Synthesis Lectures on
Information Concepts, Retrieval, and Services

SYNTHESIS
COLLECTION OF TECHNOLOGY

Rhiannon Bettivia · Yi-Yun Cheng ·
Michael Robert Gryk

Documenting the Future: Navigating Provenance Metadata Standards

Springer

Synthesis Lectures on Information Concepts, Retrieval, and Services

Series Editor

Gary Marchionini, School of Information and Library Science, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

This series publishes short books on topics pertaining to information science and applications of technology to information discovery, production, distribution, and management. Potential topics include: data models, indexing theory and algorithms, classification, information architecture, information economics, privacy and identity, scholarly communication, bibliometrics and webometrics, personal information management, human information behavior, digital libraries, archives and preservation, cultural informatics, information retrieval evaluation, data fusion, relevance feedback, recommendation systems, question answering, natural language processing for retrieval, text summarization, multimedia retrieval, multilingual retrieval, and exploratory search.

Rhiannon Bettivia · Yi-Yun Cheng ·
Michael Robert Gryk

Documenting the Future: Navigating Provenance Metadata Standards

Springer

Rhiannon Bettivia
Boston, MA, USA

Yi-Yun Cheng
New Brunswick, NJ, USA

Michael Robert Gryk
Farmington, CT, USA

ISSN 1947-945X ISSN 1947-9468 (electronic)
Synthesis Lectures on Information Concepts, Retrieval, and Services
ISBN 978-3-031-18699-8 ISBN 978-3-031-18700-1 (eBook)
https://doi.org/10.1007/978-3-031-18700-1

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

RB would like to dedicate her contributions in this book to her family.
YYC would like to dedicate her contributions to her family and friends.
MRG would like to dedicate his contributions to the Locke family in thanks for loaning him the Echo key.

Preface

It is fitting that a book on provenance should begin by talking about its own origins. The origins of our collaboration began when we were colleagues at the University of Illinois, School of Information Sciences circa 2016–2018. At the iSchool, we worked on various aspects of provenance in our multiple roles as students, researchers, educators and scholars. Irrespective of the setting, one deceptively simple question always arose: why use PREMIS rather than PROV and vice versa?

The question is one of those questions that seems simple, while in reality it hides a multitude of flexible and changeable answers. It is like asking why someone chose an Apple computer over a Windows computer. Some responses will refer to the performance of the architecture, the software that is included or a perceived ease-of-use. Yet in the end, the true, heartfelt answer often comes down to a simple preference: *I like Apple* or *I like Windows*. That type of answer is not satisfying when considering a metadata standard. It doesn't seem sufficient to say one *likes* PROV or one *likes* ProvONE or one *likes* PREMIS. We are driven towards a more objective, professional response. This book is the latest step in a long journey we have made in addressing questions about how one chooses to structure provenance and why.

We will perhaps date ourselves at some point in the future when we say that this project began in earnest in the before times of the COVID-19 pandemic. As it was, we organized a workshop on provenance metadata for the 15th International Digital Curation Conference, held in Dublin, Ireland, on February 17, 2020. The workshop was entitled *Navigating through the Panoply of Provenance: Metadata Standards useful for Digital Curation*, and its goal was to help educate practitioners about these three metadata standards, PROV, ProvONE and PREMIS, along with their pros and cons for various purposes. The first half of the workshop covered the three standards along with exercises on their use. The second half of the workshop was a hands-on session in which the participants were tasked with creating a provenance record for a topic. The participants were split into separate groups such that one group used PROV while another used PREMIS to tackle the same underlying data. Once complete, the groups then swapped metadata records and attempted to crosswalk between PROV and PREMIS. We had a delightful time discussing provenance standards, crosswalking metadata by hand, and of course, sipping Irish

Fig. 1 Photos of workshop activity at IDCC 2020, Dublin, Ireland

stouts. It was also, unbeknownst to us at the time, the last of our in-person professional engagements for a couple of years (Fig. 1).

The workshop was a success, and much was learned by the participants as well as by us. The biggest revelation from the participants was a renewed sense of confidence in tackling provenance issues at their home institutions. Documenting provenance in a digital world means working in technologically complex environments, even when the objects in question are analog. Participants stated that, after the workshop, they were ready to try to implement standardized approaches or to mediate more fruitful conversations between content specialists and IT infrastructure personnel.

