

# White Paper Report

Report ID: 99121

Application Number: HD5104110

Project Director: Niek Veldhuis (veldhuis@berkeley.edu)

Institution: University of California, Berkeley

Reporting Period: 3/1/2010-2/28/2013

Report Due: 5/31/2013

Date Submitted: 6/10/2013

**Final Performance Report for NEH Digital Start-up grant [HD-51041-10](#):**

**Berkeley Prosopography Services: Building Research Communities and Restoring Ancient Communities through Digital Tools**

**Niek Veldhuis, Principal Investigator; Laurie Pearce, Project Manager; Patrick Schmitz, Technical Lead**

**Report submitted: May 20, 2013**

BPS supports research in diverse humanities research agendas focused on the identification and analysis of individuals and social networks in corpora of administrative and legal texts. Although individual legal and administrative documents present snapshots of activities at specific times and places — real-estate purchases, adoptions, divisions of inherited property, etc. — they afford only limited and static perspectives on the society in question. BPS provides support for prosopographical and social network research with tools that help the researcher resolve a problem central to all prosopographical research: the disambiguation of namesakes, a particularly complex task in corpora that consist of thousands or tens of thousands of documents. BPS enables the corpus specialist to apply a set of probabilistic rules that emulate the rules of thumb s/he already uses to determine the likelihood, probability or impossibility that two or more identical name instances refer to the same individual. These rules reference attributes of the individual in question: maximum length of active life, profession, father's name, or place of activity.

The second means by which BPS facilitates prosopographical research is through its representation of the data as graphs that render family trees and social networks. These graphs provide access to social dynamics by illustrating that certain key figures dominate or that one or two individuals acquired a powerful position by mediating between two social networks. The interactive capabilities of the graph significantly enhance its heuristic value by allowing experimentation, effectively allowing the user to ask “what if?”. Filters will enable users to focus on people involved in one type of transaction, to center the graph on one individual, to exclude certain individuals or groups from the graph, etc. These actions may reveal patterns that are not obvious in the full graph or that the researcher may not have suspected to inhere in the data. Users may also experiment with identity assertions, by overruling the probabilistic engine in order to evaluate the effect of such a change on the graph. For textual corpora that include several thousand texts, such experimentation is impossible with traditional research tools.

In its NEH Digital Startup Phase II grant narrative, Berkeley Prosopography Services (BPS) set forth a program to develop its innovative open-source prosopographical toolkit and demonstrate its usefulness and generalizability to a community of scholars with diverse research interests. Development focused on three tracks: Technology Development, Corpora Preparation, and Fostering Community. As we identify each of these tracks in the course of addressing the topics of this final report, we aim to indicate the degree to which all of these components are essential to and support the project’s development.

## **Project Activities**

### Technology Development:

The goal of the Technology Development section of the grant was to focus on the development and deployment of a set of services that support the application but that can also be re-used for other applications (following Service Oriented Architecture principles). At the start of the grant period, the analysis and modeling of these services was largely complete. Development of database schemas, service logic and REST services support to expose functionality remained to be implemented. The goals set forth in the grant proposal for this area of project development were ambitious and wide-reaching, and included the development of a broad set of services and associated web application functionality, covering corpus management, user management, and workspace management, disambiguation and graph building functionality, reporting and visualization support, and support for creating and sharing assertions. Nearly all of this functionality was realized, although some integration work remains to tie certain areas together. The assertion support became an area of discussion with other projects, which

contributed to a stronger and more reusable underlying model, but has also delayed some aspects of the implementation.

The only change in personnel was our invitation of Davide Semenzin, a MSc candidate at the University of Utrecht (NL), to serve as an intern and participate in the development of the SNA graph visualization in Spring-Summer 2012. Working under technical lead Patrick Schmitz, he was responsible for implementing the SNA visualizer and the family-tree generator.

#### Corpora Preparation:

The primary focus in corpus preparation was completion of the demonstrator corpus, [HBTIN](#), a group of some 500 legal texts written on clay tablets in the cuneiform script. PM Pearce and U.C. Berkeley Department of Near Eastern Studies graduate students worked throughout the grant period, preparing legacy Word files for inclusion in [Oracc](#) according to the shared standards of that collaborative cuneiform project. The recursive nature of the workflow for Oracc text preparation identified a number of BPS dependencies on Oracc which, in turn, led to the development of two aids for corpus validation in the BPS environment: a “TEI view” (created by the BPS team, and contributed to Oracc) and a “Validate TEI summary” (developed by BPS technical lead Schmitz). These simplified interfaces that enable the humanities researcher to check the accuracy and completeness of the TEI mark-up and of the BPS parsing, of the identification of name instances and filiation statements contributed not only to project functionality, but represented one instance of developing collaboration between projects, one discipline-specific, the other service oriented.

External factors that impacted the preparation of the demonstrator corpus have been resolved and that corpus preparation will be completed in Summer 2013. Workshop participants are eager to see a full implementation of BPS that would support individual and collaborative research agendas, but most have deferred Oracc corpus preparation because of the complexity and time-demands of the Oracc workflow. Based on their feedback, we have identified two areas to address in order to support further development in BPS tools: (1) identify ways to simplify preparation of cuneiform text corpora that will be affiliated with the Oracc project. We will explore WYSIWYG TEI-editors as a means of simplifying the decoration of TEI in BPS pre-processing modules.; (2) develop ways to transform the content of pre-existing databases into BPS-compliant TEI. Further simplification of corpus preparation will have a positive effect on developing the BPS user-community across a range of scholarly disciplines.

In many fields of archival studies, corpus preparation is an on-going process. Documents previously unknown to researchers turn up in museum, library, and private collections — as has already happened with the HBTIN demonstrator corpus with the identification of additional components of the corpus in the collections of the Musée des Arts (Geneva) and the Princeton Theological Seminary. The ease with which BPS integrates new data and generates new analytic visual models allows for organic and speedy integration of a new model of results and reduces the problem of publications becoming obsolete, a problem endemic to humanities research.

#### Fostering Community:

During the grant period, we have actively fostered the development of a research community among end-users of BPS: humanities scholars who work with texts and documents that include much information about individuals in particular times and places and who wish to identify those individuals and the family and social networks in which they operate. However, a second community has also developed as a focus of our communications and exploration of future collaborations: technologists involved in digital humanities and related projects who are looking for more robust and sustainable development models, and specific innovations in our approaches.



In the humanities realm, our community-building began with a workshop in March 2010 held on the UC Berkeley campus. Participants included technologists from Berkeley IST and researchers from the community of cuneiform scholars (See Appendix 6 for participants in each workshop). The purpose of the workshop was three-fold: (1) to explain the user-model for the web application under development and gather additional usability feedback; (2) to explain the community model for supporting further development and sustaining the project in the longer term; and (3) to jump-start the preparation of additional corpora, using the HBTIN demonstrator corpus as an example.

The workshop began with an introduction to the fundamental concepts of social network analysis (SNA) and the potential of open-source, generalized and extensible digital toolkits to facilitate and deepen the scope of humanities research agendas. The participants readily embraced the BPS concept as well as the notion of community support for development. They then jumped into the task of preparation of text corpora as TEI to be processed using the BPS tools. As all of the participants of this workshop were cuneiform scholars, corpus preparation focused on developing corpora according to the best practices and shared standards in place in Oracc ([oracc.org](http://oracc.org)), a consortium of digital presentations of text corpora from ancient Iraq.

The second workshop, in April 2011, strengthened community bonds both near and far. Some initial workshop participants returned and additional members of the cuneiform scholarly community joined the effort. Out of this workshop has emerged a core of five researchers who work with text corpora which are temporally and geographically proximate to each other. A major component of the workshop was "hands-on" exploration of the tools. Using the demonstrator corpus, participants explored the corpora features of the service, sorted names, roles and viewed name citations. An interesting result emerged when participants began to explore the implications of the flexible data model in BPS, which allowed them to filter on features of persons and activities in ways that had not been practical before (e.g., finding women who acted as witnesses); in essence, BPS allowed them to ask *new questions* about their corpora.

Participants gathered, including some via Skype, for a morning of assessment of the existing features and functionality and the creation of a "wish-list" for future development. A serendipitous outcome of this assessment phase was that workshop participants realized many ways that the questions at the heart of their research agendas overlap, and that the prosopographical and social network methodologies promoted through the use of BPS would support further collaboration, particularly in the broadening of research questions. The workshops and the developing BPS tools have spurred the exploration of collaborative research agendas, such as a study of the networks of business-men whose interests position them in multiple ancient cities. This second workshop demonstrated BPS's ability, and the users' receptivity, to support community engagement in the evaluation and assessment of the toolkit's functionality and suitability to the tasks.

As part of fostering community, we initiated development of a communication plan. The BPS wiki (<https://wikihub.berkeley.edu/display/istds/Berkeley+Prosopography+Services+Wiki+Home>) remains the primary source of posting project progress. Our primary audience consists of workshop participants and members of the Berkeley digital community (including the School of Information, IST (Central IT), and Computer Science departments). Although we did not make use of social media or of blogs during the grant period, one outcome of our self-review at the end of the grant period was the launch of a blog ([berkeleypros.wordpress.com](http://berkeleypros.wordpress.com)), where both the PM and technical lead have committed to contribute, both as part of the project development and documentation, as well as fostering our community.

We have taken full advantage of campus digital humanities activities to spread the word about BPS's development and to foster the development of the campus digital community. Presentations made in connection with these community building efforts are described in the section dealing with publicizing the results of the program.

Finally, technical lead Patrick Schmitz participated in several events associated to Project Bamboo (a Mellon-funded project developing IT infrastructure in support of Humanities Research). In particular, he presented two key innovations (an abstraction for configurable rules, and our model for *assertions* on a corpus) at a Tool-Builders workshop. The assertions model in particular was the subject of considerable interest, and discussions continue about sharing this among various projects.

### **Publicizing the results of the program:**

A Fall 2012 presentation and demonstration of BPS at the UC Berkeley School of Information led to on-going conversations with Berkeley faculty (Prof. Ray Larson, Dr. Clifford Lynch) about the relationship of the probabilistic reasoning model at the heart of BPS to digital projects such as SNAC (Social Networks and Archival Context Project, <http://socialarchive.iath.virginia.edu/>). Additional efforts to publicize the project have primarily focused on participation in academic and campus digital humanities initiatives: at the UC Berkeley School of Information presentation, and in an invited participation in a panel discussion and poster session in a Fall 2012 campus-wide forum entitled "What Can Digital Humanities Do For You?" sponsored by the Townsend Center for the Humanities.

In the larger digital community, Pearce represented BPS at two NEH-sponsored programs: as an invited presenter at the NEH Institute for Advanced Topics, "Networks and Network Analysis for the Humanities," (UCLA, August 2010), and as a participant in the NEH ODH-funded "Linked Ancient World Data Institute" at the Institute for the Study of the Ancient World (NYU, May/June 2011). Schmitz and Pearce were invited to present BPS at the Pacific Neighborhood Community conference in Berkeley, December 2012, thus publicizing the program to the scholarly digital community in the Pacific Rim.

The BPS team has made small initial steps into the use of social media for publicizing the project. Pearce has posted a link to BPS, the BPS blog and relevant research interests on her academia.edu page. BPS team members agreed that increased use of social media would be an effective means by which to announce on-going project development and have just launched a blog ([berkeleypros.wordpress.com](http://berkeleypros.wordpress.com)).

### **Accomplishments**

#### Technology Development:

We completed the development and deployment of the initial set of services that support corpus management, workspace management, TEI ingest and mapping to the internal model, and disambiguation of name citations, as well as administrative support for user accounts and simple profiles. The web application exposes this functionality, as well as workspace support for configuring the graph builder model (including disambiguation), and for browsing and filtering (reporting) the results. Administrative functions support registration of users, authentication, and role-based authorization to control access to different levels of functionality. The web application has been in use by project participants for over a year now.

We recruited a graduate student to assist with service development in the areas of Social Network Analysis (SNA) and graph visualization. These services were built upon existing open source functionality, adapted to our services architecture. While considerable progress was made and prototype services for this were demonstrated to project participants, some work remains to integrate these services into the public web application. We need to complete some additional query support to provide information about persons and activities as GraphML (a commonly used XML dialect for representing graph data), and integrate the SNA and visualization services with this new functionality.

One additional area that was prototyped, but needs additional attention to integrate with the public website functionality, supports family tree visualization and exploration UI. This too is planned for the current (ongoing) phase of software development.

Finally, although we have a prototype implementation of basic assertion services (allowing a researcher to assert the dates for each corpus document), work remains to fully realize the model for assertions. As we discussed this functionality with other project teams, additional issues were identified that would make our model more robust, and would facilitate reuse in other projects; a more general solution is currently under development.

Given our good experience guiding students in implementation, we plan to recruit one or more new student developers to help complete these areas of work. These will be advanced undergraduate students in computer science who will work as interns, supervised by technical lead Schmitz. They will be recruited through the UC Berkeley Undergraduate Research Assistant Program. This program allows us to make some progress with limited finances, and provides an opportunity to contribute to the training of a next generation of technologists interested in digital humanities.

#### Corpora Preparation:

We have realized our goal of preparing two cuneiform corpora for Oracc and the generation of TEI, and in so doing, we identified dependencies of BPS on the TEI input produced for Oracc-project text corpora. These led to the development of a number of “work-arounds”, and identified segments of the input that may require individualization on the part of each corpus curator. With the resolution of external dependencies on museum regulations, the HBTIN corpus will be finalized by early Summer 2013, Seri’s Rim-Anum corpus (<http://oracc.org/rimanum>) is complete and awaits importation to BPS.

