Enhancing Evidentiary Work through the Lens of Human Centered Computing

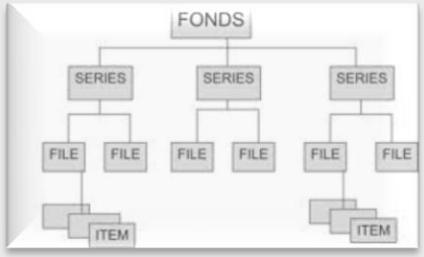
Ciaran B. Trace, Assistant Professor
Sarah Buchanan, PhD Student
School of Information, The University of Texas at Austin

SAA Research Forum • Aug. 13, 2013

Collaboration with Luis Francisco-Revilla, UT iSchool

Augmented Processing Table

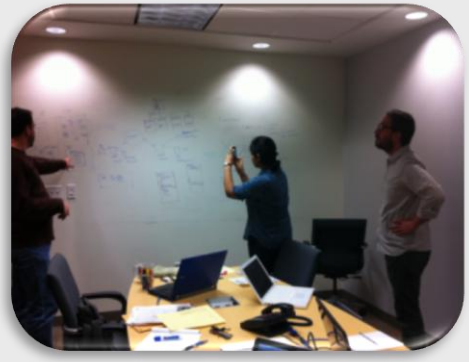
- APT is a collaborative research project bridging the fields of Archival Science and Human Computer Interaction (HCI), based at The University of Texas iSchool
- The project as a whole seeks to understand best practices and deploy surface-computing technologies towards the activity of archival processing.
  - Processing: How we prepare collections for use
- *Problem Statement*: Archival processing is costly, complex, time-consuming, and poorly understood in practice. How do archivists produce a physical and intellectual arrangement? How might technology enhance this process?



APT Web: *everest.ischool.utexas.edu/apt/*

- *Evidentiary work* involves:
  - Use of documents and associated supplemental constructs by those researchers engaged in evidence building
  - Preparation of the collection and associated supplemental constructs to support users in doing evidentiary work. Such arrangement work represents archivists' added value, and helps answer the questions: "What is it that arrangement *does* to a collection? Why arrange?"
- Arrangement creates *evidentiary layers*:
  - Archivists represent the relationships between records by constructing a hierarchical arrangement scheme
  - Places an object in a contextual relationship within the collection that reveals its value as evidence

Design and Implementation









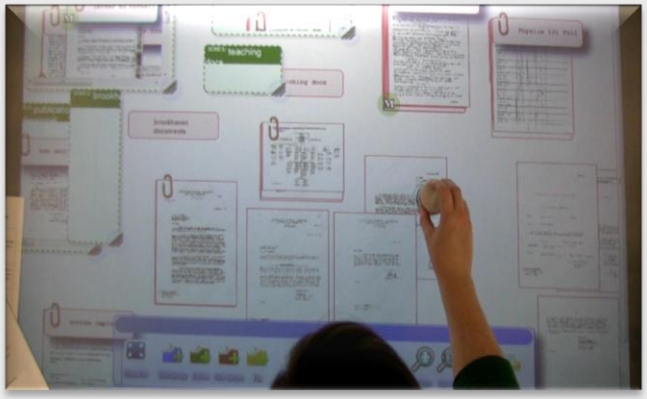
Hardware

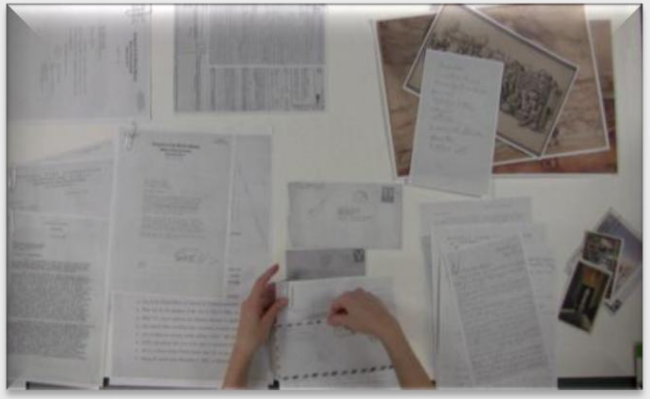
APT is a bespoke 5'x5' standing-height interactive tabletop (47" x 28" interactive area). APT uses the rear Diffused Surface Illumination (DSI) technique to provide multi-touch support.

