersive spaces, and network diagrams.

Likewise, one can ask what "data" in the digital humanities is. Well, the general opinion is that everything has some informational value. The world is informatized, and DH researchers themselves are not only researchers and elicitors of insights, but they are in-world participants making the world and information simultaneously. The information is viewed through a subjective and interpretive lens, but that interpretation is backed up by more traditionally quantitative (statistical) and computational (machine learning) methods. The work is not seen as neutral nor theory-free not independent of its contributors.

The challenge is in how to turn a "thing" in the world into something with informational and social value. This data may seem at once a range of contradictions: ephemeral and fragile and yet permanent and perpetual, subjective and objective, created for the local but also the global, inclusive of machine-scale distant reading methods as well as close human reading.

The Spirit of DH

Often, the spirit of the digital humanities is countercultural; it fights against extant power structures and norms. The DH is often value-informed and socially active, in the pursuit of social justice. DH practitioners take a questioning stance because the world is never quite as it seems. There are elusive and latent interpreted truths to be discovered or (co)created.

While its core spirit is of changing the current order, at the same time, DH work requires ever higher levels of expertise in content fields and technological methods. The work is anti-Establishment but requires some of the expertise of Establishment and accrued skills from years of development. Virtuosity in the digital humanities simultaneously requires the fiery rebel spirit and the iciness of acquired expertise. The work is often inclusive of nonexperts and the online crowds, who are sourced

for their contributions and voices and concerns in rich ways. It seems fitting that in this hypersocial age that such broadscale collaborations exist.

The spirit of DH is of experimentation along the entire work chain: theorizing and conceptualization, research, data collection, content curation, data processing, data analytics, and often open publishing (of digital corpora and collections, of virtualized experiences, of publications, and of multimedial presentations). The order of the work is not necessarily linear but recursive; for some projects, conceptually and somewhat practically, the work is continuous, ongoing, and perpetual. While some methods and technologies are borrowed from a range of fields, digital humanists have applied experimental means to every step of the process as well, with new technological platforms and tools and fresh methods. Problem-solving is achieved on-the-fly, often with the direct support and help of Web developers, coders, server administrators, data scientists, computer scientists, and librarians. This is about an artist and artisan sensibility with code.

If one asks practitioners what the digital humanities is, there is a range of answers. A common response is to go to the work, to explore it in depth, and to acquire a more inductive sense of the answer. Certainly, as DH rose from a groundswell of practice and self-definition, it may be best to understand it from a bottom-up instead of a top-down way. While practitioners are aware of each other's work, the focus is not on conformance to common practice but rather diversity. The general consensus is that DH is still emerging as an approach. The feeling is that the work is exploratory and provisional and open to re-interpretations and re-visionings.

About This Book

Data Analytics in Digital Humanities was over a year in the making, with various individuals and authoring teams conducting their research and writing it up, going through double-blind peer reviews, revising, and finalizing their work. Here, the respective authors describe their data analytics work in the digital humanities.

The book is comprised of six parts:

Part I: Design of Representational Systems
Part II: Text Capture and Textual Exploration
Part III: Engaging Social Data
Part IV: Applied Technologies for Data Analytics
Part V: Sense-Making in the World
Part VI: Support for Digital Humanities Work

The book is not comprehensive by any means. There are so many live projects and endeavors that each author and team really only have a limited perspective on the whole.

In Part I, "Design of Representational Systems," there are two works that describe data labeling. In "Semantic Web for Cultural Heritage Valorization," a research team proposes a set of ontology modules (Cultural-ON) to model the

classification of cultural heritage data, given the heterogeneous nature of cultural artifacts, both analog and digital, tangible and intangible. The authors of Chap. 1 include Dr. Giorgia Lodi, Luigi Asprino, Andrea Giovanni Nuzzolese, Dr. Valentina Presutti, Dr. Aldo Gangemi, Dr. Diego Reforgiato Recupero, Chiara Veninata, and Annarita Orsini.

Dr. Gian Piero Zarri offers an in-depth description of Narrative Knowledge Representation Language (NKRL) and its use to describe narratives as logical associations based on "elementary events." The careful specifications enable its use conceptually and computationally to capture and represent the elements that comprise narratives. Zarri's chapter is titled "Using the Formal Representations of 'Elementary Events' to Set Up Computational Models of Full 'Narratives'" (Chap. 2).

