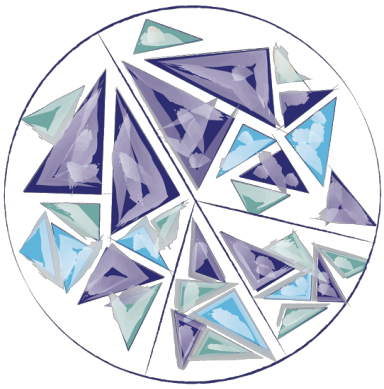

PECE

Platform for Experimental and Collaborative Ethnography

Use Cases:
Collaborative Hermeneutics and Data Sharing

Introduction

The Platform for Experimental, Collaborative Ethnography (PECE) is an Free and Open Source (Drupal-based) digital platform that supports multi-sited, cross-scale ethnographic and historical research. The platform supports collaborative ethnographic projects that address global challenges such as environmental public health and disaster prevention, preparedness and recovery. PECE provides a space for geographically dispersed researchers in empirical humanities to share primary materials (such as field notes, grey matter, photographs, and recorded interviews), and engage in collaborative and distributed hermeneutics. The PECE design team has established several analytic tools (such as structured annotation tools) for interrogating primary materials and several visualization tools (such as timelines and photo essays) for juxtaposing and drawing new meaning from diverse empirical data.

The significance of PECE for the digital humanities is three-fold:

1. we delineate the workflows through which found ethnographic material becomes “data,” providing functionality for attaching rich metadata and provenance;
2. we delineate workflows and provide functionality (through “light structures”) to support collaborative hermeneutics;
3. we provide functionality for experimentation with digitally enriched genres for expressing humanities knowledge, integrating new forms of data visualization and access to primary data.

At all of these stages, PECE provides best practice for data management per recommendations of the Research Data Alliance and the community surrounding the Open

Knowledge Foundation, customizing general orientations to accommodate special privacy requirements of empirical humanities researchers (per IRB's requirements, professional codes of ethics, etc.)[1], while simultaneously maximizing the potential for open data sharing.

PECE has been built and is governed by an interdisciplinary design group centered at Rensselaer Polytechnic Institute. PECE will be available for download and use by other research groups by December 2015. The idea for PECE grew organically from work on an experimental ethnographic platform called *The Asthma Files* - a platform that provides a digital space for geographically dispersed ethnographers to collaborate around the global asthma epidemic.

Although designed to support the particular needs of experimental ethnography projects, PECE provides a general model for the digital humanities, and particularly the empirical digital humanities (including work in anthropology, folklore, history, and related fields) that generate, collect, and analyze primary data (much of which is qualitative, textual,
and attached to found objects such as photographs or drawings produced by the people we study with), using hermeneutic techniques. PECE's "design logics" translate critical theoretical commitments (to perspectival multiplicity, for example) into digital terms.

Design Logics

PECE has been built according to a series of design logics that distinguish it from other digital infrastructures built for the natural sciences, for the social sciences, and even for the humanities. PECE is designed to support a specific kind of ethnographic project, situated within a scholarly genealogy in cultural anthropology that is between the humanities and the social sciences, akin to both but identical to neither. PECE draws concepts, methods, and inspiration more from particular lineages of philosophy, linguistics, literary theory, and critical cultural and feminist theory than it does from mainstream social science and social theory, with its emphasis on discovering or validating generalizable, large-scale social laws and phenomena. Although PECE is designed to support collaborative projects closer in character to the first terms in such classic oppositions as qualitative vs. quantitative, interpretive vs. empirical, or creative vs. data-driven, it ultimately seeks to straddle and upset such neat divisions.

Many of PECE's design logics (listed in Appendix 1) are implemented as what we call "light structure." Briefly, light structures reflect a key insight of "poststructuralist" philosophy and social theory, scholarship which is often misunderstood and mislabeled as "anti-foundational"

[1] It is widely acknowledged in anthropology that IRB requirements are far from adequate for dealing with the privacy considerations of ethnographers working in diverse situations often riven by violence, social stratification, and other contextual dynamics.

or even “anti-scientific.” A structure—a database, a language, a philosophy, a science, a social order—must be a light structure if it is to be productive and, importantly, open to growth and change. Each “light structure” design element in PECE, in this sense, marks a particular limit of a “knowledge representation” and the cyberinfrastructure on which it depends—marks the limit so that its structuring (i.e. both productive and limiting) effects are visible, and also open to question, and possibly transformation.