Software

The client side of the application is written in JavaScript and HTML, utilizing the jQuery and Fabric.js frameworks, running in a touch-enabled web browser. It supports both touch and mouse input.

Processing using APT (images on the left) and the baseline (image on the right)





Augmented Processing Table

- Prototype I was completed Spring 2012 and evaluated via a pluralistic walkthrough
  - *[Jeff Crow, Luis Francisco-Revilla, April Norris, Shilpa Shukla, and Ciaran Trace, "A Unique Arrangement: Organizing Collections for Digital Archives and Libraries" in International Conference on Theory and Practice of Digital Libraries (TPDL 2012). Paphos, Cyprus. September 23-27, 2012, Lecture Notes in Computer Science (LNCS) 3232, Springer-Verlag, Berlin, Germany, pp. 335-344.]*

- Prototype II was evaluated in June/July 2012 via a lab-based comparative usability study involving 16 participants

APT Research Methods

- Study Design (APT 2.0):
  - Balanced experimental design with 8 conditions:
    - 2 systems: Baseline/Paper, APT/Digital
    - 2 collections: Estelle Ishigo, Samuel Goudsmit
    - 2 sessions: each 2-3 hours (+exit interview)

Group	Task 1		Task 2	
A	Paper	Ishigo	Digital	Goudsmit
B	Digital	Goudsmit	Paper	Ishigo
C	Digital	Ishigo	Paper	Goudsmit
D	Paper	Goudsmit	Digital	Ishigo

Findings

	QUALITY		TOPOLOGY		PROCESS		
	Errors	Scores	Description	Average	Phases	Styles	Time
PAPER	✔	✔	✔	✔	✔	✔	✔
DIGITAL	✔	✔	✔	✔	✔	✔	✔
CORRELA-TIONS	✔	✔	✔	✔	✔	✔	✔

Also have demographic/survey data and interview data

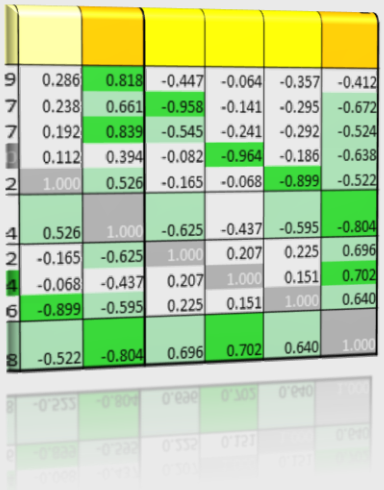
APT Results : a selection

- New metrics to understand arrangement: **Process**
  - Activity Phases
    - We identified five phases of activity completed during processing, from an initial collection review to a final presentation.
      - ***Collection Review [CR]***, ***General Document Examination [GDE]***, ***Iterative document examination [IDE]***, ***Organization [ORG]***, *and* ***Presentation [PRES]***
  - Processing Styles
    - The phase appearances clustered into five processing styles, each a “composite archivist” who completes some mixture of the above activity phases!

APT Results : a selection

- New metrics to understand arrangement: **Process**

Workflow <-> Time Correlation



more	**CR**	<-->	less	ORG
more	**CR**	<-->	less	Total Time
more	GDE	<-->	more	Total Time
more	GDE/ORG	<-->	more	Total Time
more	ORG	<-->	more	Total Time
more	Bugs	<-->	more	GDE
more	Bugs	<-->	more	ORG

APT Results : a selection

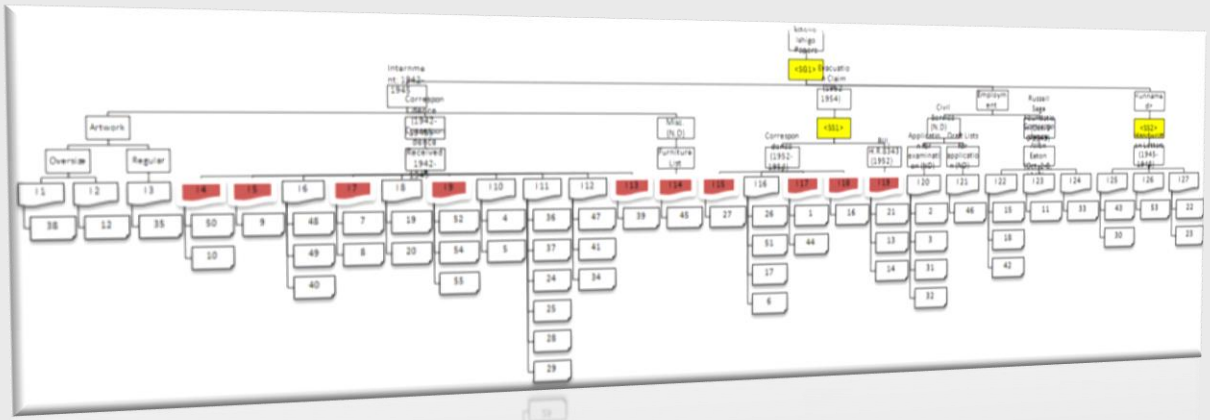
– **Quality**: Evaluated according to an arrangement's internal errors and its score. There are two types of matching errors: splits and merges.