The second part, "Text Capture and Textual Exploration," features two works related to text analytics. Joshua L. Weese, Dr. William H. Hsu, Dr. Jessica C. Murphy, and Dr. Kim Brillante Knight describe a machine learning-based classifier they developed to identify parody, an elusive challenge. Their work is titled "Parody Detection: An Annotation, Feature Construction, and Classification Approach to the Web of Parody" (Chap. 3).

Dr. Michael Percillier, in Chap. 4 "Creating and Analyzing Literary Corpora," describes a methodical process using Python and other tools to collect and process literary corpora based around a particular topic. He includes Python code to demonstrate efficient automated means to build text corpora.

Part III, "Engaging Social Data," offers two works summarizing "social" data. The first is about information from online social networks, and the second about the social aspects of learning.

Davide Di Fatta and Roberto Musotto apply integrated sentiment analysis (iSA) to online social networks in Chap. 5, "Content and Sentiment Analysis on Online Social Networks (OSNs)." They consider various ways to apply their research insights to practical applications, like marketing through social media.

In Chap. 6: "The Role of Data in Evaluating the Effectiveness of Networked Learning: An Auto-Ethnographic Evaluation of Four Experiential Learning Projects," Jonathan Bishop employs a light auto-ethnographic approach to study the role of data in evaluating networked learning across four experiential learning projects. One central question is whether there are more effective designed electronic stand-ins for educator-learner interactions to promote learning.

In the fourth part, "Applied Technologies for Data Analytics," one work focuses on computational linguistic analysis based on psychological features in texts. Another focuses on capturing research insights from related tags networks.

In Chap. 7, "Psychological Text Analysis in the Digital Humanities," Ryan L. Boyd describes insightful uses of the Linguistic Inquiry and Word Count (LIWC) tool for text analysis. His long-term uses of LIWC and direct contributions to the tool's co-development (in LIWC2015) make him a singularly appropriate author for this excellent work. Boyd argues that computational linguistic exploration of psychological aspects in language is an untapped and highly promising area of research.

In "Parsing Related Tags Networks from Flickr to Explore Crowd-Sourced Keyword Associations" (Chap. 8), Dr. Shalin Hai-Jew describes how crowd-based folk tags applied to digital imagery on a content-sharing site may be used for creating collective mental models of large-scale phenomena.

Part V is about "Sense-Making in the World." In this part, Dr. Cobi Smith describes the use of technological methods to harness crowd-sourced imagery during natural disasters. The methods and tools described in this chapter, "A Case Study of Crowdsourcing Imagery Coding in Natural Disasters" (Chap. 9), may have broad applications in the digital humanities.

Dr. Glenda Alicia Leung's "YouTube Comments as Metalanguage Data on Non-standard Languages: The Case of Trinidadian Creole English in Soca Music" (Chap. 10) suggests new analytical applications of interaction comment data related to shared, social videos. This work captures the energy and power of nonstandard lived language in connecting people on Google's global video-sharing platform.

Chapter 11 is about "Creating Inheritable Digital Codebooks for Qualitative Research Data Analysis." Here, Dr. Hai-Jew describes the research importance of sharing digital codebooks, particularly with the popularization of Computer Assisted Qualitative Data AnalysiS (CAQDAS) tools, and she offers firsthand insights on effective processes for developing and sharing such digital codebooks. If researchers are to leave a legacy of their unique coding "fists," then computational means may offer a more convenient and efficient way of transfer.

In Part VI, "Support for Digital Humanities Work," one author highlights supports for DH work. Hannah Lee, in "Is it Worth It? The Library and Information Science Degree in the Digital Humanities" (Chap. 12), argues for the importance of library and information sciences (LIS) to practically support the work in the digital humanities.

