***Initial DeepPhe Interview Protocol***

***11 September 2014***

A contextual inquiry interview is based on the premise that observation of the work *in situ* is the best means of understanding a domain professional’s workflow and subsequently building tools to support it. During the interview, we will ask the informant to conduct relevant tasks, explaining *what* is being done, *why* it must be done, and *how* it is done. As the informant continues through the task, the interviewer(s) will ask for clarification, pose questions about contextual factors, and generally attempt to build an understanding of the work. Therefore, much of the discussion will be determined by the events that occur as the informant completes their tasks.

Below, we give some broad structure for the interview, followed by a list of questions/concerns that should be addressed. However, no ordering of these questions is implied - rather, they can be asked when convenient during the course of the discussion.

**Recording**

All sessions should be audio-recorded, either via web conferencing software or portable digital audio recorder. Written notes are fine, particularly if multiple team members are present during the interview, but recordings are often helpful. If interviews are conducted via web-conference, screen captures might be recorded as well. Artifacts such as data files might also be useful, when available.

*Note:* Avoidance of PHI is vital. Participants should probably not be sharing any data that contains PHI. If any displays or artifacts involving PHI are inadvertently introduced into the discussion, any screen capture or recording modes that might capture the PHI should be disabled as soon as possible. Alternatively, the PHI data might be removed entirely from consideration.

**Interview Introduction**

The introductory component of the discussion is designed to provide basic context and to start conversation:

*Introductory script:*

Hello. My name is XXXX. As you know, you’re here today to participate in a contextual inquiry study to understand the requirements for the visual analytics tools we will be developing. These tools will help researchers explore and interpret phenotype data that will be extracted with our software. We would like to understand users’ information and workflow needs to maximize the usefulness of the software.

During the session, you will be questioned and observed. We will ask some general questions about your work, and more detailed questions about specific tasks. Data will be collected in the form of notes and artifacts. With your permission, the session will also be audio-recorded. The session will last approximately one hour.

In this session, we would like to explore in detail how you currently identify, store, and manipulate specific phenotype data and link this to genotype data in your work. We’d like to understand how this work is done, any challenges you face, and what you would hope for in a new tool to make work more efficient and easier to accomplish. If possible, we’d like you to demonstrate the tools that you currently use on some meaningful data. We’d like you to identify a representative dataset and task that you are currently working on or have worked on previously to demonstrate how you currently do things.

Do you have any questions?

Before we begin, do we have your permission to record this session? (IF YES, BEGIN RECORDING)

To start off, we’d like to get a sense of your background and research.

1. Respondent background:
   1. What is your training in? What degrees have you obtained, and in what areas?
   2. What is your current position? How long have you been at this position?
   3. How much of your time do you spend on cancer data analytics?
   4. How much of your time is spent on clinical work?
2. Problem description:
   1. What questions are you currently trying to answer related to cancer analytics?
   2. What sort of data are you dealing with?
   3. Is this a research or clinical effort?
   4. Who is this effort supported by?
3. Data:
   1. Which tools do you currently work with for your data?
   2. What are the major data challenges that you struggle with?
4. Major non-data challenges
5. What are the major challenges that you struggle with that are not data related?

**Scope of work**

*Script:*

Now we’d like to understand the scope of your work. We’re going to describe some different groups of people, and we’d like to understand which group or groups you identify with.

* NLP Developers: taking stated researcher information needs and developing tool for extracting those details from clinical text.
* Domain specific application developers: taking output from NLP and presenting it in ways that meet the needs of end-users.
* Information Brokers: translating between the stated needs of cancer scientists and the representations and vocabularies used by NLP developers.
* Informatics Researchers: Working with end-users, NLP developers, and domain specific application specific to interrogate data, explore new representations, or otherwise extract more utility out of the data.
* Population health scientist: looking at data in the aggregate, trying to draw out large scale patterns across many patients.
* Integrative cancer biologists and modelers – trying to develop causal stories explaining disease progression.
* Clinical translational scientist – linking care and interventions to outcomes
* “Dry Bench” translational scientist – applying algorithms and tools to understand characteristics of disease.

Which of these groups do you most identify with?

**Task identification**

*Script:*

Now we’d like you to identify a specific dataset and tasks that fit within the scope of the work. This should be representative of your work and illustrative of difficulties you encounter. Please choose something that does not involve sensitive data.

Have you selected a task or tasks? Please explain the task.

Note: If participants identify multiple important tasks, they should be prioritized so that the session will start with the highest priority and continue from there as time allows.

**Observation**

*Script:*

Now I’d like you to work through the task. I’ll be observing you and asking questions as you work to gain an understanding of exactly what you’re doing and why you’re doing it.

Questions to be asked at any point during the interview, including a few that are focused on specific stakeholder groups. (Some answers may become apparent based on respondent comments)

* + Which statistical methods (if any) do you use?
  + Which file formats are used? Which file formats are preferred?
  + How do you record findings or insights?
  + How do you form hypotheses?
  + Given data that might support a hypothesis, how is other data handled if it is not consistent with the hypothesis?
  + Are there any attempts made to explicitly rule out alternative hypotheses?
  + Which data tools/repositories are used?
  + How is EHR clinical text managed alongside genomic information and other supporting data?
  + When, why, and how do you make use of the research literature to inform understanding?
  + How do you build accurate models of phenotypes, given complexity of clinical text, particularly with respect to temporal ordering?
    1. How do you validate these models?
    2. How do you represent uncertainty in your models?
  + How do you determine that patients or treatments are “similar” to each other? How confident are you in this determination?
  + Which standardized vocabulary/vocabularies do you use?
  + How do you communicate results to peers? What output formats are needed in your work?
  + How do current tools succeed? How do they fail? What are current tools missing?
  + For Information Brokers: How do you translate between needs of researchers and worlds of NLP and/or basic clinical records?
  + For NLP developers: How do you evaluate success?

Note: During the session, the interviewer might be tempted to make suggestions regarding potential designs or improvements to tools and/or workflow. Such suggestions can be useful, but should be presented carefully, so as to avoid any implicit promises.

*Repeat this process for additional tasks as time allows.*

**Debrief**

The last part of the interview will be used to review general impressions, ask about observations, and to confirm understanding of what was seen during the observations.

1. Ask about general patterns that were observed.
2. Ask for clarification of any inconsistent patterns that may have been observed.
3. Attempt to draw the respondent out - try to get them to reflect on their work.
4. Give the respondent an opportunity to raise other concerns and to ask questions about the work.

*Sample Script:*

It seems that you typically XXXXX while working. And that you struggle with XXXXX. Is that accurate?

Are there any challenges that you have previously encountered that weren’t observed today?

What would you like to see in a new tool?

Are there any other comments you would like to make?

Thank you very much for your time. We really appreciate your input.

**After the Interview**

Whenever possible, try to schedule time for review of observations immediately (or as soon as possible) after the interview session. This review time can be useful for recording any observations that might not have been noted during the interview. It is important that this session happen as early as possible, as comments that made a strong impression might fade within a small number of hours.

Subsequent analysis will involve reviewing notes and recordings. Recordings should be reviewed to identify informative comments and sequences of operations. Noting time stamps and summarizing incidents of interest is encouraged as a means of extracting value from recordings without incurring the cost of transcription. Review of notes will inform the development of workflow, cultural, and sequence models that will eventually inform design.