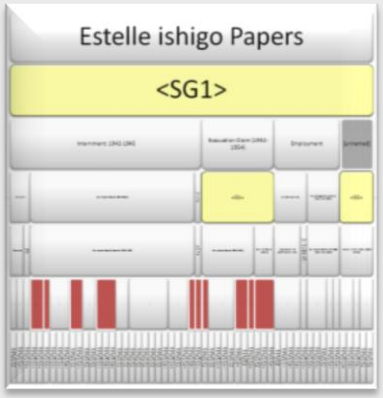
APT Results : a selection

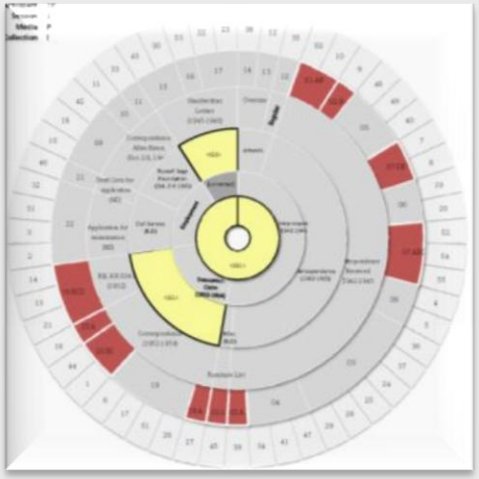
- New metrics to understand arrangement: **Quality**
  - We developed new metrics with which to **score** the quality of an arrangement:
  - Three of these metrics resulted in a ranking:
    - ***Narrativity*** (whether the arrangement accurately reflected the life of the creator)
    - ***Record-keeping*** (whether the arrangement reflected the record keeping practices of the creator)
    - ***Materiality*** (whether the arrangement expressed the types of material contained in the collection)
  - Two of these metrics resulted in an absolute value or score:
    - ***Perspective*** (whether the arrangement explicitly reveals the meta-level context of the materials)
    - ***Coherence*** (whether the arrangement follows the archival rules and established a coherent structure).

APT Results : a selection

- New metrics to understand arrangement: **Topology**







Statistical Significant Differences for APT:

- More real and total sub-groups ($p<0.02$)
- More phantom series ($p<0.02$)
- Less phantom sub-series ($p<0.04$)

APT Findings: Interview Data

- Preliminary themes identified in the course of arranging:
  - Strategies for completing an arrangement:
    - P5: “I look at the biography and the timeline, and I try to figure out what was important to that particular, in this case, person.”
  - *Meaning* of different activities (pile, clip, label, match):
    - P1: “I create piles according to, I guess type of document.”
  - Distorted principles?: Evidentiary layer or ‘subsequent user’
    - P2: "If there’s almost no original order … I’m thinking more along the lines of the researcher. If there’s some kind of original order, more the creator, because they put them in that order for a reason."
  - Better work through APT:
    - P16: "I could see dates. And like those, the three that I kind of got lazy about at the end, I could put those together correctly more quickly.”