So what is the state of data analytics in the digital humanities? Based on these collected works, it is a rich and evolving one, driven by local research needs and some global ones. DH practitioners harness technologies to complement and augment the human abilities of perception, expression, analysis, and memory. The technologies used range from open-source and free to closed-source and proprietary, and these tools are cobbled in creative and complex sequences for understanding and analysis. The skill comes not only in the applied techniques but also in the insights surfaced and the sharing of the applied and innovative techniques. While some data analytics results are "reproducible" and "repeatable" based on computational means, the assumptions underlying the DH research and data analytics are very much drawn from qualitative data analytics:

that all human phenomena have potential informational value, depending on researcher perspective and context (and vision and skill);
that human researchers are wrapped in their own subjectivities, for better and for worse, and benefit from deeper self-understandings through practiced reflection;
that all data are filtered through subjective lenses and self-informed understandings;
that data "measures" are limited, imprecise, conditional, and contested (and yet still are insightful);

that the researchers are part and parcel of the research and inform the research and the data findings and their applications;
that the communal context for the humanities is a critical part of the work (as research participants, data consumers, and researchers);
that DH research is subsumed to human needs, interests, and values;
and that social justice and practitioner ethics apply to every context.

There are inherent digital humanities truisms about data as well. For example, writings considered classically "fictional" contain the seeds of truth, and traditionally non-fictional works may not have as much truth as advertised. In this light, there are critical interrogations to be made, to understand the multihued shades between truths. Data are extracted in fresh ways, with manuscripts tagged and mapped for patterns (linguistic, spatial, psychological, cultural, and others), dialogs mined for insights, genders explored, and cultural practices probed, across and through time. All human residua contain substance that may be interpreted and informatized for new ways of seeing, feeling, and being.

DH cartographers describing data analytics in the digital humanities are in the early phases of defining and mapping this space. In theory and in practice, there are numerous other potentials that have yet to be explored, applied, and shared.

Data Analytics in Digital Humanities offers an early look at this topic, with hand-sketched DH data analytics "maps" of fine granular detail for particular defined needs but which does not yet have the cardinal directions defined or accepted compass roses.

Manhattan, KS, USA Shalin Hai-Jew

Acknowledgments

Data Analytics in Digital Humanities would not exist were it not for the work of the respective authors. I owe each one a deep debt of gratitude. Also, thanks to Joelle Pitts, Dr. Brian Rosenblum, Nancy Hays, Dr. Bryan Alexander, Dr. Glenda Alicia Leung, and others for their help in publicizing the original "call for chapters" through their respective professional and social networks. There is a lot to be said for sharing social networks, especially for such a "big ask" as writing a chapter. At Springer, Susan Lagerstrom-Fife was seminal in the acceptance of the initial book proposal, and Jennifer Malat and others on the Springer team were critical in ensuring that this work came together. Thanks to all! Their support has enhanced the quality of the text and ultimately made this work possible. Mr. Dinesh Shanmugam and Ms. Sarumathi Hemachandirane both provided very important support in bringing the book to final published form.

Contents

Authors' Biographies

Luigi Asprino is a Ph.D. student in computer science and engineering at the University of Bologna under the supervision of Prof. Paolo Ciancarini and Dr. Valentina Presutti. He is also a research assistant at the Semantic Technology Laboratory (STLab) of the National Research Council (CNR), in Rome. In October 2014, he earned a master's degree in computer engineering from Sapienza—University of Rome. In January 2015, he joined STLab. During his period in STLab he actively contributed to the following Italian and European projects: FOOD, MARE, MiBACT, MARIO. His research interests include knowledge integration, web service composition, ontology design, and semantic web.

Jonathan Bishop is an information technology executive, researcher, and writer. He founded the Centre for Research into Online Communities and E-Learning Systems in 2005, now part of the Crocels Community Media Group. Bishop's research and development work generally falls within human-computer interaction. He has over 75 publications in this area, including on Internet trolling, cyberstalking, gamification, cyberlaw, multimedia forensics, Classroom 2.0, and Digital Teens. In addition to his B.Sc. (Hons) in Multimedia Studies and various postgraduate degrees, including in law, economics, and computing, Bishop serves in local government as a councilor and has been a school governor and contested numerous elections, including to the UK Parliament and Welsh Assembly. He is a fellow of BCS, CILIP, the InstAM, the RAI, the RSS, and the RSA, senior member of IEEE, and a member of the IMarEST with MarTech. Bishop has won prizes for his literary skills and been a finalist in national and local competitions for his environmental, community, and equality work, which often form part of action research studies. In his spare time, he enjoys listening to music, swimming, and playing chess.

