

# BPS Development Meeting Notes

[Task List](#) | [Templates](#)

**June 1, 2015**

**Agenda:**

1. Davide's hire
2. Davide update us on work done.
3. Next step(s) for Davide
4. Laurie's work with Micaela (CTP) and Esther deGroot
  
5. hire for DH Mellon (with CTP)

**Action points:**

**November 19, 2013**

Exploratory meeting with Babylotec project/team from Berlin (Vordersasiatisches Museum, Cornelia Wunsch and Tobias Schmidt)

**November 5, 2013**

Agenda:

- NEH timeline
- Leiden recap:
  - tech fail
  - Davide
- Berlin Babylotec
  - diagram
  - planning
- France-Berkeley opportunity

Action points:

**October 9, 2013:**

Laurie, Niek, Patrick, Asad Ahmed (UC Berkeley NES, Islamicist)

We invited Asad to share with us how he uses prosopographical data and SNA in his research, and to explore the possibility of his participating in BPS's development of support for integrating databases into the intake of data.

The main points:

1. Asad has used Excel to construct databases and Pajek to run SNA analytics. The formats make some searches difficult to perform. The sources he uses are primarily genealogical records, which transcribe family lines, recording information about the members of a family, and then proceed to list members of the second generation and describe their genealogical information. Typically, each member of a generation is listed in chronological order (oldest son, next oldest son), and then the associated info is provided.

2. Limitations of the sources:

Ambiguities in names of the form:

- 1. Mohammad son of Mohammad of Mohammad
- 2. Sources relying on each other
- 3. The sources rework earlier sources and there are deliberate, as well as unintentional, forgeries.

3. Other potential areas in Islamic studies in which BPS could be useful:

- intellectual transmission, as in the attributions of chains of authority in reports from the prophet (in the Hadith). Asad's colleague at Stanford is an authority on this area; Asad would bring him into the conversation.

Patrick observed that many of the assumptions that Asad makes in working with his data are similar to those BPS works with.

Observations and follow-up:

- Asad's colleague could be contacted for an exploratory conversation.
- There was general agreement that our resources should not, however, be diverted to development cycles that lead in directions other than those BPS is currently focusing on.
- BPS would be interested in serving Asad's research needs, helping to simplify the entry/processing of prosopographic data, and increasing the functionality of the research agenda.
- We would invite Asad to a/the meeting(s) in November with the Berlin team to explore the shape of his database(s), and the heuristics he uses in disambiguating.

Action points:

- Laurie invites Asad to Nov. meeting(s) with Berlin team
- Request that Asad articulate the heuristics he utilizes in disambiguating namesakes
- Invite Asad's Stanford colleague to one of the early Berlin team meetings?

**October 2, 2013:**

**Laurie, Niek, Patrick**

**Agenda: consider various areas for development and support, to sustain continuous development and implementation of BPS**

- **Strategic Planning:**
  - I-School

- Funding
- Connecting the pieces

### Action Points:

- Campus Resources:
  - URAP: Laurie to review our call with Stefanie Eberling to increase applicant pool
  - Quinn: Laurie and Patrick: how do we ID recruitment possibilities/individuals for recruitment
  - API and software engineers: PLS to contact for leads
  - dlab: as a clearing house for internships. PLS will report following meetings with Cathryn Carson
  - Campus initiatives on digital leadership: PLS
  - PLS will contact EECS head, visualization course for nterns
  - PLS may have access to some funds for coding support
- I-School:
  - Niek to contact dean AnnaLee Saxenian

Time-line to be developed:

### Observations:

- Niek raised points for consideration in building models for the BPS future.
  - a day-to-day tech lead working under Patrick as chief architect
  - finding monetary support and manpower to direct and handle more of the day-to-day technical work.
- Patrick agreed these points are worth pursuing. He noted that people have been interested in the BPS problem space, and that faculty help would be a plus in grants
- Niek suggested that the ISchool faculty could be integrated to the research problem of BPS. He has had a very informal conversation with ISchool dean AnnaLee Saxenian, and will follow up with more formal conversations, including points raised in this planning session.
- Patrick: mentioned that while Project Bamboo has stalled, [Bamboo DiRT](#) has gained traction, and may draw collaboration with other research initiatives. We will explore these in conversation with Quinn.
- NEH Digital Implementation grant: LEP will review schedule and prepare for resubmitting the grant in the next CFP (deadline February 19, 2014)

### September 25, 2013:

#### Agenda:

- pre-planning for Berlin Babylon project team visit
  - scheduling blocks of time
  - arranging skype intros prior to Nov visit
- BPS strategic planning
  - I-school, at Niek's initiation
  - approach Islamicists in NES as another set of potential users
- DocEng 2013 review

**Action points:**

- Laurie sets up skype call with Berlin team and Niek: week of October 14
- Laurie invites Assad Ahmed for exploratory conversation re: data-base and his workflows
- Laurie sets up meeting with Niek and Patrick for strategic planning prior to meeting with Anno
- Patrick compiles notes on DocEng 2013 workshop presentations that were useful (slide stacks in BPS dropbox folder)