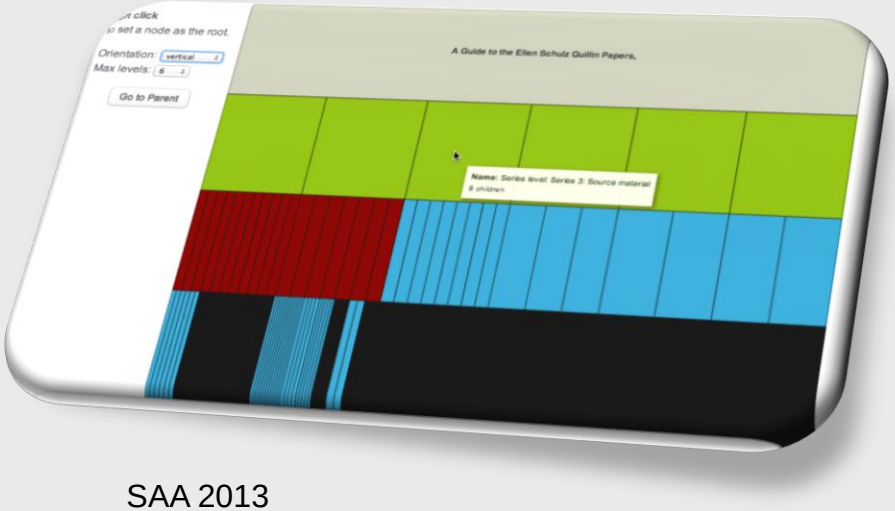
Conclusion: *"to err is human"*

- Opening the black box of arrangement:
  - A human activity performed imperfectly
  - Similar activity phases yet dissimilar outcomes
  - Ciaran's "subgroup crisis"
- Are archival principles observably put into practice?
  - High time to teach arrangement differently?
- Build technology to support these principles
  - Scaling up APT:
    - Very large collections
    - Regional repository level [VADA]
- **Ref.**: Crow, Francisco-Revilla, Norris, Shukla, Trace, "A Unique Arrangement: Organizing Collections for Digital Libraries, Archives, and Repositories," in *2nd Conference on Theory and Practice of Digital Libraries* (2012), pp. 335-344.

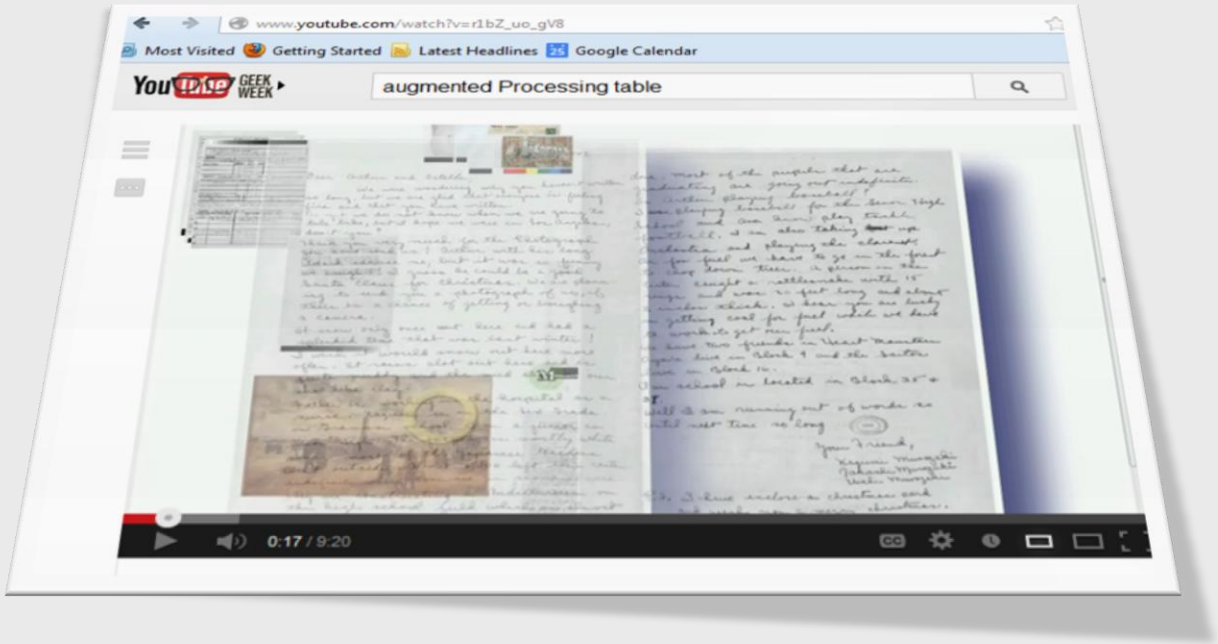
APT: Further Research & Next Steps

- Visualizing Archival Data (VADA)
- Capitalizing on an understanding of archival arrangements as hierarchical visualizations
  - At the level of a collection
  - At the level of an archival institution
  - At the regional consortia level through EAD conversions

A collection:



APT: Demo



Search "Augmented Processing Table" on YouTube