Ryan L. Boyd is a Ph.D. candidate in social/personality psychology at the University of Texas at Austin. His cross-disciplinary research revolves around using computational techniques to explore and understand cognitive-behavioral links, primarily the link between a person's language and their mental worlds. Specifically, Ryan's work on the psychology of language includes high-dimensional personality assessment, values, psychological forensics, motivations, and health, among other

topics. Ryan has taught multiple workshops on machine learning, data mining, and language analytics and is considered a leading expert on the psychology of language. He is the chief data scientist at Receptiviti and is the creator of several language analysis programs, including the Meaning Extraction Helper. Ryan is also the co-creator of LIWC2015, one of the most widely used psychological language analysis tools in the field of text analysis.

Davide Di Fatta is a Ph.D. student in economics and management at the University of Messina (Italy). He collaborates with the SEAS Department at the Polytechnic School of the University of Palermo (Italy). His main research fields are, on the one hand, e-commerce, e-marketing, digital marketing, and social media marketing; on the other hand, business systems, systems thinking, and social networks. Davide is editorial assistant for many international journals such as Kybernetes and IJMABS (in systemic field); IJEMR and IJMS (in marketing field). He is also Junior Member Manager for the Business System Laboratory (BS Lab), a nonprofit organization for the promotion of research and collaboration between the universities and the enterprise systems.

Aldo Gangemi is professor at Paris 13 University, and researcher at CNR, Rome. He has founded and directed the Semantic Technology Lab (STLab) of ISTC-CNR. His research focuses on Semantic Technologies as an integration of methods from Knowledge Engineering, the Semantic Web, Linked Data, Natural Language Processing, and Cognitive Science. Publications: https://scholar.google.com/citations?user=-iVGcoAAAAAJ. Software: http://wit.istc.cnr.it/stlab-tools.

Shalin Hai-Jew has worked on a number of digital humanities projects over the years, each harnessing a variety of methods and technologies. She has B.A.s in English and psychology and an M.A. in English from the University of Washington (Seattle), and an Ed.D. from Seattle University (2005). At the University of Washington, she received the Hugh Paradise Scholarship; at Seattle University, she was a Morford Scholar. She worked for The Boeing Company as a Faculty Fellow for two summers. She lived and worked in the People's Republic of China from 1988 to 1990 and 1992 to 1994, the latter 2 years with the United Nations Volunteer Programme of the United Nations Development Programme (UNDP). She works as an instructional designer at Kansas State University. She has published several authored texts and edited a number of technology-related texts. She reviews for a half-dozen academic publishers. Her research interests include online learning, computational research methods, and social media.

William H. Hsu is an associate professor of Computer Science at Kansas State University. He received a B.S. in mathematical sciences and computer science and an M.S. Eng. in computer science from Johns Hopkins University in 1993, and a Ph.D. in computer science from the University of Illinois at Urbana-Champaign in 1998. He has conducted research in digital humanities since 1998, where at the National Center for Supercomputing Applications (NCSA) he was a co-recipient of an Industrial Grand Challenge Award for visual analytics of text corpora and unsupervised learning for topic modeling. His research interests include machine

learning, probabilistic reasoning, and information visualization, with applications to digital humanities, cybersecurity, education, geoinformatics, health informatics, and medical informatics. Published applications of his research include spatiotemporal event detection for veterinary epidemiology, crime mapping, and opinion mining; structured information extraction; and analysis of heterogeneous information networks. Current work in his lab deals with geospatial data science and analytics, data mining, and visualization in education research; graphical models of probability and utility for information security; and domain-adaptive models of natural language corpora and social media for text mining, link mining, sentiment analysis, and recommender systems. Dr. Hsu has over 100 publications in conferences, journals, and books.