**June 4, 2013:**

- abstract for ACM paper
- date for review of Prosop workshop, skype with John Nielsen
- TEI–assess what needs to happen to import new
- Issues list/JIRA switch
- URAP requirements
- d-lab job listings

**May 14, 2013:**

- NEH final doc
- Santa Cruz conference in October: topic/set time line/responsibilities
- d-lab job listings
- URAP requirements
- review issues list and set meeting for JIRA
- summer time line
- tei upload and view: diagnose what's not working

**May 7, 2013:****Agenda:**

- **review NEH final doc**
- Santa Cruz conference in October: topic/set time line/responsibilities
- d-lab job listings
- URAP requirements
- review issues list and set meeting for JIRA
- summer time line

**Action points:**

- **Patrick edits tech section(s) of NEH final report**

**March 4, 2013:****Agenda:**

- review script outline for teleconference

- review needs for practice session on March 8
- background review for discussion at I-School on March 22 afternoon
- template updates

**Action points:**

**February 12, 2013:**

**Agenda:**

- confirm tele-conference date and practice date
- review king, date and PN markers in templates
- update PLS on HBTIN work on son/descendants and NENNI
- content of March 22 I-School convo with Cliff Lynch, Ray Larson, et al.

**Action points:**

**November 5, 2012:**

**Agenda:**

- ECAI timeline
- NEH Grant
- Townsend humanities poster session
- tele-conference needs

**October 29, 2012:**

**Agenda:**

- ECAI timeline
- how LEP will get demo to run for Townsend program on 11/13
- report on reactions from previous workshop participants
- needs list of tele-conference
- eat, drink and be merry (after we conclude the business) and wish Davide safe travels, hurry back!

**October 22, 2012:**

**Agenda:**

- develop ECAI paper timeline and rough-out main focus
- schedule graphML how-to for Laurie with Davide

**October 15, 2012:**

**Agenda:**

- review presentation
  - comments from audience
  - self-review
  - ECAI
- outline NES brown-bag presentation, decide deck and what slides need prep (Laurie brings preliminary work to meeting)
- project review
- code hand-off
- tying loose ends up
- tele-conference with previous workshop participants: schedule Cal dates and contact others
- "Aufwiedersehen" social event

Action points:

PLS: ECAI paper deadline?

Tying up loose ends:

- code handoff: everything in repository
- java docs, what do with output? install in another tree? in source tree? BPS site?
- final report from Davide: preferably before leave
- SNA checked in and integrated to build already
- PLS to move repository to gitHub, DS create branch for experimental stuff

Family trees: how to produce graphML. directed graph (fathers can have multiple sons, sons not have multiple fathers. how to represent mothers) nodes and arcs

LEP: graphML, work to create some

tele-conference: LEP to initiate: gotomeeting, DS suggests google hangout Europeans 5-6 or 8-9, suggest W/F starting Jan. 22.

check-list of outstanding threads: DS by end of next week

DS help checking in to service?

LEP: make sure Davide on staff list of wiki

PLS: affiliate DS to continue in VSPA

### **October 12, 2012: I-School presentation**

We invite you to look at our presentation: [ISchool Oct2012.pdf](#)

### **September 24, 2012**

Timeline for I-School:

- Fri. Oct. 12: Talk
- Mon Oct. 8: final check
- (10/4-5 PLS OOF; 9/24 -10/2 Davide OOF)

- Thurs. 4th: DS and LEP slide review and discuss talk
- Mon. 1st: PLS & LP: slide drafts review: establish order and ~timings, sequence; sketch from Davide

### Action points:

LEP contact DS for Oct. 1, 4 contributions

### September 18, 2012

Planning for I-School and ECAI talks:

- consider tech report (unpub) for feedback
- I-School: time on IT4DH focus
- PLS: heritage of assertions model (paper for class at I-School), how it has slowed us down, configurable rules?
- ethnog:
  - how found
  - how long get to point
  - problems
    - other things to do
    - funding
    - people who want a piece of this
    - dependencies
    - can make progress: Davide, PLS, LEP interns, foundering of "I-School generalist", not much middle ground between tech and humanities
- flesh out more how assertions model will be used and useful, knowing we won't have much experience with it
- solicit reactions of which key problems will be of interest for which venues
- have application report but don't have scientific evalu of assertion model, new functionality, what eval against for usability
- frame as conversation, less scripted and emphasize "conversation and partnership"

Secondary goal: Davide's presentation: making progress, how gather requirements, tools needed, how refining designs, what learned about perf, challenges of integration, what to continue to look for going forward. (wind up declare done or where are research questions that come out of/relate to this? resuability: how much? modeled correctly for other corpora? other kinds of questions? linked data? architecture challenges.