#### Fostering Community:

This was a particularly successful area of the program. Not only did we draw together an international group of cuneiform scholars committed to supporting and engaging in the development of BPS, but we interested colleagues in the department of Near Eastern Studies at Berkeley who already utilize SNA in their research into early Islamic intellectual history. As additional functionality is achieved, we are confident that additional scholars from diverse fields will be attracted to BPS as a research tool.

We have also developed a presence in the technology community, both on campus and at large. The BPS architecture and probabilistic modeling that supports assertions may contribute added value to preparation of corpora broadly defined as historical as in the archival and library records that are the foundational corpora of the SNAC project, which is largely driven by the collaboration of several technology communities.

## **Audiences**

### Technology, Corpora, Community

The primary audience for BPS is the community of humanities researchers who work with text corpora that contain personal names and prosopographical information about the individuals in the corpora. A second audience includes technologists supporting digital humanities projects, who are interested in robust and sustainable development models and specific technology innovations that can be reused more broadly.

At the start of the grant period, the only project associated with BPS was Pearce’s HBTIN demonstrator corpus. The completion of a second corpus demonstrated the positive outcome of a supportive team approach to corpus preparation. The workshops brought together a dozen researchers, one-half of which have continued with corpus development; all participants continue to provide feedback on BPS developments. The public website went from just a few test users to several dozen registered users,

and many more anonymous visitors. There have been nearly one hundred views/downloads of BPS presentation slide-decks from Pearce's academia.edu page. Implicit in BPS's mission to support research in a variety of humanities disciplines is the expectation that BPS would draw in investigators from other institutions. The complete list of participants and the institutions they represent appears in Appendix 5. It is noteworthy that the participants are affiliated both with major centers of cuneiform research (Berkeley, Chicago, Leiden, SOAS) and also with institutions in which they have little or no immediate contact with colleagues in the discipline. BPS thus strengthens community bonds.

## **Evaluation**

Our project plan did not call for a formal overall evaluation. In addition to feedback from informal evaluation of workshops and from the webinar hosted as the project closed, the members of the BPS team conducted a self-review at project's end. All of this input contributed to our identification of a number of strong positive outcomes as well as areas in which we did not fully meet our goals.

### Technology:

The technology evaluations that concluded our workshops produced ideas for generalizing aspects of the model and for adding new features, such as date conversions, additional import models and multiple workspaces. These were prioritized by workshop participants (as can be seen on the issues list and we have taken their rankings into account as we proceed with our development plans. The issues list on the project wiki affords us a measure of milestones completed and a prospective on work to be done.

Part of our software development methodology included periodic review of the software developed. Issues raised in these reviews were addressed in subsequent development sprints.

### Corpora Preparation:

Our goal to have the demonstrator corpus and one additional corpus prepared and ready for importation into BPS were met. The additional texts to be processed for the HBTIN project demonstrate the flexibility of BPS, as their integration serve as proof of concept that BPS is extensible and thus facilitates on-going research.

We anticipated some, but not all, of the difficulties associated with corpus preparation and the generating of TEI through Oracc. This led us to consider additional tools to support corpus preparation in the pre-processing stage of BPS use. The reservations our workshop participants expressed toward complex TEI tagging routines made clear the need for, and receptivity to, easy-to-use tools within the BPS toolkit environment.

### Fostering Community:

In this start-up phase of the project, we defined the public primarily as the campus community (which includes members of the Department of Near Eastern Studies, the Townsend Center for the Humanities and the School of Information), the community of cuneiform scholars and scholars in other humanities disciplines. A community we hoped to develop is that of the digital humanities (technologists).  
Response

A webinar conducted at project's end produced very positive feedback. The participants, some of whom were just newly acquainted with BPS, were both complementary and pointed to additional attributes they would like to factor in for their research projects. One participant raised the issue of assigning unique IDs to particular individuals in antiquity. This raises the issue of how BPS will deal with stable URIs and Linked Open Data, as the premise of probabilistic disambiguation carries with it the notion that the application of different parameters by a single or different researchers will result in namesakes being identified as the same or not the same in separate research investigations. The application of SNA

principles and the functionality of a graph visualization was deemed a welcome addition to the study of the ancient Near East.

## **Continuation of the Project**

### Technology:

U.C. Berkeley continues to fund the core infrastructure for the public website, and for collaboration tools hosted at UCB. The ongoing cost for the site is quite modest, and can be easily sustained by the institution. The collaboration tools are provided as common good infrastructure to the campus. Other tools such as the software repository are provided by third-parties that support open source development. In addition, the primary participants continue to devote time and energy to the project, although at a somewhat reduced level between funded development phases.

On the technology side, several projects have indicated an interest in collaborating on the development of support for a generalized model of assertions, with provenance tracking and support for a publish/consume model for collaboration.

Members of the Social Networks and Archival Context project (SNAC – funded in part by grants from the Mellon Foundation, and NEH) have indicated that based upon several shared discussions and a joint workshop, they would like to continue to collaborate on ideas and tools for disambiguation in our respective domains.

### Corpora Preparation:

The addition of corpora from cuneiform scholars and from other disciplines, such as Islamic intellectual history, are expected. Each addition to our list of corpora provides opportunity to refine and further implement the BPS toolkit.

### Fostering Community:

A number of potential users have expressed interest in establishing more concrete and on-going collaborations with the BPS toolkit and team. A workshop participant now on the faculty at Rijksuniversiteit Leiden has suggested that we pursue funding for collaborative international projects in the next year and has invited PM Pearce to present the work of BPS to her research unit. Colleagues who curate corpora with characteristics similar to those of the demonstrator corpus expressed the benefits that could accrue from great collaboration on corpus preparation, normally a solitary task for which a corpus curator is responsible. A curation collective among members of this cohort of users will be developed in the coming year. The anticipated members of the collective reside across the United States and in Europe, so all meetings will need to take place via Skype or similar programs. These will be scheduled monthly, and will have as their goal shared resolution of issues and dependencies on the Oracc corpus preparation routine that impede corpus development. Pearce and Schmitz will continue to develop ties to the community of technology specialists, within and beyond the Berkeley campus as they participate in inter-departmental dialogue and in local and international digital conferences.

## **Long Term Impact**

### Technology:

The work to share ideas, approaches, and innovations, with the technical community, has raised the profile of U.C. Berkeley in the community of digital humanities developers. The collaboration between campus faculty and central IT staff in a model to produce reusable, sustainable solutions in support of research, has generated interest across a number of different contexts, and drawn interest from many peer institutions. Based in part upon discussions of the BPS project, technical lead Schmitz was invited to

speak at the University of Wisconsin, Milwaukee, in their “Digital Futures” series, presenting new models of IT support for research in the social sciences and humanities.

#### Corpora Preparation:

BPS will have a long-term impact on prosopographical and social network analysis on corpora of administrative and legal texts across multiple disciplines. Its architecture and open-source code ensure that data from any period of documentation that is formatted as TEI can be processed and analyzed. The ability to upload sequential iterations of a corpus — for example, as new documents are recovered — simplifies research and analysis, as disambiguation of namesakes and the generation of graph visualizations can be done nearly automatically. It will enable researchers to ask questions of corpora that are too large to be analyzed by hand.

#### Fostering Community:

Within the community of cuneiform scholars, BPS has increased acclaim for UC Berkeley as a leading institution for digital cuneiform scholarship. Berkeley already belongs to the constellation of major digital research endeavors in the realm of cuneiform scholarship: PI Veldhuis is one of three directors of the Oracc consortium. The project is seen as a valuable resource for analysis of the text corpora produced in one of the world’s most ancient writing systems and BPS thus confirms Berkeley as a place which supports and develops cutting-edge and innovative tools to humanities resources. The positive reception of BPS in outreach fora prompts us to continue to develop BPS and to look to other projects focused on SNA to explore collaborative development. We will undertake to initiate such investigations in the coming academic year.

## **Grant Products**

#### Technology:

The project website ([berkeleyprosopography.org](http://berkeleyprosopography.org)) was set up with two levels of access: “Reader” enables the casual visitor to the site to explore the site. The Reader can: view all corpora loaded to BPS, view and explore corpora contents, examine document details including the Name-Role-Activity instances attested in each document, link to external sites which support alternative views (simplified TEI, images and line art) and/or provide additional metadata (Oracc: [oracc.org](http://oracc.org); Cuneiform Digital Library Initiative: [cdli.ucla.edu](http://cdli.ucla.edu)), filter and sort results of loaded corpora. There are two levels of access for registered users. Users who are corpus curators have the ability to modify the contents of a corpus and define the rules to be invoked in the workspace where the probabilistic modeling occurs. They and all other registered users can set model parameters and assign weights (as a percentage) to the validity or likelihood of a disambiguation rule.

The graph visualization is an important technological product, not only because it provides an attractive and interactive interface with data, but because it affords researchers the opportunity to explore subsets of a corpus’ data that might not be apparent in traditional (hand-generated) analyses. For example, clusters of individuals (as seen in Appendix 4, Figure L) can be difficult to identify in densely populated graphs of a complete data set.

#### Corpora Preparation:

The metadata contained in the decoration of the TEI generated in preparing corpora for use in BPS may be considered a secondary grant product. Should that TEI and its associated meta-data prove to have value for future digital research environments, our open-source architecture and mission of reusability and extensibility would support its integration into those environments.

Cuneiform researchers interested in utilizing BPS for prosopographical and social network research may opt to prepare the corpora according to the best practices and shared standards of the Oracc

consortium. Thus, their corpora add directly to the digital resources available to an entire community of researchers. Thus, even in this initial phase of BPS's development and early implementation, we facilitate the development of basic research materials (here, text corpora from specific times and places).

#### Fostering Community:

While community is not a tangible deliverable, it is nonetheless an important outgrowth of this phase of BPS development. The on-going interest, both from the humanities researchers and the technology community, reflects the added value BPS contributes to research agendas.

As the grant period of our NEH Digital Start-up Grant concludes, our assessment of our progress and the feedback and support of our two communities leads us to continue developing BPS. The NEH Digital Humanities Start-up Grant provided significant support in the form of seed money that enabled us to demonstrate proof of concept in the form of functionality and community development. BPS is well-positioned for the next stage of implementation, in which we expect to build on the initial functionality and incorporate the additional ideas and integrations suggested by the growing community of users and collaborators. We will also undertake a formal evaluation of how the tool works for researchers, and how well it supports or expands scholarly workflows. We are seeking additional near-term funding for this next phase of work (including support from the NEH, in the form of a Digital Implementation Grant, for which we have applied in the 2013-2014 CFP), but are also exploring long term sustainability models in which the BPS project and its growing community are self-sustaining.

## Appendices

Appendix 1: BPS Poster: our narrative in brief text and pictures

Appendix 2: Presentation slide deck, Spring 2013

Appendix 3: BPS Architecture Diagram

Appendix 4: Highlights of the BPS User Experience and Workflow

Appendix 5: Workshop participants

Appendix 6: Cited URLs



## Appendix 1. The BPS Poster

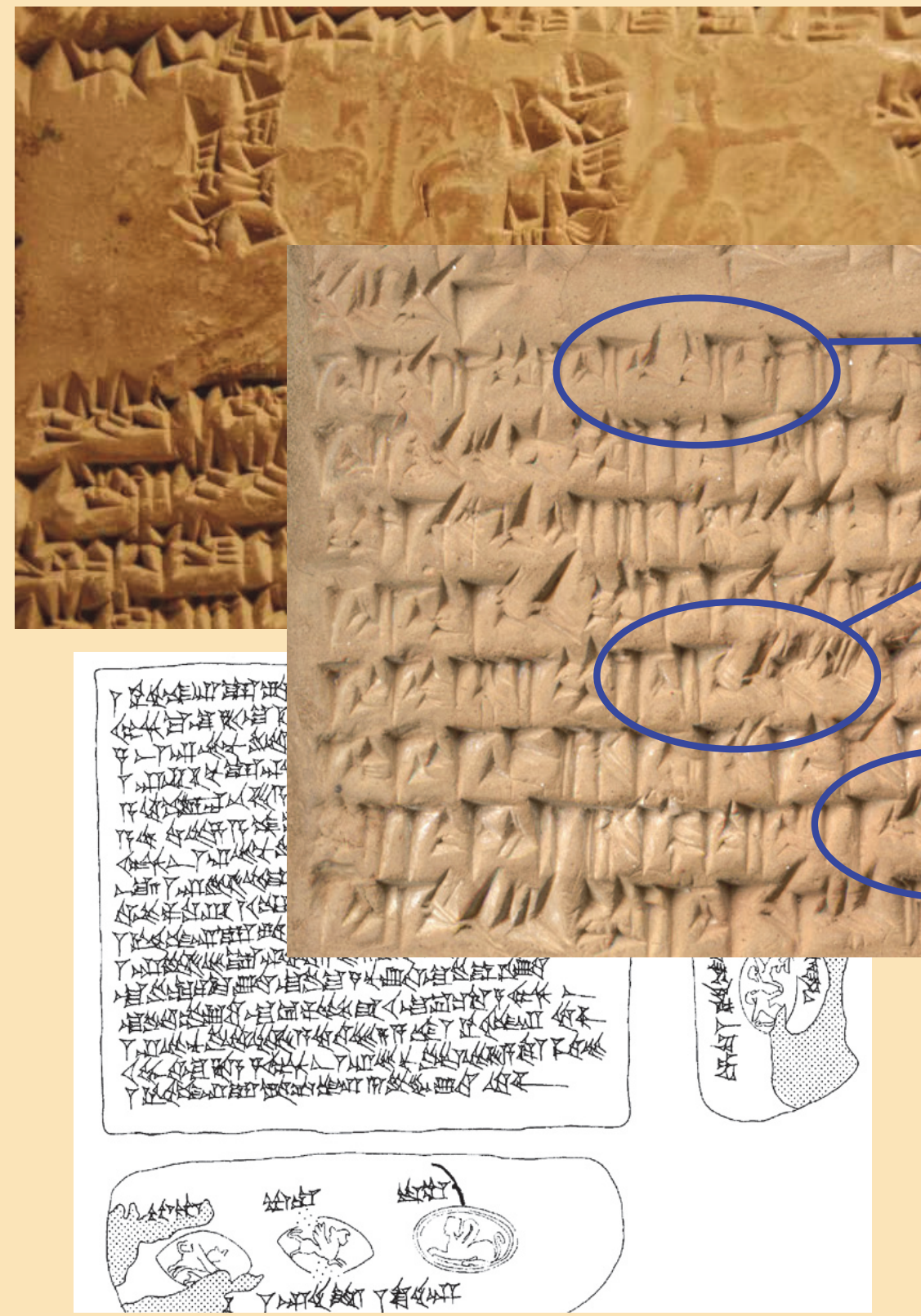
This poster presents the BPS narrative in four quadrants:

1. Context: BPS appeals to researchers who are interested in eliciting names, roles and activities from documents that record legal, economic, and administrative activities.
2. Problem: BPS provides new tools and approaches to traditional time-consuming methods of assessing individual corpora. Generalized approaches to fundamental research problems provide strategies and resources for many humanities research agendas.
3. Open-source and extensible architecture: The architecture of BPS supports a natural workflow progression, from corpus preparation to data-processing, name disambiguation and generation of graph visualizations of social networks.
4. Application: The generalizability of BPS ensures that it supports research in a variety of disciplines, grounded in texts from the ancient through the modern worlds.



