Plying Diverse Data: Flexible Digital Infrastructures, Metadata Narratives, and Visual Life Oral Histories Archives

Vivian L. Wong, Ph.D.
University of California, Los Angeles (UCLA)
RDA/US Data Share Fellow, Archivist, & Filmmaker
vivwongis@gmail.com | www.vivianlwong.com

INTRODUCTION

For creating communal spaces to connect research in ways that encourage scholarly collaborations and discoveries my research examines how digital platforms and metadata are (ap)plied to describe and analyze the unstructured data of visual life oral histories. It employs the Platform for Experimental and Collaborative Ethnography's (PECE) flexible digital infrastructure and analytical tools to produce an oral history archive that is searchable across diverse visual media collections and within individual media (re)sources. I also discuss the challenges of archiving visual life oral histories and consider future research directions for my work.

TRANSFORMING ORAL HISTORY DATA TO METADATA NARRATIVES

Oral histories by definition and design have narrative structures built into their interviews; they document and collect the histories of people in their everyday lives. Creating structured analytics that reveal and examine the stories people tell about themselves, composed around particular themes, key life events, and relevant topics, allow annotations to exceed labeling and identification. Annotations of oral history interviews with analytical question sets create metadata that are narratives structures. These metadata narratives reveal the patterns of life and make meaning of human experiences; and are searchable in an individual's interview and across oral histories collections, adding extra context to different types of metadata.

JAPANESE SCIENTISTS ORAL HISTORY PROJECT [JSOHP] ARCHIVE

Artifact | Record	Structured Analytics| Metadata Narratives
COLLECTION OH-1	TYPES OF QUESTION SETS
• 24 unstructured oral history interviews with astronomers and physicists in the U.S., Europe, and Japan [most interviews – 1:30 to 2:30 hrs. TRT and longest interviews – 4 to 5 hrs. TRT]	**EVENT-BASED:** Question sets that identify the key events of person's life and career.
	TOPIC-BASED: Examples are education, parents, role models & mentors, teaching, laboratory research, leadership, work environment etc.
Amount of Data • 45 hrs. of video footage= approx. 1 terabyte hard drive	**THEMATIC:** Explores the themes in one's lifestory. E.g. family, community, discovery (scientific, personal), adversities – losses & setbacks, success & satisfaction, belonging & solidarity, serendipity etc.
Format: Quicktime MOV	

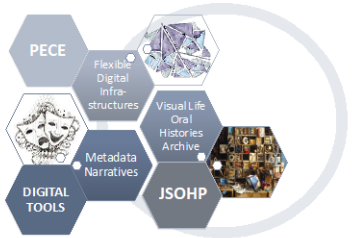
JSOHP Archive Design Logics: complex personhood, crafting selves, situated knowledges, incommensurate ontologies, epistemological differences, and divergent cultural aesthetics

CONCLUSION: MORE METADATA RESEARCH, APIs, PIs, & PECE

PECE Data Types are digital tools that create robust metadata. They organize queries into directed structure analytics that interrogate, interpret, and annotate data – or artifacts in PECE's vernacular. The JSOHP Archive uses these tools deliberately to provide intellectual access to its oral histories down to the conversations of interviews. More research is warranted to explore PECE's structured analytics as metadata: the processes of creating metadata that are metadata themselves. How can PECE address its interoperability and improve its scalability? E.g., Developing an API (Application Program Interface) is an important issue to discuss. It would allow PECE's analytics to run independently of the platform and encourage others to adopt the tools for their research without requiring PECE sites. In addition, PIs (Persistent Identifiers) are built into PECE's infrastructure and applied to its artifacts. A discussion to attach PIs to analytics and annotations will take into account data management best practices and information retrieval concerns in praxis.

REFERENCES: Artful and Stunning Cabinets of Curiosities, Decoded, NYT Magazine 2/26/16; Brown, S. (2008). 'Delivery/Service.' In Smith, M. N., et al., *Agora.Techno.Phobia.Philia2: feminist critical inquiry, knowledge building, digital humanities*. Panel at Digital Humanities 2008. Oulu Finland ; Brown, S., Clement, T., Mandell, L., Ruecheoven, D., Wernimont, J. (2016). Creating Feminist Infrastructure in the Digital Humanities. In *Digital Humanities 2016: Conference Abstracts*. Jagiellonian University & Pedagogical University, Kraków, pp. 47-50; Drucker, Johanna. "Materiality and Theoretical Approaches to Interface," *DHQ: Digital Humanities Quarterly*., 7:1, 2013; Gordon, Avery. *Ghostly Matters: Haunting and the Sociological Imagination*. Minneapolis, University of Minnesota Press, 2011; Greene, Mark A and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *The American Archivist* 68, no. 2 (2005): 208–63; Haraway, Donna. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14, no. 3 (1988):575; Kondo, Dorinne K. *Crafting Selves: Power, Gender, and Discourses of Identities in a Japanese Workplace*. University of Chicago Press, 2009; Liu, Alan "The State of the Digital Humanities: A Report and a Critique," *Arts and Humanities in Higher Education* 11, no. 1–2 (December 1, 2011): 8–41; Liu, Alan. "Where Is Cultural Criticism in the Digital Humanities," in *Debates in Digital Humanities*, ed. Mathew K. Gold. Minneapolis, University of Minnesota Press, 2012; Manoff, Marlene. "Archive and Database as Metaphor: Theorizing the Historical Record," Portal: Libraries and the Academy 10, no. 4 (2010): 394; Mckinney, Cait. "Body, Sex, Interface: Reckoning with Images at the Lesbian Herstory Archives," *Radical History Review* 2015, no. 122 (May 1, 2015): 115–28; McPherson, Tara. "Why are Digital Humanities So White? Thinking the