Consider other meetings/venues to present/publish

- ITDL
- knowledge representation
- aim for 10-pp paper, using ACM guidelines

Action Points:

- LEP will map outline and assign sections
- begin to assemble slides

## September 11, 2012

### Agenda:

- complete abstract
- each contributor present outline/bullet-points for talk(s)
- timeline for drafts, practice run-thru and revisions
- consider possible slides for ppt
- NES brown-bag lunch: October 24:
  - 12-1, 254 Barrows. (start on Berkeley time) 20-30 mins. talk, remaining time questions
- tele-demo:m
  - date/time (consider 9 time zones, potentially)
  - logistics:
    - time needed: 2hrs?
    - setting date: doodle: possible dates/times AM best, to account for time zones:  
Wednesday, Oct
    - 9/ 16; Friday Oct 12/19
    - skype or otherwise (for larger group?)
    - location
    - equipment
  - invitees:
    - previous workshop participants and supporters: Nielsen, Kozuh, Seri, Waerzeggers, Wunsch, Tinney, Robson, Cohen (Wagner, Kedar), Robson; possibly Garfinkel, Langin-Hooper, Wallenfels
    - NES: ancients (including students), A.Ahmed (NES Islamic)
    - IST: Masover, Greenbaum, Alvarado?
    - I-School?
    - Townsend?

### Action Points:

## September 6, 2012

draft of abstract completed for October presentation and ECAI conference. posted to googledocs

### Action points:

Patrick to review and contribute for meeting Sept. 11

## August 27, 2012

### Agenda:

plan paper for October presentation to I-School and ECAI conference (Dec. 7-9 at Cal)

### Brainstorm/Outline:

10 pages: (breakdown)

- 1 - boilerplate: title/abstract, references, khds (epigrapher fails here)
- .5 - Intro
- 1-2 - context, lit review
  - tech

- UCD
- DH/prosopog
- 7 - Body of Paper: Topics to cover:
  - 2-DS/PLS-Viz/UI: resource allocation, (FB for Hellenistic Babylonia as model, admittedly populist)
  - 1-DS- SNA: wrapper library, abstractions, measures, SOA/SNA efficiency, probabilistic graph issues
  - 1-PLS/DS-SOA: little:
    - semantics of ROA
    - payloads, chaining issues
  - .5-1-PLS-assertions, draw on tools paper
  - .5-PLS-rules configuration (draw on tools paper)
  - 2-LP-IT4DH: autoethnography, UCS for users w/no IT tools; and devs with no. domain knowledge
    - UCD
    - User mode?: intro: theme and frame for paper

Davide and Laurie began abstract

Action points:

- bios needed from Davide and Laurie
- abstract to be completed week of 9/3

**August 14, 2012**

Agenda:

Review Davide's charter and milestones

### **BPS Internship: August review of milestones**

Original goals and the current state of achievement:

- Creation of a flexible and interactive visualization environment for historical social data
  - – Development of an appropriate visualisation layout for family trees and community structures
  - – Development of navigation and exploration features (pan, zoom, link- age views, connectivity highlighting, genealogy expansion, etc.)
  - – Development of search and filtering options
- Creation of a Social Network Analysis framework to provide researchers with new and relevant insights about the research domain.
  - – Implementation of common SNA metrics: done
  - – Implementation of a SNA sub-structures analyser: done
  - – Development of a scalable, modular and extensible SaaS-oriented de- ployment solution: most of this is ready; needs to be checked in and integrated (error logging to be written in)

### **Revised milestones:**

1. Aug. 30:

- checking in SNA and code review and handoff
- SNA code reviewed and contributed

- java docs
- code still to write: output layer betw SNA / graph context

2. Sept. 17:

- Test code, functionality and regressions: requires 1 week, this date is not necessarily firm.
- Acceptance tests

3. Oct 12:

- Visualizer working prototype
- Niek and NES demo

4. Oct 23:

- doc review session: java docs, end user docs, package documentation, installation notes, arch notes, experience report.

5. Oct 30:

- Visualizer finalized.
- Handoff. Testing finished and corrections implemented

#### **Remaining project tasks and associated action points:**

- demonstrate integration of SNA/viz/probabilistic engine/UI: Davide consider this and develop schedule for the architecture meeting week of Aug. 21
- requirements for viz and then consider with Patrick. will fall into scope of SNA. PLS will adapt. intermediate layer
- develop list of individual deliverables in viz:
  - viz packages: family tree layouts (at this point, smple single root, tree)
  - Laurie to provide desiderata for Aug. 21 meeting
  - list original scope for viz, reprioritize
    - Laurie and Davide provide for Aug. 21 meeting
    - Presentations:
      - ISchool:
        - Patrick: check schedule of Friday seminars through Oct.
        - story should include heavy dighum component: LEP involved
      - Near Eastern Studies:
        - department brown bag
        - specialized session for ancients
        - LEP: check with M. Larkin for possible dates, try to piggyback sessions

#### **August 8, 2012**

Mid-internship demo sets out to:

- Describe progress in developments of BPS architecture and implementation of SNA (&viz)
- Clarify status of web service and distinguish between web-service and local version

- Demonstrate what information can be extracted from a subset of corpus texts and the analytics that fuel the viz
- produce and demonstrate the graph