Kim Brillante Knight is an associate professor of emerging media and communication at the University of Texas at Dallas. Her research focuses on digital culture, negotiations of power and the formation of identity, particularly in relation to marginalized groups. In her book, *Viral Structures in Literature and Digital Media: Networked Counterpublics and Participatory Capitalism* (forthcoming from Routledge), Dr. Knight addresses the role of digital media as it circulates outside of broadcast paradigms and empowers or oppresses subjects in network society. Her latest book, *Fashioning Makers and Counterpublics: Critical Making and Public Humanities*, is under advance contract with the University of Iowa Press. Her public work can be found at kimknight.com, thespiraldance.wordpress.com, and fashioningcircuits.com. More information may be accessed at: https://www.utdallas.edu/ah/events/detail.html?id=1220419626

Hannah Lee is a double alumni of UCLA, obtaining her B.A. in English and MLIS. A doctoral student at Fielding Graduate University in Santa Barbara, the University of California, Los Angeles (UCLA), she is focusing on specializing in the future developments of information systems and technology in the library and information sciences field. Her dissertation examines an organizational approach to the book publishing industry in the United States. She works as a librarian, research analyst, and informationist in the Los Angeles area. Some of her interests include copyright, intellectual property, book arts, mentoring, and volunteering for nonprofit organizations.

Glenda Alicia Leung is an independent researcher and sociolinguist from Trinidad and Tobago. She received her B.A. in English from the University of Florida and her M.A. in applied linguistics from Ball State University before earning her doctorate in English linguistics at the University of Freiburg, Germany, in 2013. Her doctoral thesis "A Synchronic Study of Monophthongs in Trinidadian English" was a quantitative acoustic and sociophonetic study that reported on vowel distribution in contemporary Trinidadian English. Her articles have appeared in *World Englishes*, *Multilingua*, and *Journal of Bilingual Education Research and Instruction,* as well as in the *Varieties of English Around the World Series* from John Benjamins. Her research interests include creole language in performance, English language learning, and language acquisition.

Giorgia Lodi is currently a research assistant at the Institute of Cognitive Sciences and Technologies (ISTC) of the National Council of Research (CNR)—Semantic Technology Laboratory (STLab) and a consultant at "Agenzia per l'Italia Digitale"—AgID. She works in such areas as open government data, Linked Open Data, Semantic Web Technologies, Distributed Systems, Big Data, and smart communities. She received a Ph.D. in computer science from the University of Bologna (Italy) in 2006. Publications: https://scholar.google.com/citations?user=W6gWCGUAAAAJ

Jessica C. Murphy is Associate Professor of Literary Studies at the University of Texas at Dallas. Her research focuses on early modern literature, culture, gender studies, and digital humanities. Murphy's publications include a book, *Virtuous Necessity: Conduct Literature and the Making of the Virtuous Woman in Early Modern England*, published by the University of Michigan Press in 2015; journal articles; solo-authored and co-written book chapters; and a co-edited collection.

Roberto Musotto is an Italian Barrister who works on commercial and property law, trusts, and organized crime. He is a Ph.D. student in economics at the University of Messina and a research and teaching assistant at the University of Palermo in Political Economy. His fields of research interest include spatial and social network analysis, organized crime, commercial law, and game theory.

Andrea Giovanni Nuzzolese is a Postdoctoral Researcher at the Semantic Technology Laboratory (STLab) of the National Research Council (CNR) in Rome, Italy. He received a Ph.D. in computer science in 2014 from the University of Bologna (Italy). His research interests include Knowledge Extraction, Ontology Design Patterns, Linked Data, and Semantic Web. He has been a researcher in the EU-funded project IKS (Interactive Knowledge Stack) and a main developer of Apache Stanbol software stack (reusable components for semantic content management). He has published scientific papers in international journals and conferences.

Annarita Orsini with a master's degree in law, is an employee of the Italian Ministry for Heritage and Cultural Activities and Tourism currently working in the specific domain of the tourism for the definition of common standard data models for key data such as point of interests, accommodations, and tourist guides. In recent years she was head of unit of the Directorate General Organization of the Italian Ministry for Heritage and Cultural Activities and Tourism working in such domains as Open Data, Linked Open Data, technical interoperability for public administrations, to cite a few.

Michael Percillier is a postdoctoral researcher at the University of Mannheim, Germany. His areas of specialization include corpus linguistics, World Englishes, literary linguistics, and historical linguistics. His research has covered the comparative study of institutionalized and learner varieties of English in Southeast Asia, the representation of nonstandardized varieties of English in literary texts, and the contact effects of Old French on Middle English in the medieval period.