# Tools for Social Network Analysis of Ancient Texts

## Context



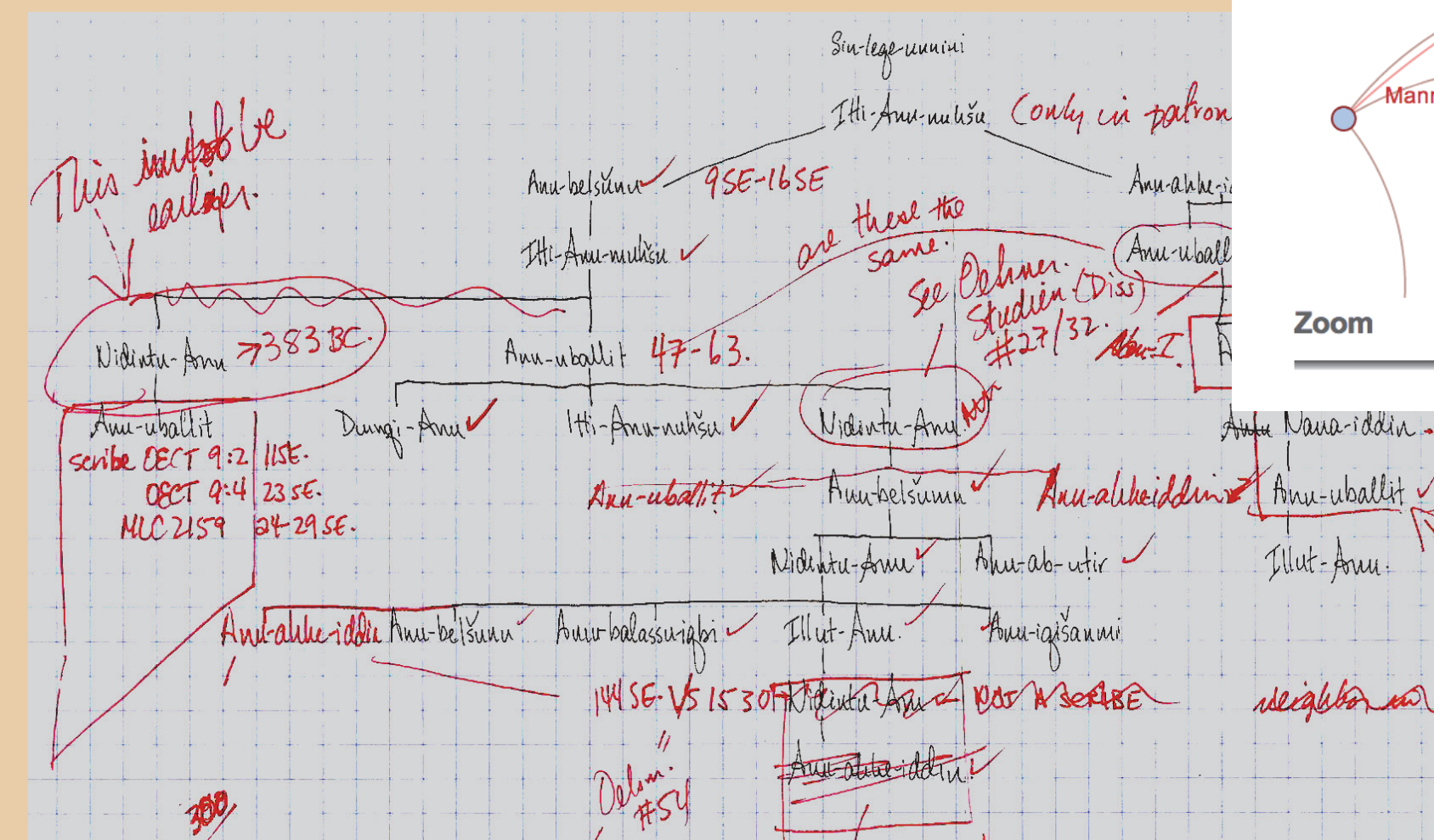
Anu-uballit (seller)  
 Labaši (buyer)  
 Kidin-Anu (witness)  
 Zoros (witness)

Hellenistic Babylonia: Texts, Images, Names  
<http://oracc.museum.upenn.edu/hbtin>

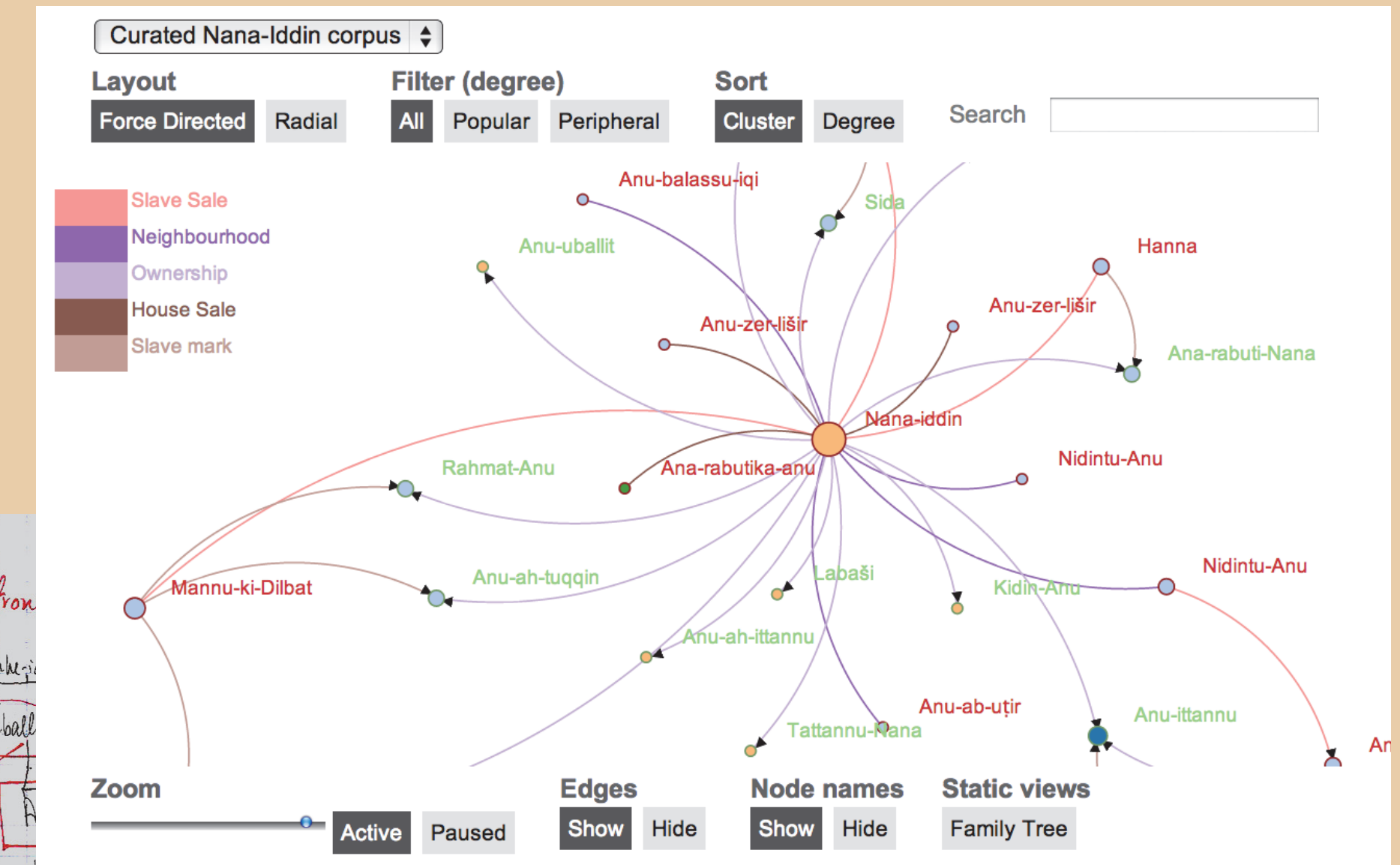
## Prosopography:

- Find family relations
- Disambiguate namesakes
- Build social networks

## How to get from hand-drawn tables and family trees...



## Problem



...to automatically generated, interactive visualizations?

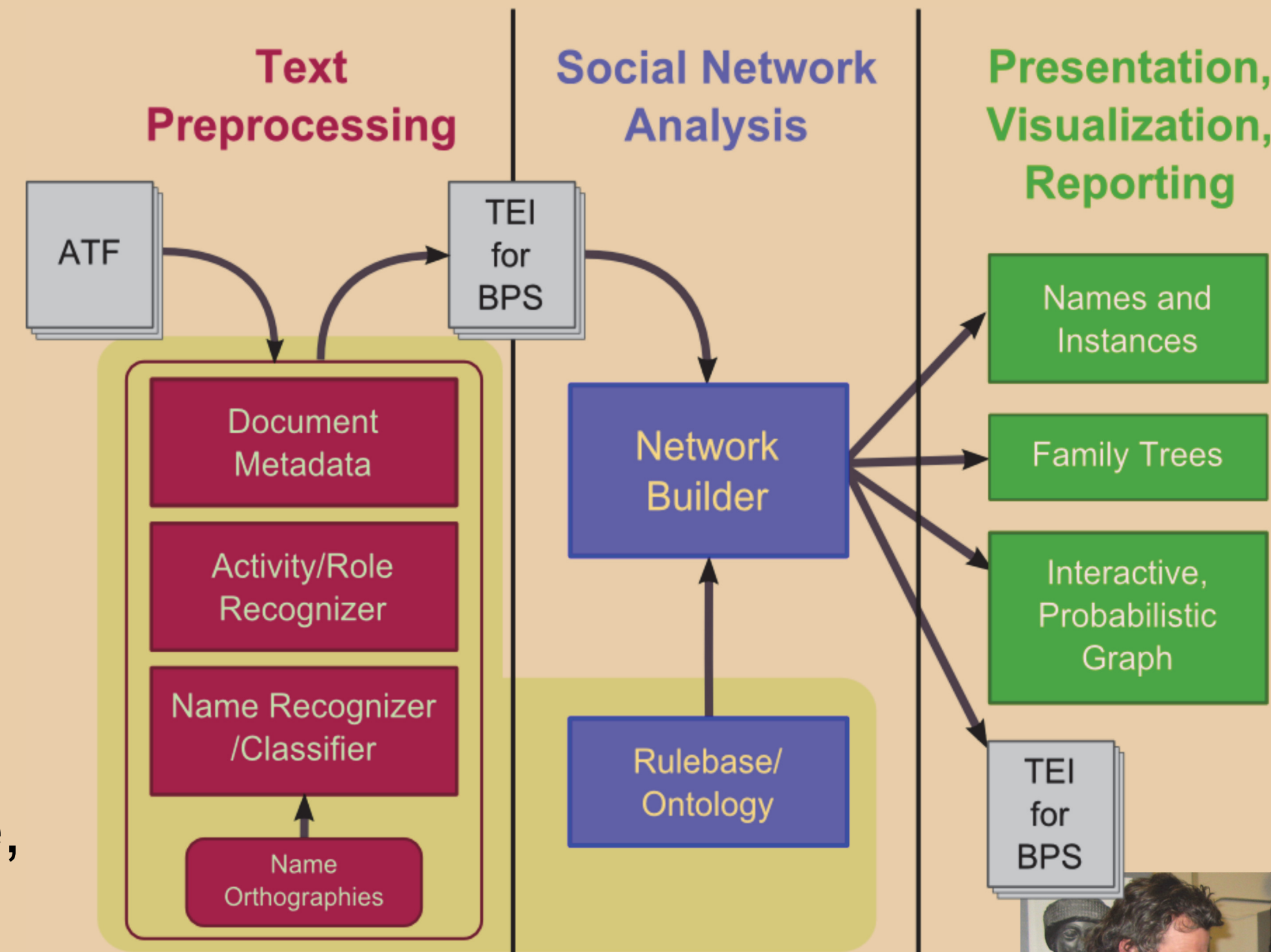
## Berkeley Prosopography Services

- Name, activity recognition
- Probabilistic disambiguation, graph construction
- Pluggable visualizations
- User feedback incorporated into recognizer, graph
- Service Oriented Architecture, abstracted for re-use