**Patrick:**

spoke about where we are and what remains to be implemented. He laid out the course and progress of Davide's internship. He has approached the apprenticeship as an opportunity for Davide to develop project development and management skills that will serve his professional development. To meet these goals, the beginning of the internship included a great deal of background reading and much discussion of the architecture and conceptual foundations of BPS. Once the overall picture was understood, considerable attention was given to decoupling the components from the overall architecture. This was an outgrowth of and contributes to the project demands that the architecture be corpus agnostic and generalizable. Learning how to develop and deploy stacks on the local hardware required a fair amount of time early on. Early on, Davide and Laurie also met without Patrick, giving Davide the opportunity to interact directly with a humanities researcher. Patrick spent a great deal of time on these fundamental steps and thus paved the way for Davide to move into the phase of developing the components that will function in the SNA reasoner. This meant, however, that Patrick was not able to devote as much time as he had hoped to the processing of many of the tasks on the issues list, many of which were direct outgrowths of the March 2011 workshop. BPS was hit by a number of known spammers and much time was needed to address the security issues here. However, incremental progress in BPS functionality has been made and the newest additions appeared just before the workshop; Laurie will review the features in the next week or two.

**Laurie:**

gave a brief lead-up to demo. She explained that she hand-selected the texts with which Davide worked: documents from the Nana-iddin family archive (most of which has been published in [Doty 1978](#)). A subset of these texts was chosen because it contains known connections among the persons mentioned in the corpus, and presents evidence from a variety of text-types. These texts would provide a maximum payoff in demonstrating results at the same time that the data was easily accessible for Davide. At this point, BPS is not able to process the TEI, so Davide hand generated the (corpus agnostic) graphML that fuels the components of the SNA reasoner. She pointed out that while Davide is working locally, the demonstration is running as a web-service, that is the visualizations are generated by the software and are not mock-ups. The demonstration will show that we have the capability to handle the data correctly and to generate useful and meaningful outputs.

**Davide:**

began his presentation with comments about which metrics he chose to present. In response to Niek's questions, he clarified that all of the metrics and the computations conform to standard SNA practices. He made clear that several of the metrics can be computed in a variety of ways but he opted to utilize only one method—although the architecture would allow for the implementation of other computational modules should a research so desire.

In presenting the graph, he showed the graphML he created from the text information. This is the only step in the demo where manual insertion of data was done. Once the graphML was produced, he ran the libraries of computation modules and Gefi UI rendered the graphs. The central graph illustrated the centrality of Nana-iddin (and the relative sizes of the nodes indicated the degree of each node.) The ability to choose different layouts of the same graph and the ability to focus on document views were shown. Niek raised the question of how the central graph (and particularly the document view) improved the researcher's tool kit. During this discussion, Davide dove into the code and implemented the ability to generate searches for requests on "Neighborhood", "Slave sale", "Ownership", among others. Choosing any of these searches returned graphs of the relationships over the entire demo corpus. Of particular interest was the graph returned on Slave Sales, which produced four distinct graphs (see the center graph in the second row in the photo below). Not only did this demonstrate the functionality of the SNA parser, but also was a testament to Davide's comprehension of how to implement the research needs of humanities scholars.

Davide pointed out that further development will take place in an interactive mode, testing and developing and retesting. This will mean that not all functionalities will be completed but that the progress will continue to insure that project architecture, development and functionality continue to remain integrated and mutually supported.