Structured Analytics

Analytic techniques in the empirical humanities differ from those in social science fields that may collect similar data, and are more akin to those used in literary and philosophical research, relying primarily on hermeneutics (interpretation for explanation and evocation rather than representative or statistical sampling for identification and validation). The goal is not to develop a concise and consistent view of an object, but to produce and explore multiple views of an object, leveraging “epistemological pluralism” (Keller 2002; Turkle and Papert 1990). Indeed, providing multiple, different interpretations and ways of representing particular phenomena (the sociocultural causes and impacts of the Fukushima nuclear disaster, or the impact of genetics research on understandings of environmental health, for example) is the key task for humanities researchers. Computational advances that support open-ended, underdetermined engagement with digital content that enables (even encourages) drift and transmutation in the way content is identified and taken up in analysis, are thus required.

PECE’s structured analytics were thus designed to facilitate multiple “readings” of empirical material. Structured analytics are shared sets of evolving questions designed to support a researcher in critically engaging with an artifact (see Appendix 2 for example of a structured analytic). By responding to these analytics, researchers collaboratively and analytically *annotate* an artifact; the annotation then becomes attached to its corresponding artifact and thus travels with it as it appears in different contexts.

Many annotations, especially early in the work of a project, are of relevant published material, but any object can be annotated: an image, an interview or interview excerpt, a *Nature* article. The annotations are structured to ease sharing and comparing of “notes,” and to pull analysis back to a project’s “shared questions.” As a researcher writes an annotation, the entry window presents her with a series of these questions, to which all other project collaborators are also responding. A researcher may ignore some questions and write extensively on others, but each response becomes an object in the PECE, and thus available in a structured way for recombination with other annotations on other materials.

Open Data, Open Knowledge, and Privacy

PECE was designed specifically around the urgent need of promoting data sharing and collaborative research in the empirical humanities. Its development team is committed to the project of advancing open access, open data, and free/open source development in academia.

Since the inception of the project, the PECE design team has encountered challenges in finding the balance between the need for strong privacy and data protection and the goal of fostering collaboration among remote research groups, project members, and non-academic collaborators.

In respect to this general guidelines for data sharing, PECE follows the Open Knowledge Foundation's Open Data definition, observing three general guidelines for design and implementation of PECE's data management policies: 1) data must be discoverable and indexable through the web; 2) if the data is not machine-readable and distributed in an open format, it is not reusable; 3) open data must not not have legal restrictions for its usage, repurposing, and redistribution. For the purposes of data management, the PECE design team has adopted the OKF definition of "Open Knowledge" in working with the ethnographic data produced: "Open knowledge is what open data becomes when it's useful, usable and used".

Given these general principles for open data sharing, the team is currently working to specify, document, and design software solutions to promote openness while protecting research co-participants' privacy. In contrast to other domains in humanities research, PECE's data can be highly sensitive and potentially hazardous to human research subjects and research co-participants. Therefore, the team must attend to ethical and privacy issues carefully in the design and implementation phases, combining strict IRB restrictions and Open Data guidelines in data management practices. Large portions of the data PECE produces can be made available and reused by other researchers working on similar research topics. Balancing the need for openness and the importance of data protection and restriction, the team described one functional requirement for data encryption and another for (machine-readable) open data (using a flexible, international copyright license, CC-BY-SA 4.0 International) in two use cases.

Use Cases

PECE0.2. *Annotations are structured as a list of shared and extensible research questions.*

PECE0.4. *Data Sharing with Creative Commons BY-SA 4.0 International as Default License*

PECE0.9. *Sensitive data is encrypted before being stored and decrypted whenever opened for viewing, modification, or deletion*.

Appendix

1.Design logics

Alt-Ontology: This is the most abstract and elusive of our design logics, perhaps best glossed by the ancient and continuing conundrum of philosophies of Being (ontology at its most essential, transcendental, and reductionist) versus philosophies of becoming (that tend to be nominalist, particularist, and process-oriented). “Alt-Ontology” is a sign to mark how ontologies are more achievement than they are origin, and that they always contain an injection of alterity or “otherness” that renders them incapable of full closure, and always open to change. A formal, stable ontology and its controlled vocabulary and semantics are foundational to all digital platforms supporting knowledge representations and related projects across the natural sciences, social sciences, and humanities. The overall design of PECE relies, as it must, on such ontological devices but at the same time seeks in a variety of ways to keep their essential inconsistencies and strangeness visible and open to question. In the most global or general sense, PECE tries to render any “ontology,” including its own, as a constantly open and difficult question that must also simultaneously be settled and operationalized.