Project Website:

<http://www.berkeleyprosopography.org>

## Solution



## Apply BPS to:

- Islamic Intellectual History
- Multiple cuneiform archives
- Trade across cultures
- Captioned iconography

## Ask new questions:

- Unexpected social trends
- Cultural transmission

## Project Team:

Niek Veldhuis (veldhuis@berkeley.edu)  
 Laurie Pearce (lpearce@berkeley.edu)  
 Patrick Schmitz (pschmitz@berkeley.edu)  
 Davide Semenzin (davide.semenzin@berkeley.edu)  
 Steve Tinney (stinney@sas.upenn.edu)  
 Eleanor Robson (er264@cam.ac.uk)

## Berkeley Prosopography Services



## Extend tools:

- Multiple input formats
- Authority tracking

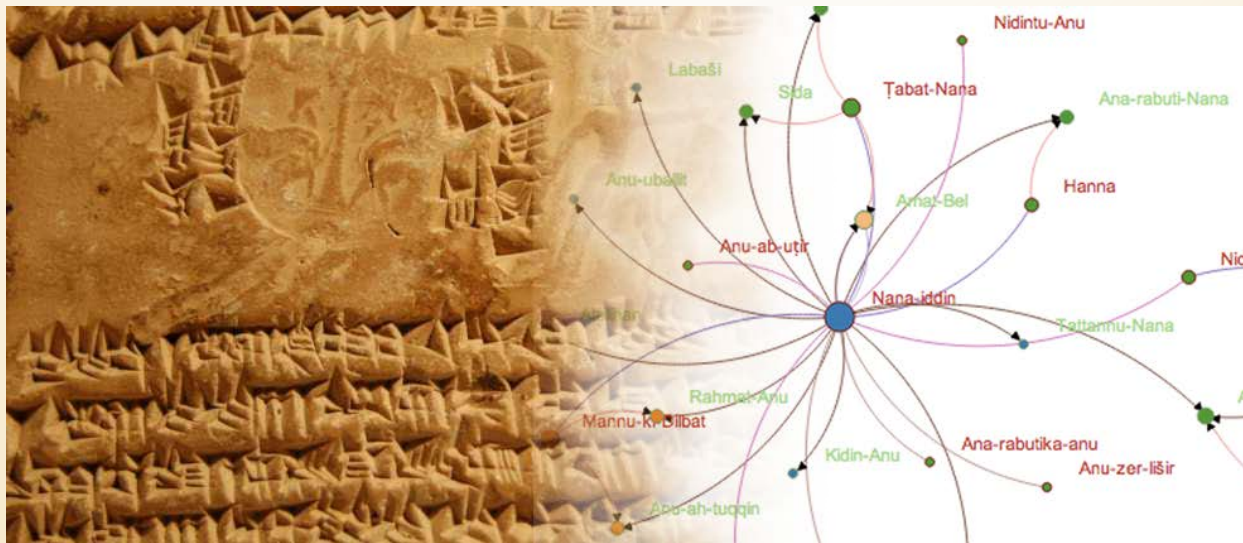
## Potential



## **Appendix 2. Presentation slide deck, Spring 2013**

This slide deck was compiled for presentation at the UC Berkeley School of Information Friday Seminar Series in March 2013. The session focused on the potential for BPS's probabilistic modeling of assertions to contribute added functionality to the disambiguation workflow of the Social Networks and Archival Context Project (SNAC). As a result of this presentation, we will pursue collaboration with Prof. Ray Larsen and Dr. Cliff Lynch in the coming academic year.

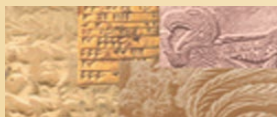
# Ancient Families, Modern Tools:



## Berkeley Prosopography Services

Dr. Laurie Pearce, UCB NES  
Prof. Niek Veldhuis, UCB NES  
Patrick Schmitz, UCB IST/RIT  
Davide Semenzin, Utrecht Uni ICS

I School Friday Seminar w/ SNAC  
UC Berkeley  
22 March 2013

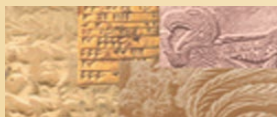


# What is Prosopography?

- Identifying people referenced in corpora: onomasticon
- Building genealogies: family lineages
- Recovering relationships: social networks

## Dependencies:

- Scope and condition of media & data
- Disambiguating namesakes
- Finding family relations
- Recognizing activities and roles
- Controlling chronological framework





# Project Content: Hellenistic Uruk

## Hellenistic Babylonia: Texts, Images and Names *University of California, Berkeley*

### Welcome

More than 3,000 cuneiform clay tablets document the intellectual, religious, scientific, legal and economic activities in Hellenistic Mesopotamia. Originating primarily from Uruk and Babylon, these texts show that although Alexander the Great and his successors transformed much of the cultural landscape of western and central Asia, they left many native practices and institutions intact. Hellenistic Babylonia: Texts, Images and Names presents to Assyriologists, Classicists, ancient historians and others the evidence necessary for study of Mesopotamia at the time when traditional culture came under the powers of the Hellenistic world.

Three primary areas of this website include up-to-date and readable publication of the materials necessary for an integrated study of Hellenistic Mesopotamia:

» **Texts**: transliterations and translations into English of texts from the major sites of Uruk and Babylon.

» **Images**: drawings and photographs of seal impressions on Hellenistic cuneiform texts.

### Pages

Home

About

Texts

Images

Names

### More Information

Bibliography



530

8-20

3

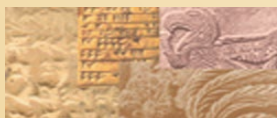
10,000

legal texts

name citations/text

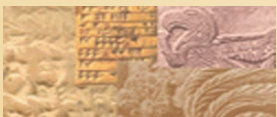
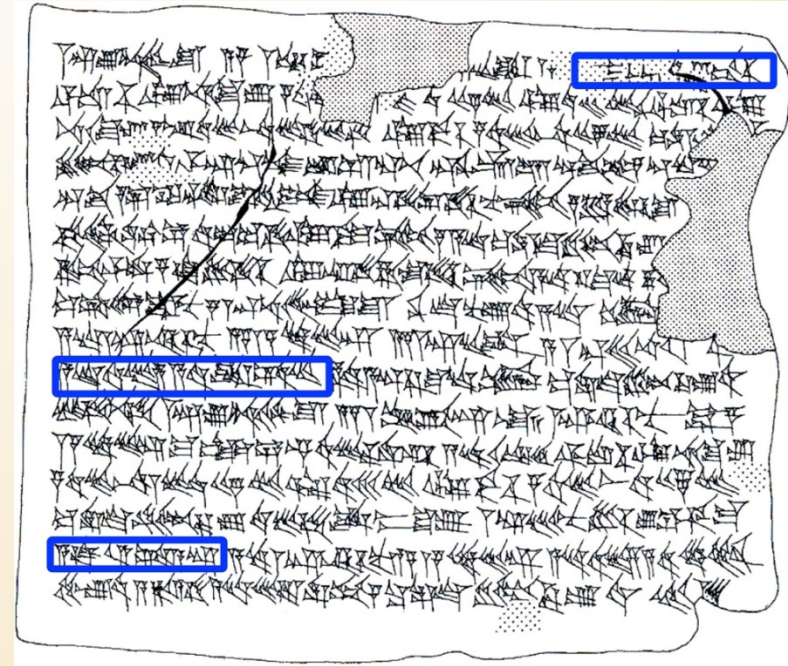
individuals/citation

name instances



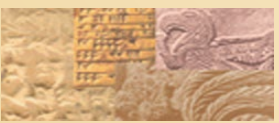
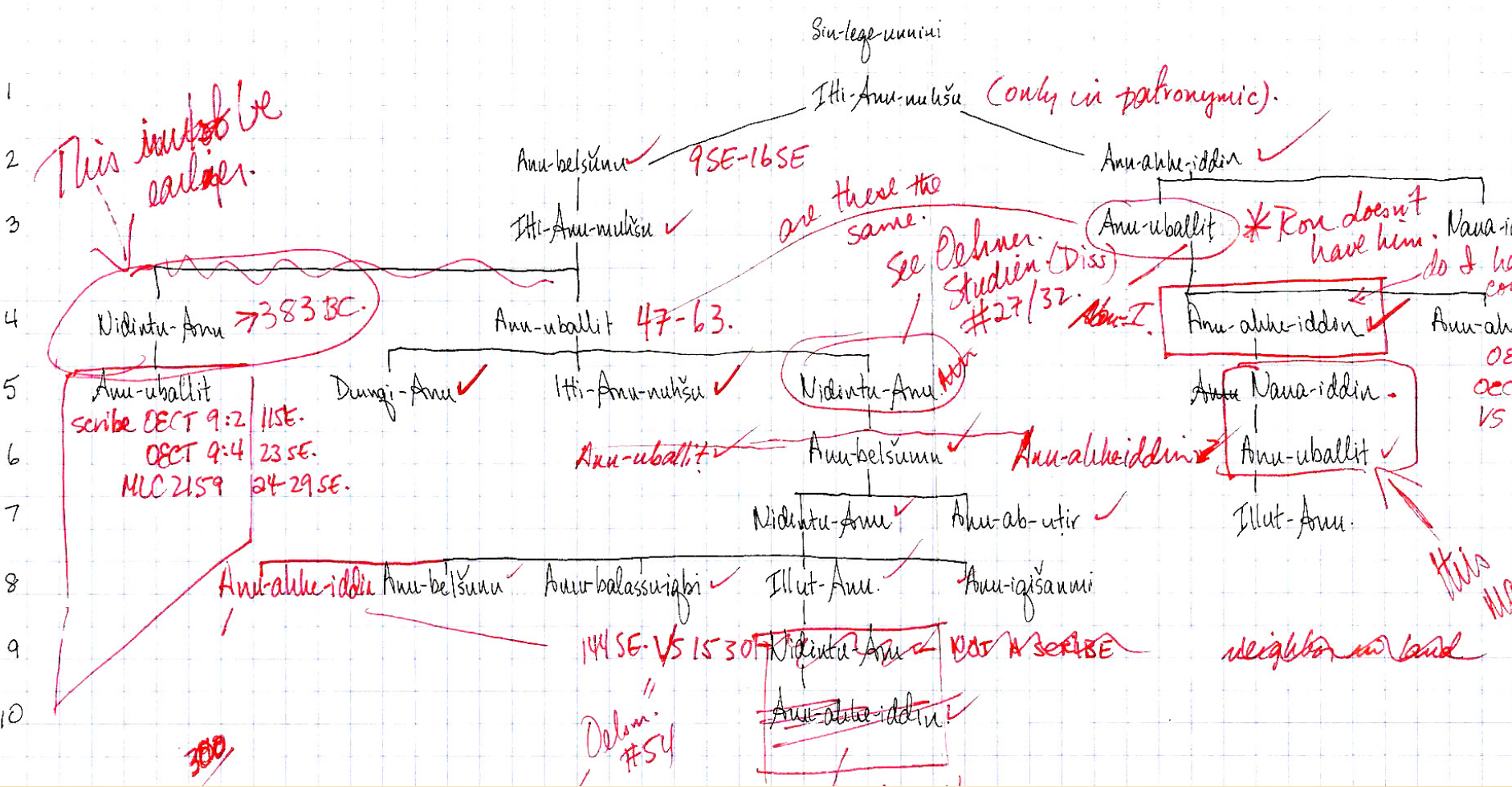
# Data Mining in Uruk Legal Texts

- **Boilerplate text**
  - **repetitive patterns**
    - A sold to B
    - C guarantees for D
  - **attributes**
- **Onomastic data**
  - **standard naming pattern:**  
A / son of B / son of C // descendant of D
  - **papponymy:** name child for (male) ancestor





# Cuneiform Prosopography: Ancient Texts, Ancient Tools

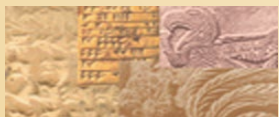




**Needed:  
a better way  
new tools**

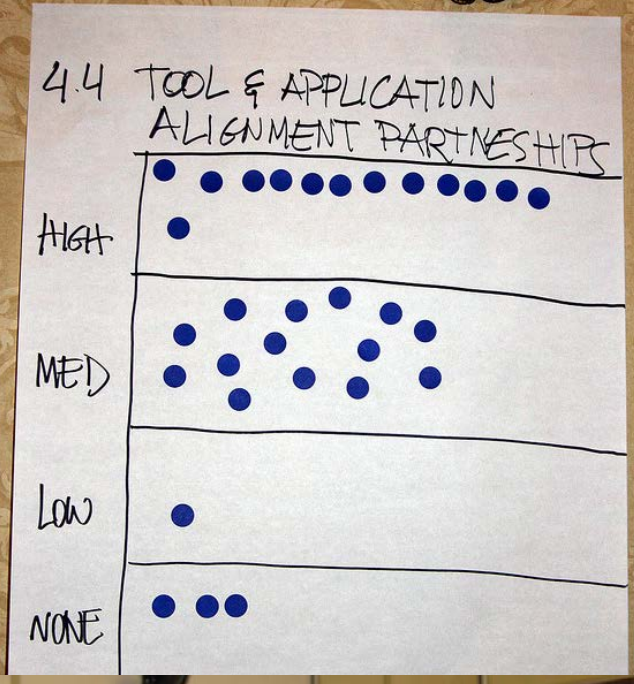


- value-added research
- increase scope of questions/constraints
- framing of new questions