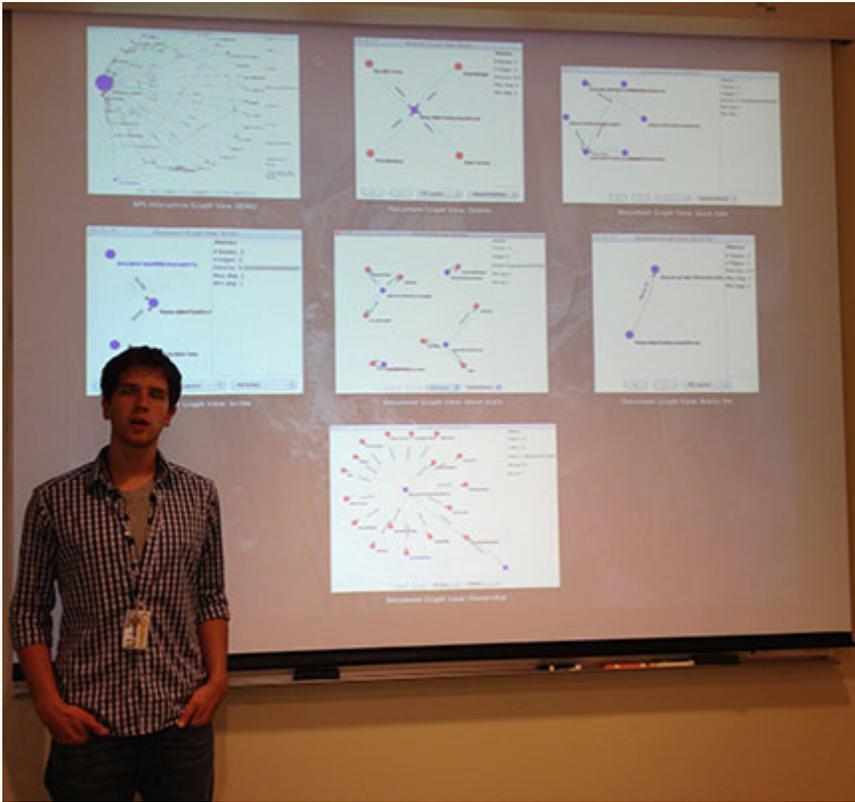
**Prospective:**

Next week: realign DAvide's charter and milestones

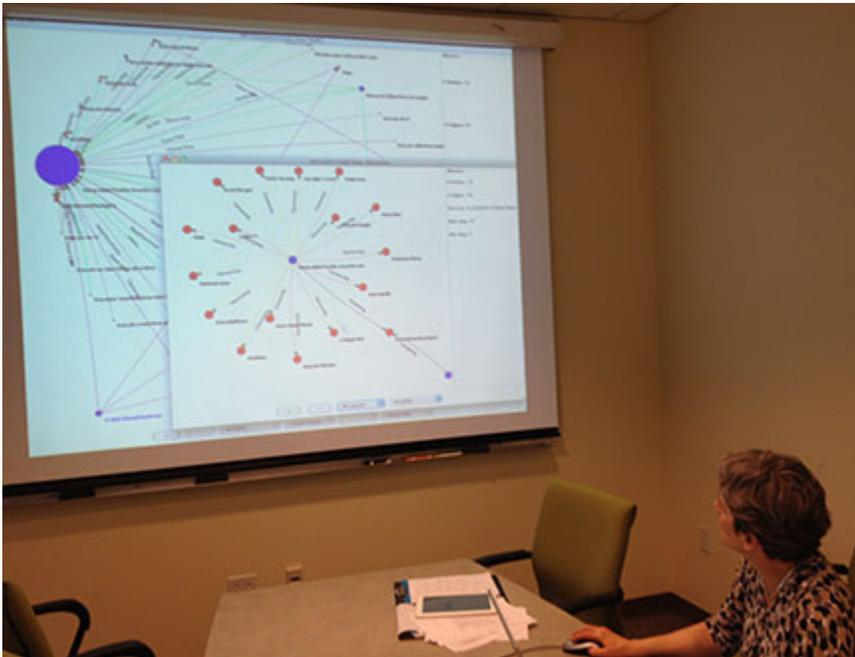
September: another demo with increased functionalities

**Images from the demo:**

Davide demonstrates how requests for different relationships produce different graphs. He implemented these various search capabilities on the fly, in the course of discussion about the utility of the overall graph and the need for BPS to provide "value added" to the humanities researcher. This is a clear indication of what the SNA graphs can easily demonstrate and how it will aid the humanities researcher.



Laurie explores the graphs.



August 6, 2012

## Meeting Agenda:

- Admin paper with Davide
- confirm room assignment and equipment
- Review Davide's project charter, prepare revised list of milestones: (postponed until Aug 14)
  - end August
  - mid/end Sept
  - mid/end October (end of Davide internship)
- Preparation for August 8 demo. (to Niek)
  - fill in outline (Aug 8 demo: <https://docs.google.com/document/d/1DBM5bjrNty1ic8uaQLtpGhZpAc66aklG2A1-78Rdfdg/edit#heading=h.1y4dwajeg293>)
  - review focus points
  - view preliminary presentation (Davide designing in prez)
    - SNA metrics
    - graphML
    - viz
  - selected corpus:
    - rationale
    - expected analysis/output
  - priority list for Patrick

## Action points:

- for August 8:
  - Davide develop graphML to show additional relationships
  - review googledoc for topic for each team-member's presentation
- for August 14:
  - review demo
  - establish milestones
  - taking stock at mid-internship

\*May 21, 2012|\*

Meeting:

Agenda:

\*Davide: settling in: reimbursements

Update:

Davide about ready to run stacks locally

PLS finishing up clean up. Better handling for names: TEI modelling for names, with name glossary at beginning. Was only using normalized names, so now changing processing. Concern if TEI from other projects is in different formats.

Flagging errors: willing to absorb certain PN errors.

Logging primitive: would like more standard log output, wants parse log specifically (for rebuilt TEI). Would like to see publication data incorporated into output.

Small things to work on:

--- marriages: persons not all being handled. 2 nested persons. pre-process

---need marking up of scribes: pre-processing

---kings

---right edge: PNs are not witnesses

LEP: how to treat ina aszabi

Follow-up tasks:

LEP ask STinney about roles in ATF if outputted to TEI

LEP: basic rules wives, false witnesses, scribes, king's names, neighbors + P nos. where they occur: child page added following meeting with this information

need to build simple graphml for DS to work on

PLS will work on specific use cases with DS.

DS and LEP to collect views. Send DS Caroline's workshop paper. To meet for initial round of discussions this week.

\*May 14, 2012\*

Meeting May 14:

Agenda points:

\*calendar review

\*Davide access to corpus TEI

\*What to assign student worker (C Bravo): clean data for Patrick, making notes of odd syntactic problems

Action points:

marriage needs to be documented as "gamonym"

roles such as relationships "wife", "his brother"

PLS and DS will process TEI in advance to deal with certain patterns: nested patterns; Not a temporary solution, because different projects will introduce own idiosyncrasies that have to be dealt with. Corpus specific processing in the initial layer of the architecture. Will do this in XSLT.