Additional workshops were included in the programs for the 83rd ASIS&T Annual Meeting and at iConference 2021. Our second workshop happened in March 2021, when working-from-home had become the new norm. Despite the lack of face-to-face connections, we could still interact in real time on our colorful online whiteboard, made specifically for the online conference experience. The virtual environment crafted inside a Miro Board enabled us to use the video game *Animal Crossing: New Horizons* to help

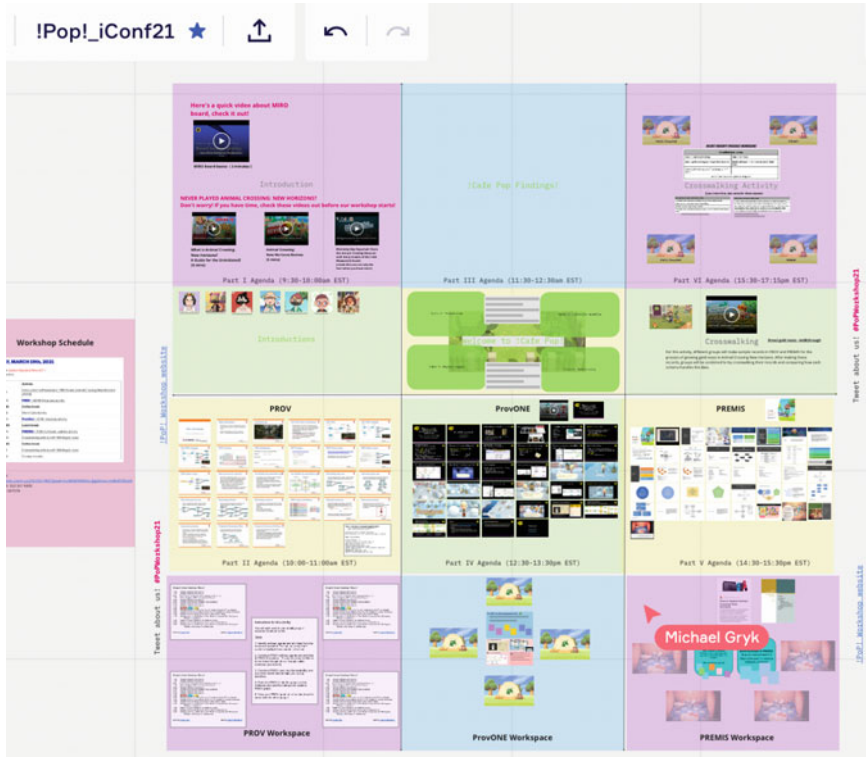

Fig. 2 Photo of workshop activity at iConference 2020, online

explore provenance and *¡PoP!*, a panoply of provenance metadata standards, with some much needed cuteness and cheer during dark times (Fig. 2).

This book is the latest step on our journey, one where we hope to share what we've learned along the way with a larger audience. The Oxford English Dictionary defines the word "provenance" as origin, source, ownership of an artwork or guidance to determine authenticity. Provenance, as we know it today, is not limited to history domains. It has many faces in different fields: phylogeny concerns the tree of life of species; genealogy studies the ancestry of families; stratigraphy dates layers of sediments. The idea of provenance transcends disciplines, and this book *Documenting the Future: Navigating Provenance Metadata Standards* is intended for anyone in any field who has a keen spirit to dabble in the world of provenance. Perhaps echoing life in general, provenance is more about the journey than the destination. We don't claim to have an authoritative answer to that deceptively simple question of why we choose one provenance standard over another. Our hope is that the chapters in this book will help empower the reader to frame and answer provenance questions on their own.

We did not make this journey alone and wish to thank all of the people who have helped along the way. First and foremost, we would like to thank our workshop participants for their curiosity, energy, drive and feedback in working with these standards. We would also like to thank the University of Illinois, School of Information Sciences, and the Center for Informatics Research in Science and Scholarship for all of their support. We would like to thank Dr. Jerome McDonough for his comments on modeling provenance in PREMIS. We would also like to thank Dr. Bertram Ludäscher and Dr. Tim McPhillips for sharing their expertise regarding provenance in general and this book in particular.

Last and certainly not the least, we thank you, the reader, for joining us on our journey. Enjoy.

Massachusetts, USA — Rhiannon Bettivia
Illinois, USA — Yi-Yun Cheng
Connecticut, USA — Michael Robert Gryk
December 2022

Contents