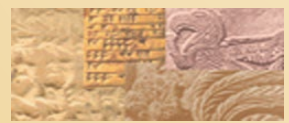


COMMUNITY - Human

- CLUSTERS:
  - VISUAL
  - TEXTUAL
  - PERFORMANCE
  - CREATORS / AWARE
  - FULL SPECTRUM OF STUDENTS
  - SPECTRUM OF USERS - EARLY ADOPTERS

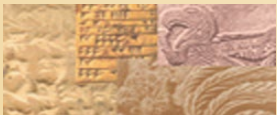


Oracc Activity data mining NLP  
 Corpus HTML TEI  
 papponymy  
 ATF Services XSLT SOA  
 decipherment  
 affiliation  
 Lemmatization



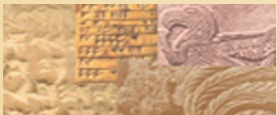
# Project principles

- User Centered Design!
- Humanities requirements
  - Hide the math
  - Allow for a high-level view
  - Have a tool that serves the user
- Technical requirements
  - Models and software must be Reusable, Generalizable, and Scalable
  - Follow best practices: REST/SOA, common platforms & standards, configurable/pluggable



# BPS innovations

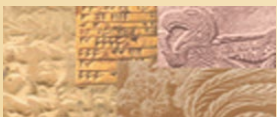
- Digital Humanities application built on enterprise engineering principles.
- Probabilistic model of disambiguation, with extensible rules
- Assertions overlay computed model
- Network analysis to explore model





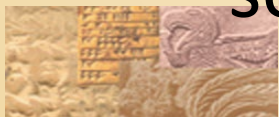
# Assertions

- Probabilistic model
  - Supports uncertainty
  - Assertions control model parameters
  - Assertions override model results
- Workspaces
  - Support hypotheses
  - Build community
  - Track authority



# Social network analysis

- Services
  - SNA engine (computes metrics, features)
  - Filters and pivots to refine scope
  - Visualization kit
- Features
  - Support probabilistic network
  - Axes and features derived from data
  - Support any incoming data, regardless of semantics



## Home Corpora Workspace

[Documents](#) [People](#) [Clans](#) [Settings](#) [Admin](#)

## Set Model Parameters

## Background on the model

The BPS analyzer will try to disambiguate among citations using the same name(s). To do this, it will basically model a new citation-person for each name it finds in a document (including fathers, grandfathers, etc. that are mentioned as qualifiers to the named actors). Then, it will attempt to collapse some of those citation-persons to get to the set of actual (real world) persons mentioned in all the corpus documents. Each citation-person is compared to other citation-persons, and a set of rules is applied to determine how likely it is that the two citations are the same person. The analyzer proceeds in two steps: first it considers all the citation-persons within each single document (*intra-document*), and then it considers the citation-persons across the entire corpus (*inter-document*).

When comparing two citation-persons, the analyzer will first require that there is no conflicting information about the two citation-persons - e.g., if they have different declared fathers, they will be considered as distinct, and will not be collapsed. The rules below allow you to configure whether specific roles must be considered to be distinct, and to control how strong the likelihood that two persons with partial matching name information are the same real world person.

## General settings:

|  |                                 |
|--|---------------------------------|
| Number of qualifications (father/grandfather/ancestor/clan) in addition to forename required to consider a name citation "fully qualified" | <input type="text" value="2"/>  |
| Assumed typical length of active business life (years)   | <input type="text" value="25"/> |
| Assumed typical separation of generations (years)  | <input type="text" value="15"/> |

## Step 1: Intra-document rules:

Rule Steps 1A, 1B, and 1C collapse citations within a single document.

## Step 1A: Consider equally qualified names

|   |  |
|---|--|
| Collapse equal, fully qualified citations<br>(e.g., "PN <sub>a</sub> , son-of PN <sub>b</sub> , in-clan CN <sub>c</sub> "<br>and "PN <sub>a</sub> , son-of PN <sub>b</sub> , in-clan CN <sub>c</sub> ") | <input type="text" value="Always: 100%"/>      |
| Collapse equal, partly qualified citations<br>(e.g., "PN <sub>a</sub> , son-of PN <sub>b</sub> " and "PN <sub>a</sub> , son-of PN <sub>b</sub> ")   | <input type="text" value="Conservative: 30%"/> |
| Collapse equal, unqualified citations<br>(e.g., "PN <sub>a</sub> " and "PN <sub>a</sub> ")  | <input type="text" value="Aggressive: 75%"/>   |

## Step 1B: Consider compatible, but not equally qualified names

|   |  |
|---|--|
| Collapse partly qualified citations with compatible, fully qualified citations<br>(e.g., "PN <sub>a</sub> , son-of PN <sub>b</sub> " and "PN <sub>a</sub> , son-of PN <sub>b</sub> , in-clan CN <sub>c</sub> ")   | <input type="text" value="Conservative: 30%"/> |
| Collapse unqualified citations with compatible, more qualified citations<br>(e.g., "PN <sub>a</sub> " and "PN <sub>a</sub> , son-of PN <sub>b</sub> , in-clan CN <sub>c</sub> ",<br>OR, "PN <sub>a</sub> " and "PN <sub>a</sub> , son-of PN <sub>b</sub> ") | <input type="text" value="Aggressive: 75%"/>   |

# Home Corpora Workspace SNA

Documents Persons Clans Activities Roles

## SNA Visualizer

Views

Curated Nana-Iddin corpus

Layout

Force Directed Radial

Filter (degree)

All Popular Peripheral

Sort

Cluster Degree

Edges

Show Hide

Node names

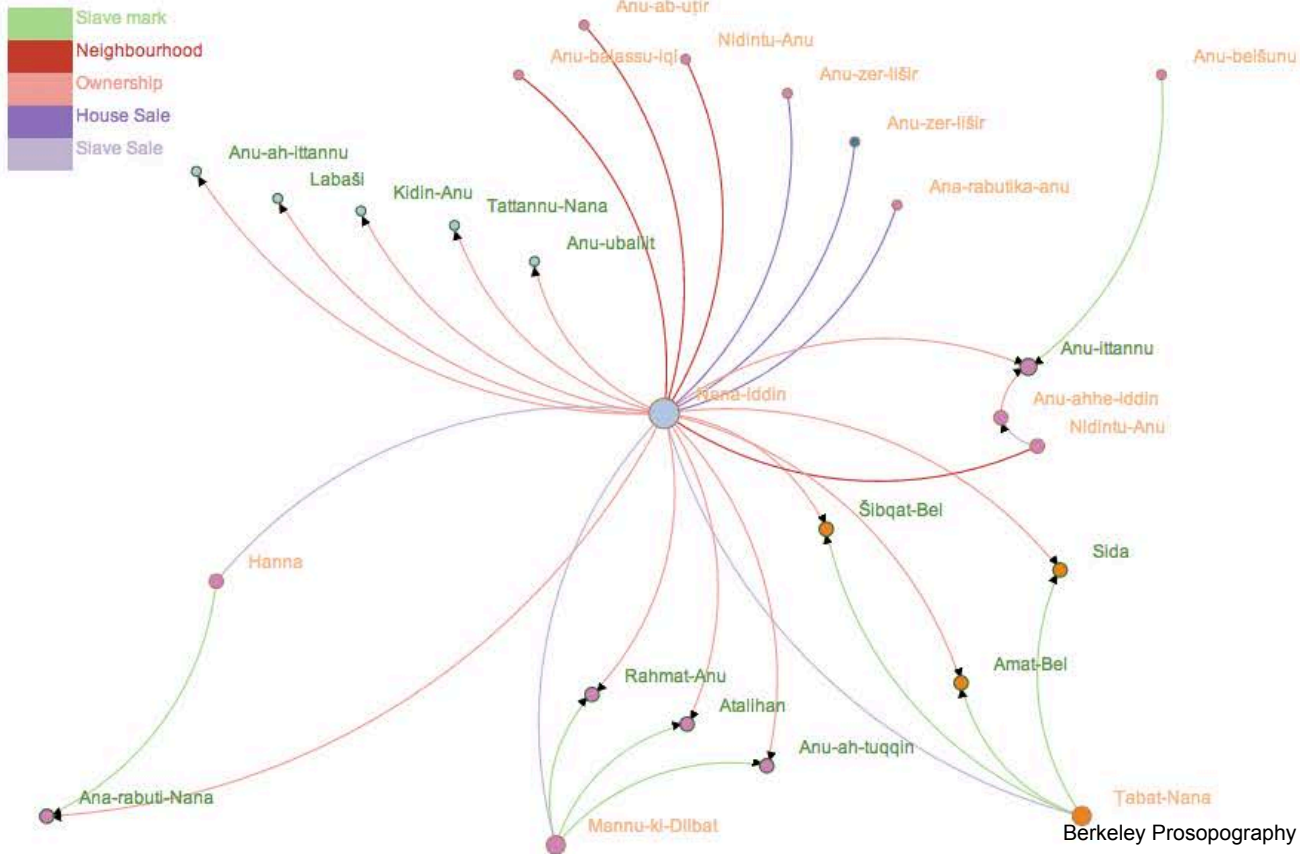
Show Hide

Static views

Family Tree

Zoom

Active Paused



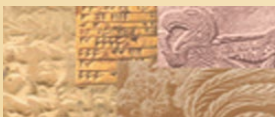


# Questions, discussion

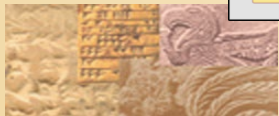
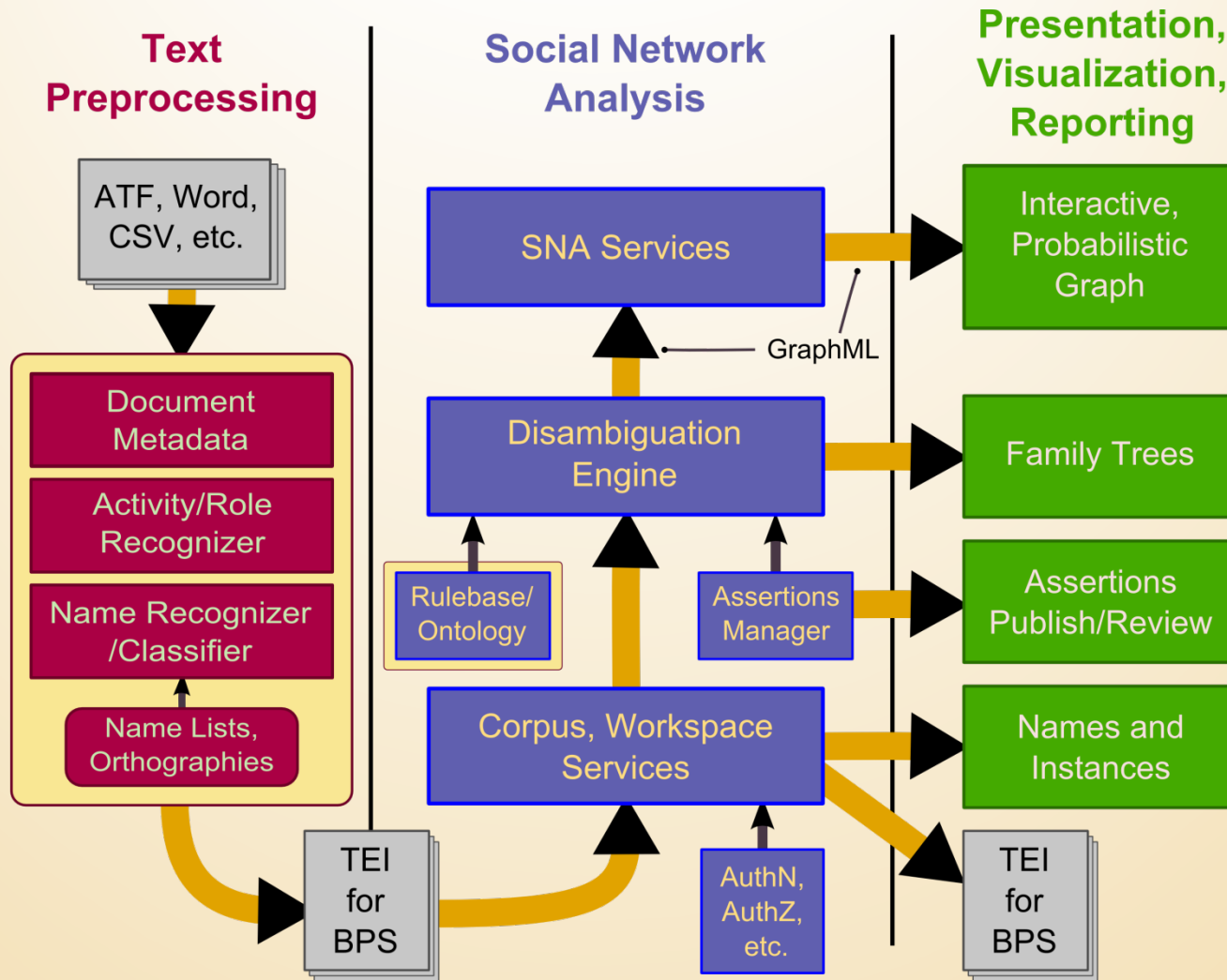
<http://www.berkeleyprosopography.org>

Links below available from About page of site.