Language Historicity and Pragmatics: PECE was designed with a distinct understanding of language: not as a transparent, self-evident medium, but a practice that is both underdetermined and overdetermined by specific historical conditions of production and contexts of use. This understanding of language draws from different genealogies in the study of narratives and discourses, ranging from an understanding of the linguistic sign as a site that is fundamentally constituted by conflict and dispute of multiple perspectives (Mikhail Bakhtin, Valentin Voloshinov) to historical approaches of discourse as an articulatory practice (Michel Foucault, Michel Pêcheux, Dominique Maingueneau, Norman Fairclough) and, more recently, to the anthropological study of language ideology exploring what counts as a language for determined social groups (Michael Silverstein, Paul Kroskrity). PECE allows for interpreting and understanding the ways in which history animates language by describing the conditions of possibility and the linkages between different language practices. Using collaborative tagging coupled with structured annotations, PECE allows for researchers to collaborate in their efforts to describe and foreground the multiple, heterogeneous, and contentious constitution of language in the context of a particular historical formation and across different institutions, communities, and knowledge practices.

Valuing Noise: an essential and even primary goal of digital infrastructure in the sciences is to separate signal from noise. Through a series of exclusions, extractions, and condensations, Big Data is converted into Small (or at least A Lot Less Big) Science. And the scope of most cyberinfrastructure, even that for cultural heritage projects, is always limited in advance: data on global textile trade, for example, has no place in a digital platform documenting and analyzing indigenous Andean weaving. Ethnographers have a long history of fieldwork methodologies that insist they attend to more of the world than they think they are supposed to, that insist they collect more data on more people and things and their relationships than they think they need. PECE is designed to facilitate projects with ever expanding and evolving groups of collaborators, who contribute different data sets and different interpretive habits and goals. PECE allows for and encourages the continual addition of new types of data, representing new topics and domains, not previously defined as significant or pertinent. PECE is also designed to present a researcher, at various moments of the research process, with data or analyses from other researchers working in a different area. Someone researching the development of immunological theories of asthma, for example, might have their attention drawn to an interview with an atmospheric chemist who studies ozone levels in Houston. PECE leverages or augments the alterity already present in alt-ontology, in other words, always collecting more data than it knows what to do with, asking researchers to invent new analytics for this “noise” (signal’s future anterior).

Pursuing Differential Reproduction: valuing "noise" and inventing new ways to make sense of it is, in some ways, only a re-iteration of how the natural sciences proceed. Biologist, philosopher, and historian of the life sciences Hans-Joerg Rheinberger understands the growth of scientific knowledge in terms of "experimental systems:" carefully defined structures that can reliably reproduce known phenomena and knowledge, but configured in such a way, or coupled to other less well-defined systems, that they can produce novel, unexpected results to be explored and elaborated, and in turn become part of a new reproductive apparatus. PECE is designed to support this experimental quest for "differential reproduction," providing both content and structures that are both stable yet open to play and experimentation. A key example is our annotation module. Many digital humanities platforms allow for the annotation of text and other media, but almost always in an entirely unstructured way. PECE researchers can choose from a set of annotation modules that present each user with a "lightly structured" set of questions; they can respond to the question as they see fit, they can skip questions, or they can add new questions that become part of the structure that subsequent researchers can respond to. Multiple interpretations of the same artifact are collected in this way; moreover, each question and response within an annotation is made discoverable and available for later, higher-level analysis in comparison to other annotations made by other researchers. Thus, our "light structure" is a close synonym to Rheinberger's "experimental system."

Explanatory pluralism: in facilitating multiple interpretations of the same artifact, the PECE annotation module supports another design logic, the promotion of explanatory pluralism. Interpretive differences are a signature feature of most humanities research, as they are in the tradition of cultural anthropology in which PECE is situated. The expectation in the humanities is that different researchers will develop alternative understandings of the same object or event; the expectation in the natural sciences is that researchers will converge on a single, "best" perspective. In actual fact, however, such convergence either takes an inordinately long time, or is forced by reasons other than Reason. Feminist philosopher and theoretical biologist Evelyn Fox Keller has shown that the capacity to entertain and develop multiple theories for complex objects and phenomena has historically been crucial to the vital growth and development of the life sciences. Yet most digital infrastructure in the natural sciences has an implicit or explicit goal of conveying researchers as quickly as possible to a single possible answer. By design, PECE encourages the creation and assembling of multiple interpretations, hypotheses, and theories in the firm belief that such explorations of multiple possibilities is necessary for the kinds of complex conditions PECE researchers seek to understand.