PLS will add DS to wikii

List of fiddly things to fix: PLS and LEP to go through notes and compile into wiki or project-tracking software

consider adding feedback mechanism per document within workspace of BPS

problem of problems list being on wiki run by berkeley runs into problem with other projects' access

LEP: harder measures of what questions are asked, what are must/should/nice haves, how does vizualization have to work (how do family trees present and what do they show)

Process for raising and resolving issues in terms of assessment of tools. Set up documentation for tracking this process. Combination of high-level loose specs and then user stories. DS will look into the possibility of an instance tracking program. Look at google.docs. DS likes pivotal tracker---good for back-tracing etc. Looking eventually to move to github.

LEP: pull out software from NEH seminar in LA

\*May 4, 2012\*

Davide Semenzin arrived week of May 4. Meetings with Patrick: review development environments, background materials for those environments, introductions to IST staff and to Warren Hall, set up work space. Meeting with Laurie: review text processing procedures, against error logs produced by Patrick.

\*February 14, 2012\*

Third meeting of Patrick, Laurie and Davide Semenzin, an MA student from Utrecht University (NL), who will spend six months (May - October 2012) at Berkeley working in the development of the SNA reasoner and graph viz of BPS for his master's project. Davide is a student of Professor Sjaak Brinkkemper and Marco Spruit. Following a series of emails and a Skype conference between Davide and Patrick in Fall 2012, the BPS team agreed that the prolonged period of work that Davide is able to devote to development for BPS and his interest in the SNA and viz components mesh well our current development needs.

Davide met with Patrick and Laurie in order to familiarize himself with the humanities component as well as the conceptual framework of the BPS project and architecture. In three meetings, Davide drafted the initial Project Charter statement. On his return to the NL, he will meet with his advisers and refine the project statement in conjunction with their guidance and in consultation with Patrick and Laurie.

First milestone will be the submission of Project Plan/Charter on April 1.

\*August 30, 2011\*

Laurie and Patrick review prioritized list from Spring workshop. There are a number of low-cost-in-time /resources items that Patrick will be able to develop. Laurie will provide the information and data necessary.

Before addressing the priorities list: Patrick mentioned that HTML5 is getting a lot of discussion and its graph capabilities may be useful for us. He will continue to investigate. Agreed that viz is important to develop soon.

## Topics:

1. Roles: Laurie will provide templates for the various types of texts with indications of characteristic language that enables identification of individual roles and the nature of the activity documented in the text.
2. Broken/Damaged Names: require template for name morphology. Possible NLP area
3. robust IDs: referring to the IDs of PNs in the TEI. Anytime the TEI in an Oracc project changes, the signature changes. For the most part, this should not be a problem as the persons and line order don't change that much. However, for projects not using TEI generated through Oracc, and in certain situations in Oracc TEI, the name instances in a given line may change. PLS is thinking about adopting robust linking, using structures and patterns. This would accommodate situations where, if the number of name citations were to change in a particular line of a document, the robust linking would recognize that there are no longer 3 instances of that name citation and would flag the entry.
4. Place names: one of the priorities for workshop participants was the ability to identify places other than the transaction-location. The easy case is when another place name is mentioned in the text: the ATF lemmatizes for that and can be identified in the TEI. The difficult case is where places are alluded to by the presence of people known to have come from some place other than where the transaction occurred. As this is a high-cost project, it is necessary to determine whether this will produce something other than information that is nice to know.
5. Dates:
  1. conversion of Babylonian dates to Julian dates: requires tables from Parker and Dubberstein (the standard reference on this) and search for existing conversion program. There are a number of constraints on intercalation and disruptions in the regnal sequences that have to be accounted for. LEP will look for researchers who may have developed tools for integrating these chronological problems into the conversion algorithms
  2. display Babylonian format dates: this requires pulling the metadata from cdli catalogue

Graphviz: This is an area for which we are going to look for a development partner. Laurie will contact Tim Tangherlini (UCLA) for leads to presenters at the Aug. 2010 NEH digital workshop (in LA) who might be interested in developing Graphviz in conjunction with our probabilistic/assertion-based architecture.

Began identifying projects that would be appropriate for student researchers:

\*NLP to identify roles. Included here is the problem of accommodating database input from researchers and how to convert db into TEI, as an initial (or separate) phase, and then run the TEI into the BPS tools. This recognizes that db would produce sparse TEI. Problems with db conversion into TEI would be identification/correction of errors in pattern matching for roles.

\*NLP to identify broken/missing names: requires morphology of names, name authority index and output of existing names

\*May 3, 2011\*

Outcomes from tally of priorities from workshop of April 8-9: ([pdf](#) of worksheet)

Priority will be given to those items that contribute to establishing basic functionality of all components of the system. Low-hanging fruit (those items that are low-cost in terms of development time) will be done, regardless of how central they are to basic tools functioning (these are items that contribute to usability and interface that may or may not be critical to tool operation).