Valentina Presutti is a researcher at the Semantic Technology Laboratory of the National Research Council (CNR) in Rome, and she is associated with LIPN (University Paris 13 and CNRS). She received her Ph.D. in computer science in 2006 at the University of Bologna (Italy). She has been the scientific coordinator for CNR in the EU-funded project IKS (Interactive Knowledge Stack—A framework for Semantic CMS that bring semantics at the interaction level), and she has been one of the main researchers in the EU-funded projects NeOn (Networked Ontologies—methods and tool for design of networked ontologies). She has more than 50 publications in international journals/conferences/workshops on topics such as Semantic Web, knowledge extraction, and ontology design.

Diego Reforgiato Recupero has been an Associate Professor at the Department of Mathematics and Computer Science of the University of Cagliari, Italy, since December 2015. He is a programmer, software developer, automation and ICT expert. He holds a double bachelor from the University of Catania in computer science and a doctoral degree from the Department of Computer Science of the University of Naples Federico II. In 2013 he won a Postdoctoral Researcher position within the Semantic Technology Laboratory (STLAB) of the Institute of Cognitive Sciences and Technologies (ISTC) of the National Research Council (CNR) where he worked on Semantic Web and Linked Open Data; he is still an associated researcher at STLAB where he collaborates. He is co-founder of R2M Solution S.R.L. (Italian company), R2M Solution Ltd. (UK company), and Sentimetrix Inc. (US company).

Cobi Smith has a doctorate in science communication at Australian National University. She worked as a UN consultant based inside CERN on the EU-funded Citizen Cyberlab project. She currently works in state government in Australia, in emergency services information management and disaster risk reduction. After a degree in Cultural Heritage Preservation, she took a master's degree for qualified researcher in archival/documental analysis through the use of information technology and formal model (XML, XSL, XSLT, RDF). For more than 10 years she worked at the Rome Research Consortium and she gained expertise in the field of formal languages, markup languages, and ontologies. From 2010 to 2014 she worked at "Archivio Centrale dello Stato" in Rome. Since November 2014 she works at Directorate General Organization of the Italian Ministry for Heritage and Cultural Activities and Tourism.

Chiara Veninata after a degree in Cultural Heritage Preservation, took a master's degree for qualified researcher in archival/documental analysis through the use of information technology and formal model (XML, XSL, XSLT, RDF). For more than 10 years she worked at the Rome Research Consortium, and she gained expertise in the field of formal languages, markup languages, and ontologies. From 2010 to 2014 she worked at "Archivio Centrale dello Stato" in Rome. Since November 2014 she works at Directorate General Organization of the Italian Ministry for Heritage and Cultural Activities and Tourism.

Joshua L. Weese is a Ph.D. candidate focusing on CS education research. His education experience comes from work as a graduate teaching assistant as well as time spent as a NSF GK-12 fellow. Josh's research focuses on operationalizing computational thinking as well as assessing computational thinking, in outreach programs aimed toward enriching K-12 students' experience in STEM. Apart from CS education, Josh is highly active in developing data analytic tools for the physics education research community.

Gian Piero Zarri has been awarded an M.Sc. in electronic engineering (University of Bologna, Italy) and a Ph.D. in computer science (University of Paris XI-Orsay, France). After more than 30 years spent as Research Director at the French National Centre for Scientific Research (CNRS), he cooperates now with the STIH Laboratory of the Paris-Sorbonne University as a senior associated researcher. Professor Zarri is known internationally for having defined and developed NKRL ("Narrative Knowledge Representation Language"), a representation tool used in several EC-funded projects and characterized by the addition of an "ontology of events" to the usual "ontology of concepts." He is the author of about 200 refereed papers concerning Artificial Intelligence, Computational Linguistics, Information Retrieval, Intelligence and Global Security, etc., and is the editor, co-author, and author of several books on Expert Systems, Intelligent Databases, and Representation and Management of Narrative Information. Professor Zarri has co-operated intensively with industry and has worked with international organizations like the European Commission, UNESCO, and Eureka-Eurostar and with national research-funding agencies in France, Italy, Austria, Portugal, the Netherlands, etc.