Juxtapositional analytics: identification, equation, integration, and synthesis are some of the primary modes of analysis in the natural and social sciences, and in cultural anthropology as well. Drawing on the rich history of interactions between the surrealist movement in art and cultural anthropology in France, ethnographers began in the 1980s to experiment with alternative modes of analysis and presentation more oriented around difference, comparison, collage and juxtaposition. PECE's design structure seeks to leverage difference—different artifacts and data, different annotations from different researchers, different explanatory paradigms—into insight, through a variety of display mechanisms. New understandings of an event like the Fukushima disaster, for example, are generated not through its conformity with items in a Chernobyl data set, but by foregrounding through juxtaposition their differences on multiple registers, across scales. In terms used by philosopher Gilles Deleuze, PECE works through a logic of the "AND" (exemplified by Spinoza) rather than a logic of the "IS" (exemplified by Kant and Hegel).

Crossing Scales, Working Double Binds: in conversation with early cyberneticists and information theorists, the anthropologist Gregory Bateson came to deploy logical set theory as a means for understanding the contradictions, paradoxes, and double binds that inhere in any complex system of communication and meaning. Through the archiving and analysis of artefacts representing the multiple domains and scales of a complex condition like asthma or disaster—from the molecular to the global, the individual to the social, the physical to the biological to the cultural, the material to the discursive—PECE is designed to draw out the disjunctures that occur when passing from one scale or domain to another.

Transmuting ambivalences of meaning: at every level, PECE instantiates poststructuralist understandings of language. This may best be conveyed by a long quote by anthropologist Michael Fischer concerning shifts in how language, as a complex system of material signifiers, was understood over the course of the 20th century: "Behaviorist models take words and symbols to be unproblematic tokens, combined and rearranged in meaningful chains of sentences or utterances, done in turn-taking, stimulus-response sequences. Analysts can thus build up models of culture based on sets of belief statements made by actors. Symbolist models recognize that symbols are not univocal simple tokens but have fans of meanings, and that more is exchanged in any speech act than either speaker or receiver comprehends. Nonetheless, in symbolist models, symbols are still but more complex sign tokens – like overly full bouquets or pockets of fertile sediment – richly polysemic yet discrete. Indeed, the richest symbols are black holes: the entire culture is said to be condensed there. Symbolist analysts organize their models of culture around key symbols, symbol clusters, and nodes of semantic networks, somewhat like a

crystal structure. There is a reassuring sense of relative stasis or stability of the symbolic system. Structuralist, and particularly poststructuralist, models decompose symbols and metaphors into chains of metonyms or association that play out into disseminating, ramifying, transmuting dynamics, attempting to model, in the structuralist case, the semantic-symbolic parameters of variation and transformation, and, in the poststructuralist case, the transmuting ambivalences of meaning that keep texts and communication labile (unless forcibly controlled, in which case poststructuralist deconstructive sensibilities highlight the tensions and processes of alternative meanings subversive to those intended and authorized by the controls)."

Michael Fischer, Emergent Forms of Life and the Anthropological Voice (Durham, NC: Duke University Press, 2003).

2.Example structured analytic

Structured Analytic: Reading digital infrastructure

The following structured analytics were designed to critically engage with (or "read") digital humanities platforms - an ethnographic project that has helped us situate PECE in the broader digital humanities community. In order to "annotate" a digital platform, a system user would first create an artifact referencing the digital platform. Then any user on the system could annotate the artifact by responding, to whatever extent they deem appropriate, to these structured analytics through an Edit interface on the system. The responses to these questions could then be aggregated and disaggregated to visualize where thinking on the platform converged and diverged.

General

1. Who was the system built to serve, and why?
2. What functions does the system provide?
3. How are functions technically supported?
4. What other systems, platforms, or modules does the system rely on? Do other systems or platforms rely on it?

Data

1. What kinds of data do users enter?
2. How is the data stored (backend technologies, data formats)?
3. Where is data stored on the platform? In what structure is it stored? Describe the structure.
4. How is data organized?
5. How does data move through the system? What functions allow for data discovery?
6. How does the app translate user data into information?
7. What kinds of environmental, medical etc information does the app pull in and provide users?
8. How does environmental, medical, etc. information contextualize/frameuser data?
9. Are there public or private interfaces to collect data? (if so, what are the terms of use?)
10. Are the data available to the public? If so, what is the license? If not, who uses the data and for what purposes?