The model for the tools starts at the corpus level. The data from that corpus will be directed in two ways: one to the problem of disambiguation (on the machine level) and the other to the matter of assertions (and probabilistic modeling), central to the humanists' interaction with the data. Achieving disambiguation and some functionality with assertions will lead to graph and SNA output. Preference will be given to those tasks that most directly lead to the implementation of a functional, if basic, set of tools. The features identified by workshop participants as having highest priority/interest will be added on top of basic functionality and UI development.

\*April 15, 2011\*

Task List and Wish List resulting from workshop of April 8-9:

Item related to UI:

links desirable for each document

- link to Oracc for corpus and for doc (done 4-15-11)
- link to Oracc for document, photo, line drawing (done 4-15-11)

Item related to Reasoner process:

- Allon would like to see explanation of how get to results, listing rules applied to document process leading to result.

Names and Names Authority:

- Disambiguation as process of corpus curation?: For some researchers, hand disambiguation as assemble a corpus. Yoram and Laurie this is not the case for larger corpora. There are also issues between and among different corpora.

\*Short names (nicknames) and alternate ("other") names:

1. Patterns of orthographic variants of a name
2. Common nicknames
3. Individuals known by two names

\*Damaged and missing names:

Yoram wants help with working with these. Patrick said help with missing name problem depends on presence of some qualifications on the citation.

Damaged names: Patrick introduced workshop participants to Levenshtein distances. By computing these, may be able to consider possible matches. Need to understand the units of Levenshtein (probably not letters in romanized transliteration--PLS) and conventions of marking damaged morphemes in Oracc.

\*Name Authority support:

In forename refs., value of n attribute is the normalized name, whereas nymRef points to name orthography. Desirable to pull variants from nymList and put into document listing. Move normalized name to the first column and put "name orthography" in second column

\*Growing consensus names are global.

\*Chronology:

1. Allow years to be a timespan, and then make the nrad's derive from those. Note that the computation of earliest and latest must derive from the basis-timespan. This lets the father range slide around more for a given citation of a son, and would let all slide around more for a document with uncertain date.

\*Corpus management in Workspace:

1. Must be able to refresh the corpus, essentially re-importing the one we started from
2. Be able to just dump the corpus, to get back to the initial state. This is setCorpus(null), and make sure it works.
3. Looks like there may be concurrency problems. Need to figure out how to synchronize the service context caches, so we do not step on one another. Main thing should be the maps. Need to think about the others. Consider allowing single session on each workspace, and on each corpus - i.e., single write-access session.

### Tasks

Laurie: review treatment of damaged morphemes and rendering of broken names in Oracc---requires review of current standards and consultation with Oracc Steering Committee (Veldhuis, Tinney, Robson)

\*May 13, 2010\*

TEI of mini-corpus generated for tools testing. Mini-corpus of 20 texts selected for known family ties and generational continuity. All texts belong to family group published by L.T. Doty in Journal of Cuneiform Studies, "The Nana-iddin family".

\*February 4, 2010\*

### Agenda

- planning sessions and equipment needs for late March conference
- prime the pump for conference participants: what to prepare, what to bring
- review PN markings, prior to finalizing mark-up for broken names and disambiguation
- examine TEI at <http://cdl.museum.upenn.edu/hbtin/downloads/hbtin-teiCorpus.zip>

### Outcomes:

- LEP to review TEI with STinney: is TEI disregarding lemming for distinction between meanings of A (son vs. descendant)
- LEP to define tasks for conference participants:
  1. text corpus preparation
  2. documentation to review prior to arrival
  3. identify research questions
- PLS
  1. tool development: focusing on identifying/extracting relationships and family-tree building

## **January 20, 2010**

### Agenda:

- looking toward the prosopography conference late March
- update on lemming process
- update on tool development

## **December 8, 2009**

### Agenda:

- HBTIN workflow
- mid-process evaluation and human readable text

- thinking ahead to other corpora (CTIJ)

incorporate CDL into BPS. most of NLP for HBTIN being done by lemmer. some of nlp and parameter material is corpus specific.

TEI review for names, or for roles, either before processing or as process files. flexibility of processing--- discard multiple clans, error reports.

lemmatizer interface not emacs: ?

LP: think about implication of multiple corpora, rules for lemmer for each, how we work with corpora. perhaps watch workspaces in close environments to particular corpora. separate feeds to keep contexts.

workspace: model raw corpus, lemming corpus, how build lemming. have to model the whole process, incorporate them. additions to tweaking

lemming process: how to lemm text, what's happening at deeper level. what's priority for deving? web interface.