PLYING DIVERSE UNSTRUCTURED DATA



DIGITAL TOOLS: PECE DATA TYPES

Artifact: An artifact is the basic unit of data on PECE. Artifacts represent data that a researcher determines to be relevant to the site, including documents, images, video interviews, sound recordings, ethnographic fieldnotes, articles etc. An *artifact* is akin to a *record* in archival terms.

Analytic: An analytic is a question designed to elicit various viewpoints about an artifact – the basic unit of data on PECE. Questions can be categorized and grouped into *Structured Analytics*, which are sets of questions. These sets of questions can be conceived to generate the metadata about an artifact.

Annotation: Annotating an artifact involves responding to a set of shared and evolving questions (i.e., Analytic/Structured Analytics) that are designed to elicit various viewpoints about the artifact. One can conceptualize annotating an artifact as akin to archival description, describing a record in archival terms.

ENGAGING CHALLENGES ARCHIVING ORAL HISTORIES

Technology, equipment, and labor costs increase when working with video and sound recordings, especially oral history interviews that average 2 to 3 hrs. The JSOHP Archives' initial research phase includes twenty-four interviews with an archive of potentially 100+ interviews. The expertise required includes familiarity with archival processing and digital technologies to convert, edit, and transfer video and sound footage; it involves skills sets across different professional domains. The cost of media equipment and expendables increases the research budget more than paper archival supplies.

Creating access to unstructured oral histories data takes time. Annotations to the videos are made in "real time;" and it is work that does not lend itself to machine automation. Researchers write the analytics and make annotations that produce metadata, indexes, tags, and logs of the oral histories as they watch the interviews. Such details of archival processing and description for video footage are time intensive. They emphasize the process over product, which can be prohibitive when resources are limited – and even a luxury when an archive has a large backlog of unprocessed collections.

Individuals' research idiosyncrasies derived from different intellectual traditions; theoretical, methodological, and disciplinary approaches; interpretive strategies; professional identities and affiliations; research agendas and commitments etc. Although PECE's flexible digital infrastructures are designed to facilitate multiple viewpoint and supports multiple users groups making annotations and analytics. Furthermore, more robust metadata is created when more researchers participate in these processes on the PECE sites.

The Platform for Experimental & Collaborative Ethnography (PECE) provides a place to archive and share primary data generated by empirical humanities scholars, facilitates analytic collaboration, and encourages experimentation with diverse modes of publication. It also functions as a portal to a suite of open source tools useful for humanities research. PECE's design is both theoretically inflected and ethnographically grounded: platform design has been oriented by "design logics" drawn from cultural, social, and language theories, oriented by the constantly evolving needs of a collaborative ethnographic project, *The Asthma Files (TAF)*. ***PECE Design Logics***: hauntology, explanatory pluralism, juxtapositional logic, pursing different reproductions, scruffie contours & blurred focus, transmuting ambivalences of meaning , and valuing noise.

ACKNOWLEDGEMENTS: Digital Practices in History and Ethnography RDA Interest Group • Empirical Humanities Metadata RDA Working Group • Minorities, Women, and Migrant Scientists Oral History Project • PECE Design Group • Kim Fortun • Mike Fortun • Sharon Traweek • Research Data Alliance • RDA Data Share Fellowship

REFERENCES continued: Histories of Race and Computation," in *Debates in DigitalHumanities*, ed. Matthew K. Gold. Minneapolis, University of Minnesota Press, 2012; *PECE User Documentation* handbook v. update 1 March 2018; Svensson, Patrik, 'The Landscape of Digital Humanities', *Digital Humanities*., 4:1, 2010; http://worldpece.org/; http://45.76.3.16/jsohp.

Sponsored in part through grants from the Alfred P. Sloan Foundation #G-2014-13746 and from the National Science Foundation NSF ACI #1349002