- HBTIN project home:
  - <http://oracc.museum.upenn.edu/hbtin/>
- Project wiki
  - <https://wikihub.berkeley.edu/display/istds/Berkeley+Prosopography+Services+Wiki+Home>
- Code:
  - <http://code.google.com/p/berkeley-prosopography-services>
- Contact us:
  - Laurie Pearce ([lpearce@berkeley.edu](mailto:lpearce@berkeley.edu))
  - Patrick Schmitz ([pschmitz@berkeley.edu](mailto:pschmitz@berkeley.edu))
  - Davide Semenzin ([d.semenzin@students.uu.nl](mailto:d.semenzin@students.uu.nl))
  - Niek Veldhuis ([veldhuis@berkeley.edu](mailto:veldhuis@berkeley.edu))



# BPS System Architecture



# Simplified excerpt of TEI for BPS

...

```
<lb n="12"/>
```

```
bab-ba-nu-u2-tu2 a-na SZAM2 TIL{mesz} a-na
```

```
<person id="p6" sex="1">
```

```
<persName>
```

```
<forename>
```

```
{m}dum-qi2-{d}60
```

```
<note type="lem">Dumqi-Anu[ buyer]PN</note>
```

```
</forename>
```

```
DUMU sza2
```

```
<lb n="13"/>
```

```
<forename type="patronym">
```

```
{m}tat-tan-nu-{d}na-na-a
```

```
<note type="lem">Tattanu-Nana[ father]PN</note>
```

```
</forename>
```

```
DUMU sza2
```

```
<forename type="patronym" subtype="grandfather">
```

```
{m}ri-hat-{d}60
```

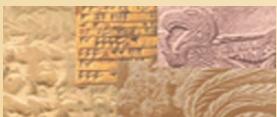
```
<note type="lem">Rihat-Anu[ grandfather]PN </note>
```

```
</forename>
```

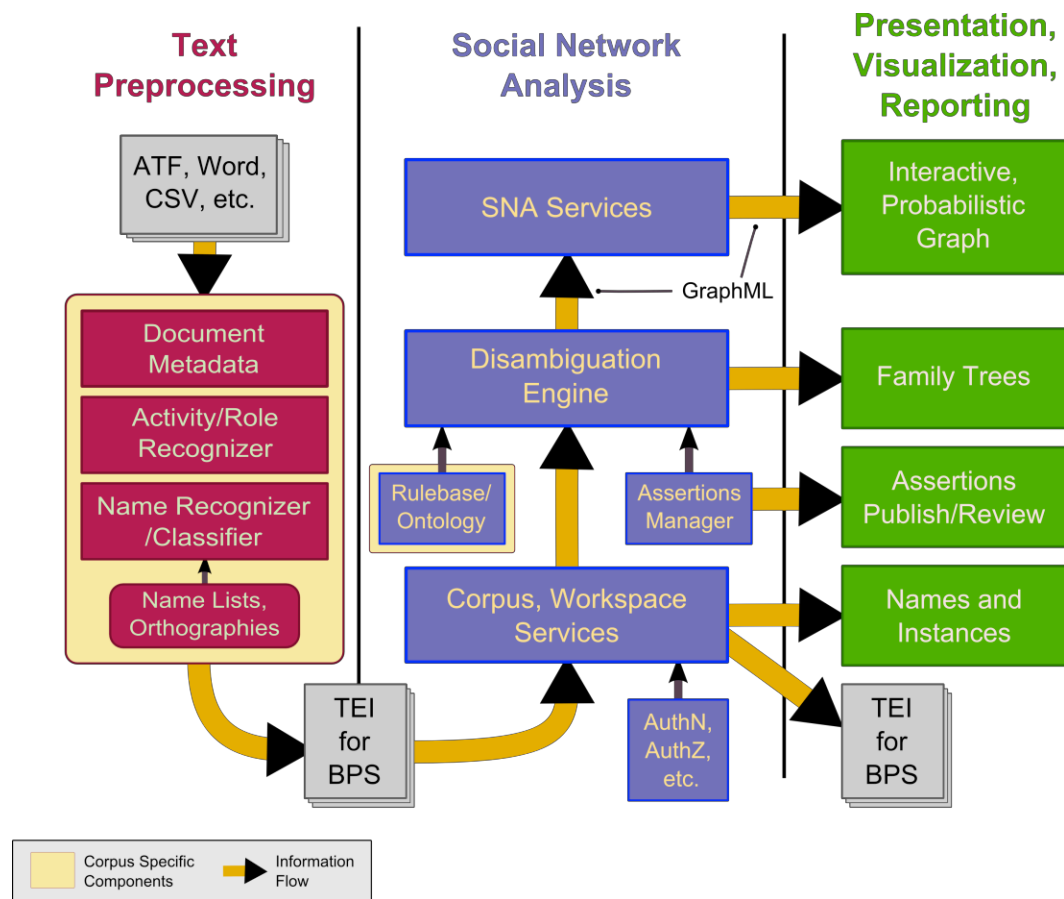
```
</persName>
```

```
</person>
```

...



### Appendix 3: BPS Architecture Diagram



This diagram presents the major components of the BPS system, and the way that information flows through the system. The three primary sections are depicted in three vertical columns, with the major information flows represented as yellow arrows among the components.

The first column, in red, shows text processing steps, in which a corpus is converted from some native format to TEI that includes elements denoting the individual documents, activities within each document, and persons that have roles in those activities. This markup may be generated by hand, or by some semi-automated processes to recognize names, filiation, roles and activities. Because each domain has different language models, many of the components in the text processing section have corpus- (or domain-) specific aspects to them.

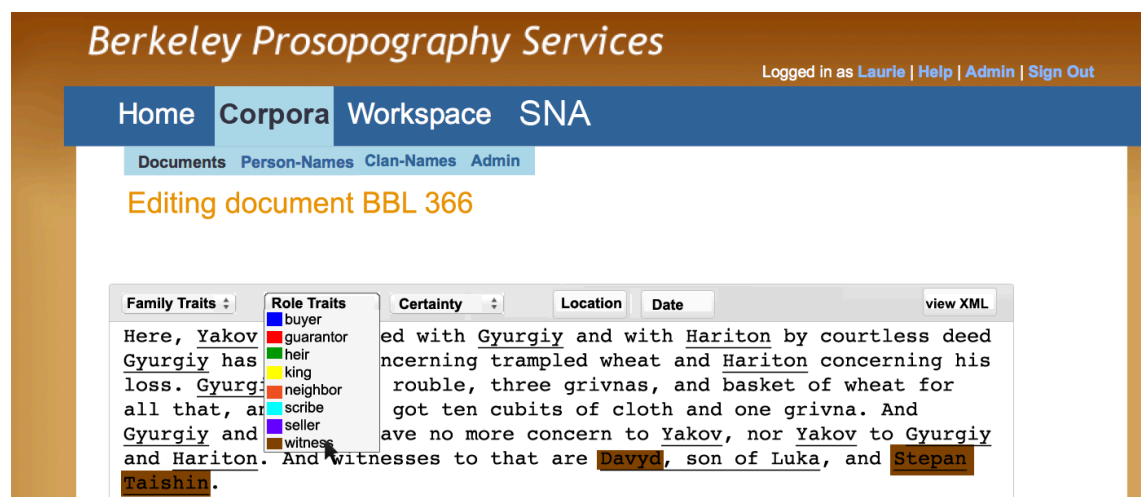
In the second major section, in blue, TEI is uploaded and parsed by the corpus services, and a native data model is built internally. The workspace services share this model, and leverage authentication (AuthN) and authorization (AuthZ) components to support login and access controls on corpus and workspace resources. The disambiguation engine incorporates rules that may be generic or may be corpus-specific, and associates the name citations in each document with actual persons depicted in the texts. It includes support for assertions that researchers make to confirm or reject the possibilities suggested by the engine. Finally, GraphML (a standard XML format) is sent to the SNA services to compute significant features of the social networks. While the basic pieces are in place for all these components, the next phase of the project will generalize and extend a number of these components.

The last section, in green, shows the presentation/visualization of results from various core model and analysis components, including the declared data model in each corpus (names, activities, etc.), assertions that the researcher has made or imported from others, family tree visualizations, as well as interactive network graphs for exploration and understanding. A workspace model will also be exportable as TEI (this may require some extensions).

## Appendix 4: Highlights of the BPS User Experience and Workflow

Effective productivity tools are easy to use and generate a natural workflow. The figures below provide an introduction to the BPS user experience as implemented, or as planned. They are organized in three sections, corresponding to the three columns in the architecture diagram (Appendix 3). Each section contains images of the interface in the order the user generally will encounter them. Captions contextualize the image in the workflow. The generalizability of BPS is assumed throughout, but special attention is drawn to those features that require specific modification by each corpus curator. Nearly all the examples illustrated are drawn from the results of processing the BPS demonstrator corpus, the legal cuneiform texts from the Hellenistic period (331-46 BCE) in Mesopotamia (ancient Iraq). These texts, available digitally in the project Hellenistic Babylonia: Texts, Images and Names (HBTIN), are processed according to ORACC ASCII Text Format (ATF) standards. ORACC provides the TEI (as described in the red column of the architecture diagram), which the corpus curator Pearce uploads to BPS, making it available for other researchers to explore as well. Keeping in mind that the details displayed will differ from corpus to corpus, we present illustrations of work completed and future development in BPS.

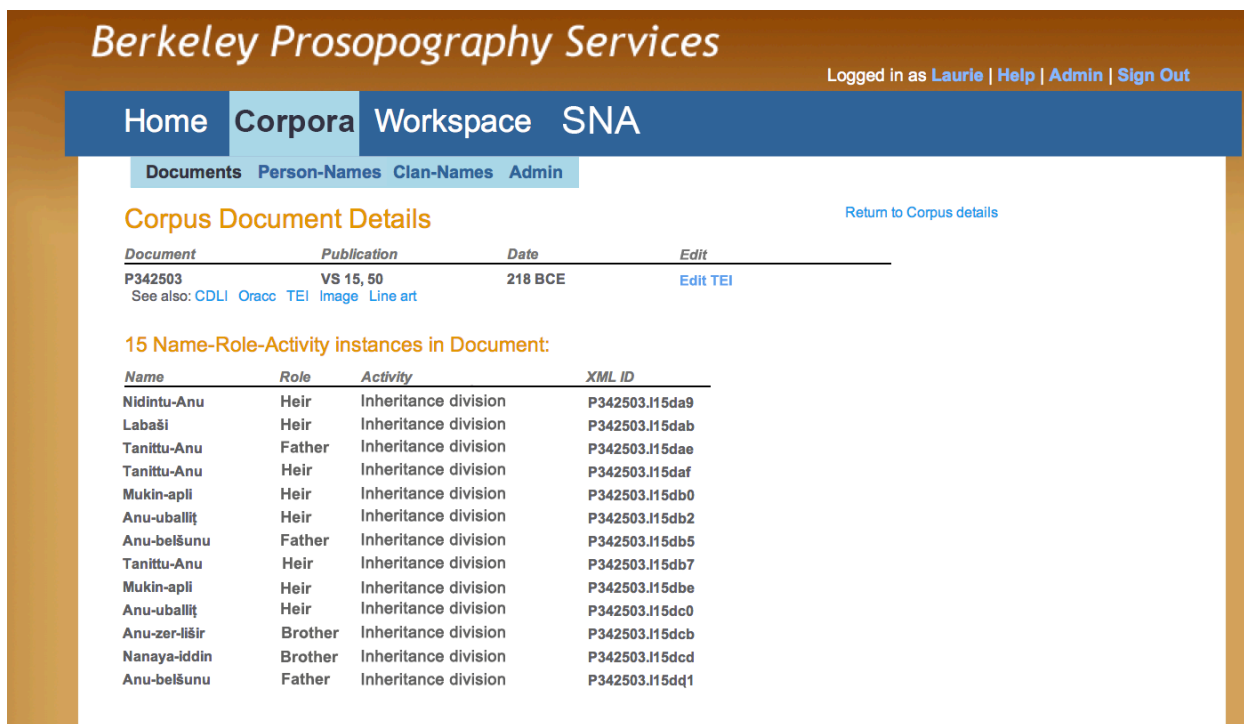
### I. TEXT PREPROCESSING



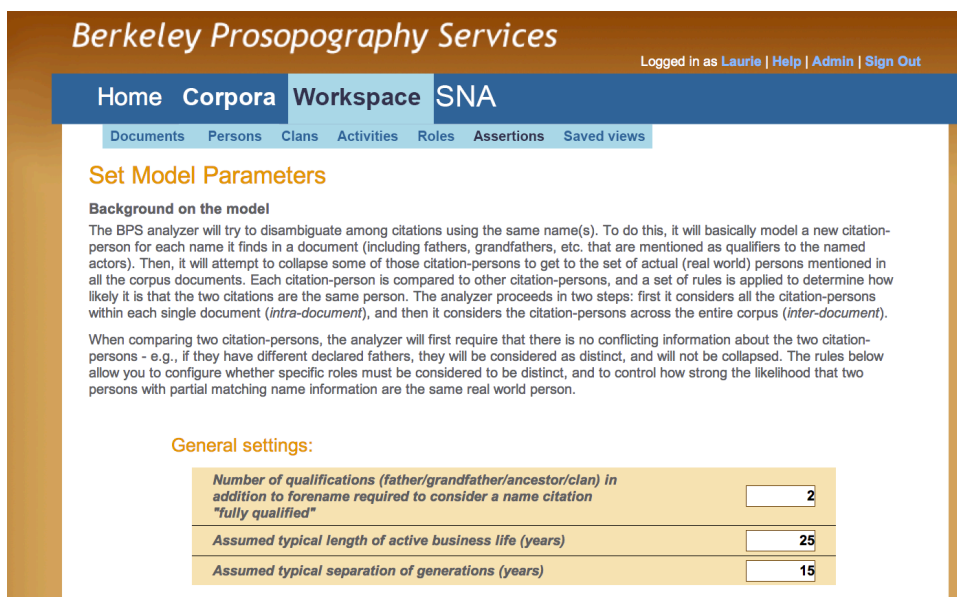
**Figure A. Proposed TEI Editor.** Text preprocessing is a corpus-specific activity that paves the way for the computation of SNA metrics. BPS takes in text decorated with TEI markup, uploaded via a simple interface in Corpora > Administration. In response to user feedback suggesting the simplification of the decorating of TEI with the inclusion of additional tags (for persons, location, roles, activities), we will explore the integration of a WYSIWYG TEI-editor. Drop-down menus and color-coding will make TEI markup as easy as clicking on the desired choice. The TEI editor, located in the Text Preprocessing portion of the BPS architecture, is a service BPS will provide to facilitate enrichment of the TEI of diverse corpora that may not be marked up as fully as the corpus curator requires for social network analysis and prosopographical analysis. The example here is of a text in a corpus of birch bark letters. Although the corpus input is TEI, the markup for in-project use does not include tags for roles within activities; here, two witnesses are tagged.