**November 30, 2009:**

Agenda:

- team-development/management issues, based on demonstrator project experience

Outcomes:

- reviewed discussion from previous meeting re: relative weight of individual in transaction where there exist duplicates or two documents such as quitclaim and another doc. relating to that transaction: at this point the code models one or more activities per text. It does not model a single activity covered by two texts. This can be added in the future and does not need immediate attention.
- Laurie looked at the XML of the files that also exist in TEI and believes that the lemmatization process is responsible in part/in full for the multiple ancestors that are showing up. The logogram A can be read either as "son" or "descendant". Laurie will confer with STinney to come up with expedient solution to the problem, recognizing that she is not going to fully lemmatize all 99 texts Patrick has available quickly enough to allow Patrick to have logical readings.
- This problem underscores the need to have human readable output. Patrick is going to work-up a quick way to pull a listing of the P##, PNs, FNs, LNs that can be checked by the mortals for errors /inconsistencies.
- The difficulties in getting to understand where the TEI wasn't making sense (there shouldn't be three ancestors for an individual) suggests that we should be developing reasonable testing and evaluation routines for mid-process assessments.

Action Items:

- list of clan names (done); ask STinney constraint can be added to lemmer rules that if a name is not a clan name, the lemmer will mark the sense of the relationship-term (A, DUMU) as "son" rather than "descendant"

**November 24, 2009:**

Agenda:

- Review action points for month:
- XML/TEI valid markup
- unicode ATF

Outcomes:

- creation/addition of BPS feedback mailing list/form

**November 17, 2009:**

Agenda:

- Review template additions
- Mark-up for literal strings
- updated XML?
- HBTIN progress

Outcomes:

PLS worked on user management on the [Berkeley Prosopography Services](#) website:

- basic profile and contact support. One can now register, edit profile, and send feedback.
- user-administration support working, allowing admin to assign roles to people

Next steps:

- to administer corpora and support basic workspace functions for corpus (list stats for a corpus, etc.)  
First version will have manually loaded into DB. Later, automated loading from an admin page.

Action items:

- list ATF markup significance of: square brackets, #, ... , etc.
- contact STinney: how intends to handle TEI markup for damage. PLS: how to interpret and add filter if ST not dealing with it. (done)
- contact STinney for legal XML for features now existing (done)
- start task list between PLS and ST: use wiki page--set up and notify by e-mail. (done)
- emacs: consider alternatives
- order of templates:
  - witness, dates (done)
  - activity: prebend (one type done), quitclaim
  - roles
- think about how identify quitclaim and sales associated. collapse the activity in them?---weight of activity, would seem to be doubled. different principals/witnesses. one activity, two docs.; cut weight in half/reunite to one if all principals same. weight on duplicates and triplicates.

**November 10, 2009:**

Agenda:

- Review action points from Nov. 3 meeting
- HART final report
- Templates for roles and activities.

Based on the [witness template](#) and text structure (clearly defined sections of main text and witness list):

1. rather than default role being unknown, probably more efficient to set default as principal or witness.
2. as roles are refined (e.g., seller or buyer), can make more specific

LEP:

cuneiformists would normally understand "principal" to be the buyer or seller. But there are other roles that will get labeled as "principal" by default that most cuneiformists would not consider to be "principals". These include guarantors, clearers of claims, neighbors of property being leased or sold. However, this should not present a problem as long as corpus curators and users understand that this is a default distinction, that more specific roles will be added to the parses and/or that roles that require too much fiddling to label automatically can be more specifically designated by hand.

Task: develop simple taxonomy of roles

Principal Witness

buyer

seller: (probably not useful to submark)

prebend seller

other kinds

generic

To designate the scribe: encapsulate UMBISAG in entire person marker. This will account for any variation in placement of optional professional designation that appear in fully designated name. There are no multiple scribes, so are dealing with a single scribe design.

Dates:

build table for chronology

month names and lengths

Parker & Dubberstein: OCR?

king's names/date ranges

Arsacid period dates

Parthian period dates

Action Items:

Laurie:

- finish up HART report (done)
- request Steve Tinney to update HBTIN XML (incrementally) (done)
- build table for chronology (Parker Dubberstein OCR?, send to PLS)
- list of month names (and lengths?)
- year ranges for kings (almost done)

Patrick:

- fetch and incorporate updated XML
- review user stories
- consider taxonomy for role designation: defaults of principals and witnesses (cf. witness template)
- incorporate witness template information into role-markup
- harmonize textmarkup on date formula on witness template page and dates template page

**November 3, 2009:**

Agenda:

Review PLS and LEP wiki-pages added for project development.

Outcomes:

PLS most pressing need: latest XML from Steve Tinney. (still dealing with problem of text presenting multiple ancestors for single individual)

Laurie to consider:

1. how text corpus clean-up impacts fundamental assumptions: pay attention to those features
2. the development of family tree user stories for the graph builder, keeping in mind:*a. steps necessary for the most primitive construction to what the final product might look likeb. how much weight to place on user needs for visualization vs. interactivity*
3. templates for markers for roles

Action points:

1. LP to contact Tinney (UPenn) for: (done 11-9-2009)
  1. latest XML
  2. frequency of 0-version updates: changes in lemmer to fix bugs, changes in data for output
  3. ideas on model for delivering new versions of corpus:# 0 zipped xml to PLS
  4. 1 updated version available at known stable URL, service fetches it
  5. 2 hit API, making this process more hands-off
    2. LP to develop templates for activity texts
    3. LP to develop rules for geographic names: co-occurrences and rules for identification
4. PLS: identify dependencies for user/service/ui stories. develop better task list.

Next meeting: November 10.