Computational Access to Digital Material

Exploring topics around engagement, ethics and resources

Leontien Talboom
Cambridge University Libraries
United Kingdom
lkt39@cam.ac.uk
0000-0001-7408-5471

Jenny Mitcham
Digital Preservation Coalition
United Kingdom
jenny.mitcham@dpconline.org
0000-0003-2884-542X

Abstract – Computational access is a newly emerging approach to access within the digital preservation community. This type of access is becoming more widely discussed, but a lot of uncertainties are present around the term. A guide is being created as part of a collaborative piece of work between the Digital Preservation Coalition (DPC) and Leontien Talboom, Software Sustainability Institute (SSI) fellow 2021 to help digital preservation practitioners to understand the topic and take some practical steps towards implementation. The creation of this guide is leading to much wider discussions on this topic, many of which will not find their way into the guide in much detail. This panel session will therefore provide an opportunity and platform to discuss these issues in greater depth with a range of practitioners with practical experience in this area.

Keywords – computational access, collections as data, access, computational methods
Conference Topics – Innovation

I. Introduction

Within the digital preservation community, the term computational access is used with increasing frequency. Many practitioners are aware that this type of access may be beneficial - both for their own use of the digital material and that of their users. What is involved in establishing computational access to digital materials is sometimes harder to establish, and practitioners can be left unsure about which steps they should take to begin to explore these technologies. In some cases there is also a lack of understanding of what the term 'computational access' actually means and how it interfaces with other related concepts such as artificial intelligence, machine learning and data mining.

Computational access, and the closely related term 'data as collections'[1], offers a new way of providing access to material to be used for computational methods. To explore this and related terms in more detail, and to provide the community with a way to get started, a guide is being created by Leontien Talboom (a Software Sustainability Institute (SSI) fellow) in collaboration with the Digital Preservation Coalition (DPC) and a broad network of experts drawn from the community. This guide aims to demystify computational access and make it a more approachable topic for digital preservation practitioners.

The guide will be covering several themes. It will start by introducing and defining computational access and related terms, such as artificial intelligence and data mining. It will discuss the benefits and drawbacks of using computational methods to provide access to digital materials and will provide some useful first steps for practitioners who want to get started. A selection of helpful case studies will also be shared to demonstrate what is possible.

While in the process of creating this guide, it has become apparent that a huge amount of knowledge and experience has already built up within the community. Recognising that this knowledge can not be encapsulated in a short introductory guide, this panel has been brought together to enable further discussion on the topic of computational access and how it can be used.

II. Proposed Topics

The panel session will start with a short introductory presentation providing some background to the topic and this session. A discussion will then be facilitated on a range of topics related to computational access and there will also be time for audience questions. The topics discussed may include:

iPres 2022: The 18th International Conference on Digital Preservation, Glasgow, Scotland.

DOI: 10.1145/nnnnnnn.nnnnnnn

Ethics around using computational tools - This is a key topic and one that can be approached from a number of different angles and perspectives - both from the user's and the digital preservation practitioner's point of view. Many of the processes associated with computational tools used to facilitate access are 'black boxes', therefore raising ethical concerns around the underlying processes of these tools.

Resources and infrastructure - Computational access is a novel way of opening up collections and could offer many interesting future uses of collections, but how should organizations with limited resources and experience manage this? What is a potential way to get started? What (digital) skills and expertise are needed to support it?

Communicating with stakeholders - Establishing computational access to digital materials involves collaboration with a range of stakeholders from different disciplines and with different levels of knowledge and interest. There is a balance that needs to be found when communicating with different stakeholders (for example vendors and research communities) and challenges that arise when communicating with these groups.

Engaging an audience - There is little benefit in facilitating computational access to digital materials if no-one uses them. The panel will discuss how to engage an audience and ensure they are providing tools, services and resources that meet users' needs.

III. Invited Panelists

A range of panelists have been invited to participate in this session ensuring a range of expertise and backgrounds are represented. Below a short description can be found outlining each panelist and their work. Jenny Mitcham will be facilitating this panel session.

James Baker is Director of Digital Humanities at the University of Southampton. He works at the intersection of history, cultural heritage, and digital technologies, and is currently researching histories of knowledge organisation in twentieth century Britain. James is a Software Sustainability Institute Fellow, a Fellow of the Royal Historical Society, a member of the Arts and Humanities Research Council Peer Review College, a convenor of the Institute of Historical Research Digital History seminar, and a Trustee of the Programming Historian. Prior to joining Southampton, James held positions of Senior Lecturer in Digital History and Archives at the University of Sussex and Director of the Sussex Humanities Lab, Digital Curator at the British Library, and Postdoctoral Fellow with the Paul Mellon Centre for Studies in British Art.

Dr. Sonia Ranade is Head of Digital Archiving at The National Archives (UK), with responsibility for digital services to records creators in government (for records selection and transfer), preservation of the digital Public Record and access to born digital records. Sonia's research interests include probabilistic approaches to archival description, digital preservation risk and developing new access routes for digital archives (both for 'readers' and for computational re-use). Sonia holds a PhD in Information Science.

Leontien Talboom is a collaborative PhD student at The National Archives (UK) and University College London. Her research focuses on the constraints faced by digital preservation practitioners when making born-digital material accessible. She currently also works as a web archivist on the Archives of Tomorrow project and a technical analyst at Cambridge University Libraries.

The panel session will run as a face-to-face session.

REFERENCES

[1] T. Padilla, L. Allen, H. Frost, S. Potvin, E. Russey Roke, and S. Varner, "Always Already Computational: Collections as Data," Final Report, 2018.