### II. SOCIAL NETWORK ANALYSIS (SNA)

SNA is the outcome of a number of inputs and computational processes. SNA is particularly successful when namesakes are disambiguated to represent individual persons. Greater certainty that a particular attribute can be positively associated with a specific person increases the accuracy of the disambiguation. The figures in this section depict existing and planned components of the BPS “disambiguator” and “engine” that operate on the corpus TEI. In the **Workspace > Assertions** menu, the researcher accesses the heart of BPS’s probabilistic model, establishing the weight of the criteria (“rules”) to be applied within and across the corpus. As each researcher brings both specialized knowledge of a corpus and a unique perspective on the data within that corpus, BPS’s architecture supports multiple assessments of the data which, in turn, produces multiple outcomes that the researchers can assess and either accept or reject. This model is particularly valuable when dealing with vague, incomplete or damaged evidence.



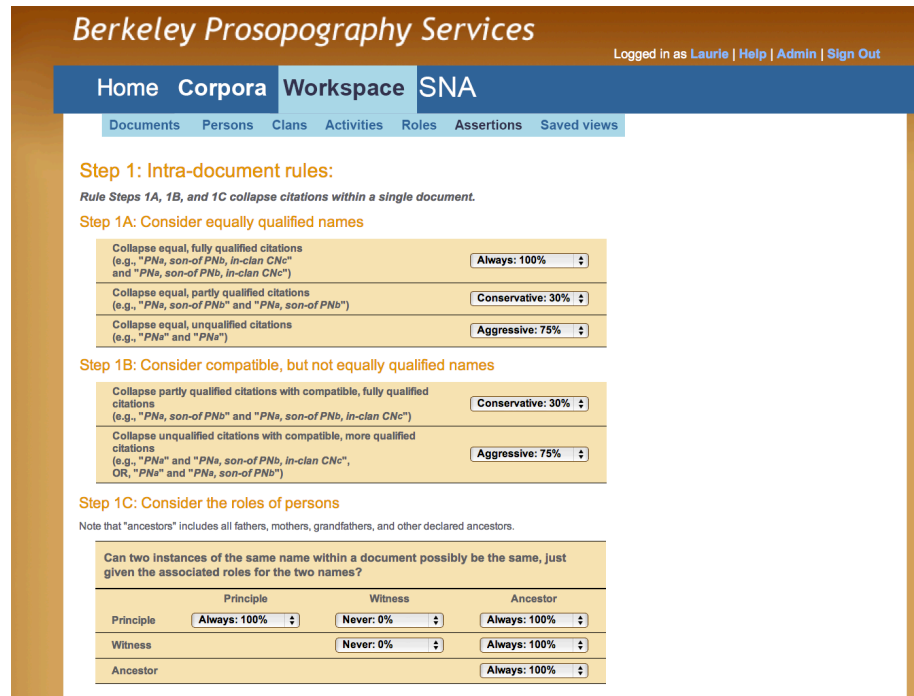
**Figure B. Screenshot of UI for: Name-Role-Activity Instances in a Document:** This view represents a basic display of the name instances, associated with their roles, as they appear in an existing HBTIN record of a single activity. In this instance, the name Tanittu-Anu appears three times, in two different roles (heir and father). The researcher uses corpus-specific criteria to determine how many real persons the three name instances represent. The corpus curator has the specialized knowledge that in this corpus a single individual can not simultaneously occupy the roles “father” and “heir” in a single document, and will be able to apply this and similar knowledge in various stages of setting model parameters (Figure D).



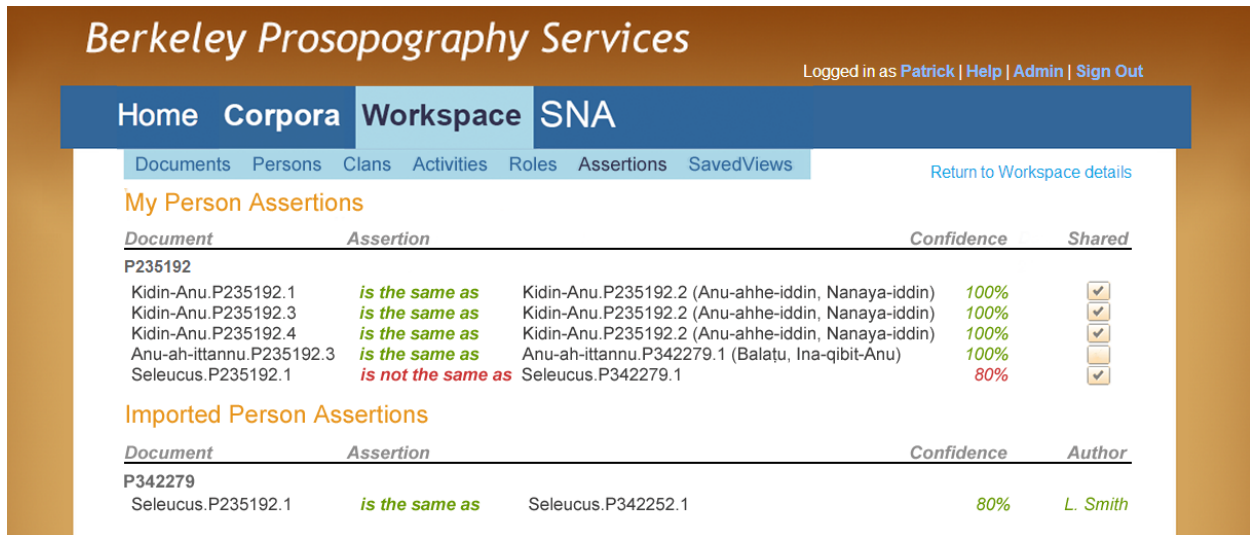
**Figure C. Screenshot of Current UI for Set Model Parameters.** The General Settings illustrated here are those applied to texts in the HBTIN corpus. General Settings are fundamental characteristics that define persons throughout the entire corpus. Specific patterns of name-giving current in Hellenistic Uruk produce multiple



instances of similarly-named individuals. To distinguish among multiple individuals named, e.g. “Anu-uballit, son of Anu-bēlišunu, descendant of Ah’utu”, the corpus curator determines the number of years that constitutes a generation. Name-instances appearing in texts separated by a larger date range are more likely to be unique persons. Other researchers may assign different values that reflect their individual understandings of the same corpus-specific General Settings. Each change in the values of any rule in the Model Parameters is the equivalent of asking “what if”? Changes will have an impact on the computation of the social network and its eventual visualization.



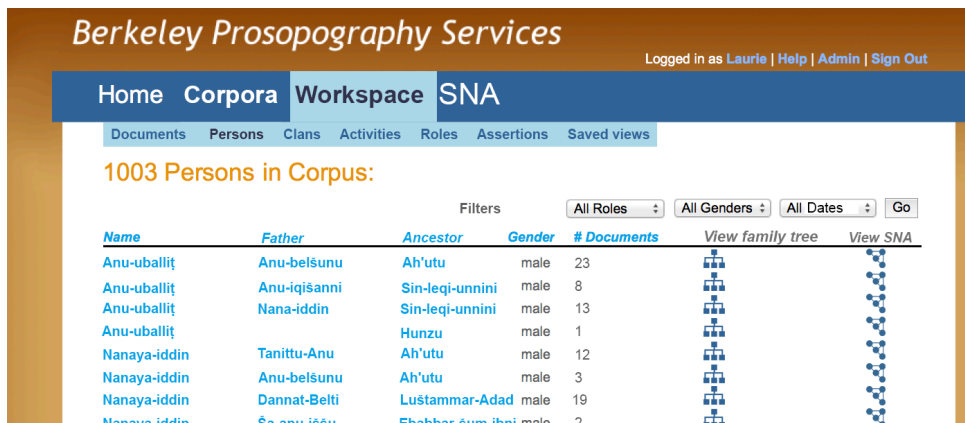
**Figure D. Screenshot of Current UI for Intra-Document Rules:** All name instances in a corpus, both within a single document (intra-document) and in documents across the corpus (inter-document), provide evidence for disambiguation. The rules used to distinguish similar name instances within a text may be the same as those used to distinguish the same name instances in texts across the corpus, in which case the wording of the Inter-Document Rules (not illustrated here) would be identical to that of the Intra-Document Rules. What may differ, however, is the probable validity of a particular rule in the Inter- and Intra-Document contexts. The user sets the probabilities in each context. These probabilities can be adjusted by the same researcher who wants to investigate the likelihood of another “what if” scenario, or by other researchers with entirely different assessments of the probable validity of the rules. The present example shows rules that reflect the researcher’s knowledge and workflow. Knowing that in Hellenistic Uruk, a principle can not also serve as a witness to single transaction, the researcher assigns 0% probability (i.e., Never) that any namesakes in these roles are the same individual, just as would be done “intuitively” in traditional prosopographical research. Another scenario in which the Inter- and Intra-Document Rules might differ would reflect the boost that would apply to the merging of two names based on co-occurrences of a person with other individuals in the document. This rule will be implemented in our next development phase.



**Figure E: Proposed UI for viewing ASSERTIONS:** The assertions reflect the application of the rules and probabilities assigned by the user. When the user interface is fully developed, results of the processing of the rules and probabilities will be represented as in Figure E, with indications of the user’s confidence in the assertion. The assertions model will allow for the tracking of authority, i.e., of the user responsible for the creation of a particular set of parameters.

### III. PRESENTATION, VISUALIZATION, REPORTING

Introduction: Presentation, Visualization and Reporting, the focus of the third column of the architecture, culminates in the production of the visualizations (graphs, family trees) that facilitate the discovery of the relationships between people in a variety of contexts. The outcome of the processes in sections I and II above is the production of a list of the names of all individuals who appear in the text corpus and who populated the social environment of the corpus under investigation—the “real people” of the corpus. The current UI supports textual reporting of names, corpora, clans, filtering by role, and gender. The user can now make choices that will finesse the graphs and diagrams that facilitate and encourage the exploration of the lives of each individual in the corpus and the activities in which he appears. The user will be able to choose to view a graph of an individual that includes relevant data from the entire corpus or the user can focus in on particular features, as illustrated below. Our graph engine produces visualizations that represent the analysis of the data resulting from the implementation of the parameters. Our users and workshop participants have identified features to be integrated into the user-interface, notably the ability to access different views through interactive buttons. Figures F through I are wireframes of features to be developed for the user interface.





**Figure F. View of Persons.** This list of disambiguated persons in the corpus is planned as a starting place for the investigation of social networks. This proposed UI builds on current functionality of the person view, adding links to the visualizations, enabling the researcher to focus in on the particular individual that interests him. The researcher's choice to view a Family Tree or Social Network graph of any of the disambiguated namesakes leads directly to the visualization centered on that individual, as in Figure M. The choice to view a family tree takes the researcher to a screen like Figure M, generated in the current iteration of the BPS visualization software.

Alternatively, the researcher will be able to select and combine attributes to generate graphs that prioritize activities, specific texts, and include multiple individuals associated with them. Figures G, H, I represent the planned expansion of the scope of BPS visualizations to provide researchers with a highly customizable interface through which the choice of attributes and individuals to appear in the graph is made.

| Document | Publication   | Date       | Select for SNA                      |
|----------|---------------|------------|-------------------------------------|
| P235192  | NBDM 91       | 232 BCE    | <input type="checkbox"/>            |
| P296677  | BRM 2, 01     | 304 BCE    | <input type="checkbox"/>            |
| P296703  | BRM 2, 25     | (211 BCE?) | <input checked="" type="checkbox"/> |
| P296715  | BRM 2, 37     | 179 BCE    | <input type="checkbox"/>            |
| P296760  | YOS 20, 008   | 296 BCE    | <input type="checkbox"/>            |
| P342141  | AoF 5, no. 02 | (211 BCE?) | <input checked="" type="checkbox"/> |
| P342142  | AoF 5, no. 03 | (211 BCE?) | <input checked="" type="checkbox"/> |
| P342143  | AoF 5, no. 04 | 152 BCE    | <input type="checkbox"/>            |
| P342144  | AoF 5, no. 05 | (211 BCE?) | <input checked="" type="checkbox"/> |
| P342145  | AoF 5, no. 06 | (211 BCE?) | <input checked="" type="checkbox"/> |
| P342147  | AoF 5, no. 08 | 212 BCE    | <input checked="" type="checkbox"/> |

Figure G. Proposed view for Document selection for SNA

| Activity                          | Select for SNA                      |
|-----------------------------------|-------------------------------------|
| Lease of field                    | <input type="checkbox"/>            |
| List of individuals               | <input type="checkbox"/>            |
| List of officials                 | <input checked="" type="checkbox"/> |
| Inheritance                       | <input type="checkbox"/>            |
| Prebend sale: temple enterer      | <input type="checkbox"/>            |
| Prebend sale: goldsmith           | <input checked="" type="checkbox"/> |
| Prebend sale: butcher             | <input checked="" type="checkbox"/> |
| Quitclaim                         | <input type="checkbox"/>            |
| Real estate sale: house           | <input checked="" type="checkbox"/> |
| Real estate sale: field           | <input checked="" type="checkbox"/> |
| Real estate sale: house and field | <input checked="" type="checkbox"/> |

Figure H. Proposed view for SNA Activity selection



Figure I. Proposed view for Person selection for SNA

## 2. GRAPH VISUALIZATIONS

2.1 A FAMILY TREE depicts a network of individuals linked through both blood and marriage lines. This graph visualization model is functional. It currently operates on an independent stack and requires only integration into the SNA engine to add further utility to the BPS toolkit.