Use

1. How is the system actually used, and what accounts for divergence between intended and actual use?
2. What pathways are users directed to take through the system?
3. What other pathways through the system seem to have emerged?
4. What user data is collected on the system, and what opportunities are there for user feedback?
5. To what extent is the work done on the system visible or transparent? What processes appear to be hidden?
6. Are there protocols in place for managing the use of the systems, and if so, are they are following industry standards or had internal groups defining policies?
7. How do users reference their use of the system? (by paying for it, by citing it as one would an academic paper as a resource)?

Support and Sustainability

1. Who built the system, with what skills, and with what kinds of social or commercial commitments?
2. What is the business model?
3. What can we tell about the sustainability (and plans for) of this system?
4. Who responds to platform issues/breakdowns/bugs?

Ethnographic practice

1. What assumptions about ethnographic practice seem to be built in?
2. How does collaboration seem to be conceived and how is it implemented?
3. What assumptions about language and knowledge are built in?

ID:	**PECE0.2**
Title:	**Annotations are structured as a list of shared and extensible research questions.**
Description:	Main actors: User, PECE Annotation Sub-System. User is logged into the system. User selects a digital object to annotate.
Trigger:	User wants to annotate a digital object or a segment of an audio, video, or text.
Preconditions:	The user is able to log in and have permission to access the digital object. The set of research questions is already stored in the system, however new questions can be added.
Steps for Main Success Scenario:	1. The user logs in and selects the digital object or segment of an object she wants to annotate. 2. User is exposed to a list of questions to choose from and, if necessary, a new question can be added. 3. User types the annotation text. 4. User tags the annotation: pre-existent tags are suggested as she types. 5. New annotation is stored with provenance metadata in the system.

Alternate scenario:	-
Postconditions:	Annotation structures (set of questions) can be retrieved as a set or per question. Annotation questions can be retrieved by author or digital document they are linked to. Individual annotations can be retrieved by question, user, digital object and tags.
Frequency of Use:	Each time the user creates a new annotation or question.
Status:	Draft
Author:	PECE project

ID:	**PECE0.4**
Title:	**Data Sharing with Creative Commons BY-SA 4.0 International as Default License**
Description:	Main actors: User, Licensing information, PECE Terms of Service and Privacy Policy, PECE permission system. User is logged into the system. User uploads any type of data and is prompted for describing the license (which defaults to CC-BY-SA 4.0 International if no license is selected by the user).
Trigger:	The system is set up to automatically prompt the user to select a license for new content.
Preconditions:	User has read and accepted the PECE Terms of Service and the Privacy Policy in order to create her account. User is able to log-in and create content.
Steps for Main Success Scenario:	1. The user logs in and creates new content. 2. The system prompts the user with a list of possible licensing terms for her content (Copyright, Public Domain, Creative Commons 4.0 BY/SA/NC International). 3. If no license is defined, the content is published as Creative Commons

	BY-SA 4.0 International. The TOS describes all public content on the system to be licensed as CC-BY-SA 4.0 International when not specified otherwise.
Postconditions:	Every digital object stored in the system has explicit licensing information as part of its metadata.
Frequency of Use:	Each time the user creates new data or uploads existing data.
Status:	Draft
Author:	Luis Felipe R. Murillo

ID:	**PECE0.9**
Title:	**Sensitive data is encrypted before being stored and decrypted whenever opened for viewing, modification, or deletion.**
Description:	Main actors: User, Encryption Sub-System, PECE permission system. User is logged into the system. User uploads a sensitive piece of information and marks it for encryption. The file is encrypted on the client side and uploaded to the system for storage. For viewing or modifying, it is downloaded and decrypted on the client side and encrypted and stored again upon closure.
Trigger:	The system is set up to automatically encrypt and decrypt files on the client side, refusing to store non-encrypted data marked as sensitive and private.
Preconditions:	The user is able to log in. User has a copy of her public key registered in her PECE profile.
Steps for Main Success Scenario:	1. The user logs in and marks a digital object as sensitive and private. 2. The system loads the cryptographic function on the client-side for encryption. 3. The client-side script requests the

	user's private key. 4. Upon retrieval of the user's private key, the digital object is encrypted and submitted to the system for storage. 5. If the user wants to view, modify, or delete a private digital object, it will be retrieved encrypted from the server, the private key will be requested and the data will be displayed encrypted to the user.
Postconditions:	After closing the digital object, it will be encrypted on the client side and resubmitted for storage on the server. Therefore, no sensitive information will be stored online.
Frequency of Use:	Each time the user logs on and manipulates sensitive data.
Status:	Draft
Author:	Luis Felipe R. Murillo