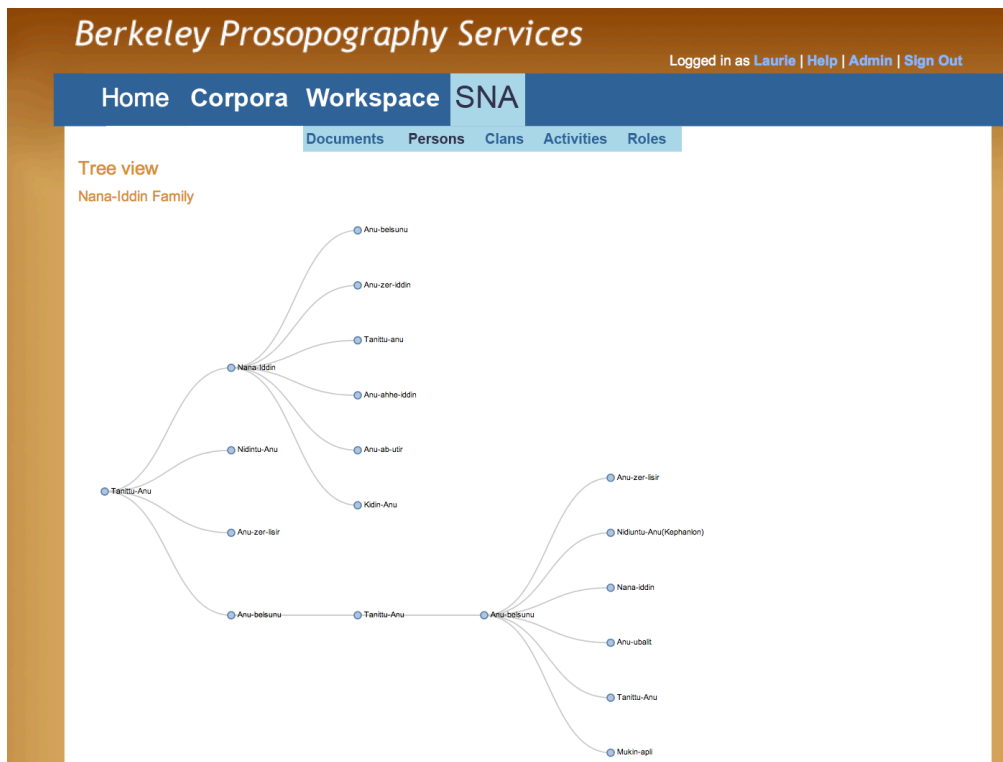
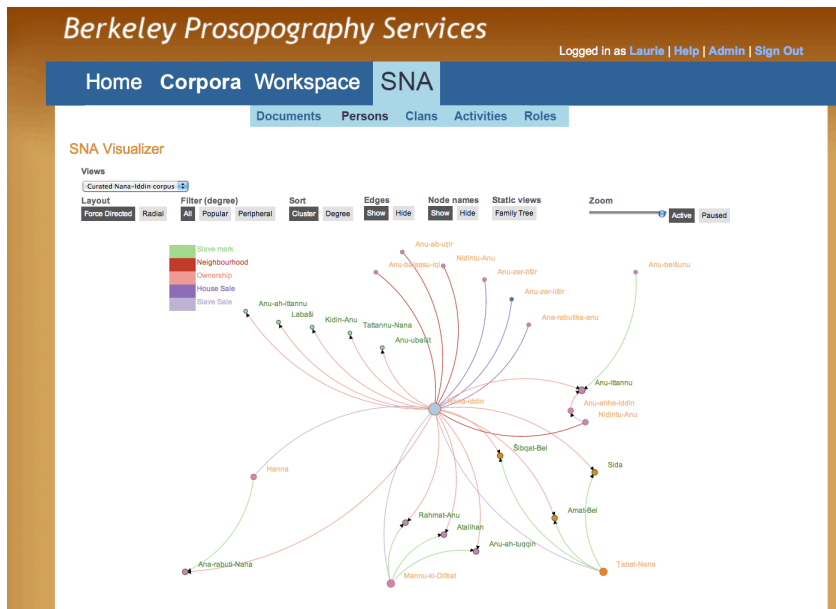


Figure J. Family tree of Nana-iddin. This family tree generator. This traditional genealogical tool presents a family structure in a clear, logical, hierarchy. As complex family trees can be difficult to display and navigate, BPS already supports and generates interactive family trees, allowing researchers to collapse or expand branches of the family tree according to the User's research interest.

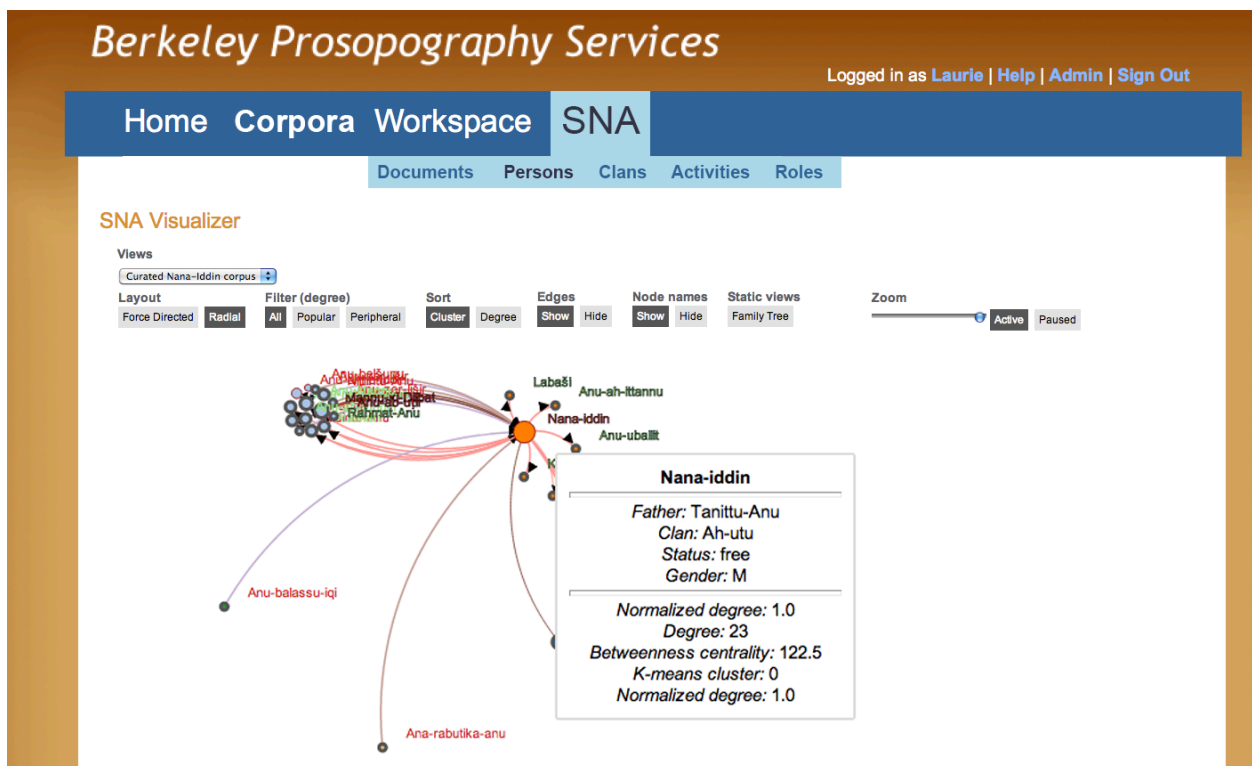
## 2.2 SOCIAL NETWORK GRAPHS

These graphs demonstrate the connections between individuals (nodes, represented as colored dots) throughout the corpus. The lines (arcs) that link them represent the activity through which they are connected. The BPS

Visualizer produces interactive graphs that can represent a social network through a variety of views, allowing the user to explore the individuals and relationships as fully as possible. The graph engine requires final integration into the BPS site.



**Figure K. Current view: Graph Centered on Nana-iddin:** This graph is centered on Nana-iddin, represented by the largest node on the graph, and reflects the social network metric of degree (how many connections exist between this person and others). The color-coded arcs represent the activity which connects each pair of individuals and the arrows reflect the direction of the relationship in those activities (e.g., from seller to buyer). The graph is interactive: the user can zoom in on the graph to enlarge it and can rearrange nodes on the screen to facilitate investigation of groups that initially appear in tightly packed configurations. Double-clicking on a node will “snap off” a sub-section of the graph, providing the user a close-up view of the network centered around the selected individual. This is particularly useful in a densely populated graph, such as in figure L.



**Figure L. Current view of Graph Showing Clusters:** This graph offers a different view of the same information as figure K. The same individuals remain distributed around the central figure, but now appear in clusters reflecting sub-groups of connected individuals. The call-out box that appears when the user hovers the mouse over a node provides additional information about the individuals (status, gender, family connections); below the ruling, the metrics used to compute the social network appear. Most humanities researchers are more interested in the diagram than in the means by which it was produced, but these statistics may be of interest to the more experienced prosopographer.

### 3. BPS WORKSPACES

The **Workspace** environment is a distinctive and innovative BPS feature. It enables the user to ask “what if?”: what if an individual is removed from a set of data—how does the resulting SNA change? How does this impact interpretation of the persons, activities, communities and cultural processes preserved in the text? It also serves as a forum for interaction among researchers, a means to track provenance of assertions, and to make public the results of prosopographical research.

| Name                      | Corpus     | Rule set     | Last modified | Visibility       | Status |
|---------------------------|------------|--------------|---------------|------------------|--------|
| UChicago Russian seminar  | BBL        | BBL-A, BBL-C | 4 Dec 2012    | Private (shared) | Active |
| AATSEEL 2013 talk data    | BBL        | BBL-B, BBL-D | 26 Nov 2012   | Public           | Frozen |
| Romeo and Juliet test     | Folger-R&J | Default      | 15 Nov 2012   | Private          | Active |
| AKK 103 Spring            | HBTIN      | HBTIN-A      | 13 Dec 2012   | Public           | Active |
| AKK 103 Spring merge test | HBTIN      | HBTIN-F      | 20 Nov 2012   | Private          | Active |

**Figure M. Proposed UI for Workspace Dashboard.** This interface is under development and will be the access point for the saved and sharing of explorations, both in progress and completed. Each of the saved views represents a different exploration. A researcher might create different scenarios to use in various teaching environments, or in order to try different rule sets on a single corpus. Combined with the date of modification, these annotations establish the provenance and provide a publication history. For humanities researchers still accustomed to print publication, this authority tracking is comparable to a written footnote or bibliography entry. In digital humanities world, such authority tracking protects the intellectual property of the researcher.

## Appendix 5. Workshop participants

Our workshops brought together specialists in digital and humanities research, who tested evaluated and made suggestions for future BPS implementations.

Participants in the March 2010 workshop:

Cohen, Yoram (Tel Aviv University)  
Crisostomo, Jay (U.C. Berkeley)  
Pearce, Laurie (U.C. Berkeley)  
Schmitz, Patrick (U.C. Berkeley)  
Seri, Andrea (Oriental Institute, University of Chicago)  
Tinney, Steve (University of Pennsylvania)  
Veldhuis, Niek (U.C. Berkeley)  
Waerzeggers, Caroline (University College, London)  
Wagner, Allon (Tel Aviv University)  
Wunsch, Cornelia (School of Oriental and African Studies, London)

Participants in the April 2011 workshop: (including participants via Skype)

Crisostomo, Jay (U.C. Berkeley)  
Kozuh, Mike (Auburn University)  
Nielsen, John P. (Loyola University, New Orleans)  
Pearce, Laurie (U.C. Berkeley)  
Schmitz, Patrick (U.C. Berkeley)  
Tinney, Steve (University of Pennsylvania)  
Veldhuis, Niek (U.C. Berkeley)  
Waerzeggers, Caroline (University College, London)  
Wagner, Allon (Tel Aviv University)  
Wunsch, Cornelia (School of Oriental and African Studies, London)

and these invited members of the Berkeley IST staff:

Alvarez, Fernando  
Greenbaum, David  
Masover, Steve

## Appendix 6. Cited URLs:

BPS blog: <http://berkeleypros.wordpress.com>  
BPS Code repository: <http://code.google.com/p/berkeley-prosopography-services/>  
BPS wiki: <https://wikihub.berkeley.edu/display/istds/Berkeley+Prosopography+Services+Wiki+Home>  
BPS: <http://berkeleyprosopography.org>  
HBTIN: <http://oracc.org/hbtin>  
ORACC: <https://oracc.org>  
SNAC: <http://socialarchive.iath.virginia.edu>