**Innovations required for the use of clinical data in genomic research.**

In order for hitherto unused clinical genomic data from patient records to be leveraged, the ILDB workgroup will need to provide solutions to the following list of key issues:

1. **Consent standards**

*[Sub-group: Snehit Prabhu. Kelly Ormond? Heidi Rehm? Sharon Plon?]*

*[Snehit: This is not my area of expertise so please add content if you are qualified. Many aspects relating to meaningful consent, patient education and engagement, and legal aspects like HIPAA protections need to be addressed systematically before the ILDB can go online].*

* 1. For legacy samples, is re-contact with patient required before use in ClinGen ILDB? If not, as a guideline to participating labs, what type of language, or specific clauses, constitute adequate consent?
  2. Going forward, what type of consent will participating labs need to be request from their patients?
  3. What privacy-related risks should be communicated to the patient?
  4. What potential research-related benefits should be communicated to the patient?
  5. Should consent requested be dichotomous (yes/no for blanket use of patient’s data), or more nuanced and multi-category? If so, how will data with partial or conditional consent be stored and used in the ILDB?
  6. Other …

**GA4GH high-level framework:**

1. Respect for individuals
2. Advance research
3. Promote health and well being
4. Foster trust, integrity and reciprocity.

**Implementation policies to operationalize framework principles**

Workgroup Oct 2014 – 7 teleconferences and drafts, inputs, created **policy paper**.

Benchmarks (maturity model), best practices, mutual trust model (1) how to manage expectation of privacy, (2) manage risks, (3) manage inappropriate access, (4). Distinguish privacy and security. Also, a **consent policy** workgroup started in Mar 2014 **–** for use of data consented to by donors or representatives. Does not focus on returning health related findings. How to get consent, what constitutes consent, how to audit consent. Concept of “data-subject”. 2 sources of data: clinical trials (pharma) and diagnostic testing laboratories. Data discovery vs. Data sharing.

Matchmaker Exchange has its own tiered consent policy.

**Data models and IT infrastructure**

*[Sub-group: Snehit Prabhu. Larry Babb? Sandy Aronson?]*

*[Snehit: For ILDB version 1, a lot of this will just be a description of the relevant pieces of GeneInsight’s software architecture.]*

* 1. Which fields of information and data types from patient records are to be captured? Which fields will be mandatory vs. optional?
  2. What controlled vocabularies will be used for each field of entry?
  3. What will the database schema (relational or document model) be?
  4. What web-based interfaces will be made available to (i) data uploaders, (ii) data administrators, and (iii) data users?
  5. How will we authenticate each type of user?
  6. What type of IT infrastructure will be used, and how will it be accessed? What IT security provisions will be made?
  7. What redundancy and data-backup provisions will be made in case of system failure/crash?
  8. What software production and maintenance protocols will be in used?
  9. Other …

1. **Data submission and QC protocols**

*[Sub-group: Snehit Prabhu. Sam Baxter? Samuel Aronson? Larry Babb?]*

*[Snehit: From my discussions with Sandy Aronson and Sam Baxter, it appears that we will require an admin team for day-to-day logistics. Besides that, I anticipate that we will also need an executive committee to make higher-level decisions, around topics like those listed below]*

* 1. For large bulk uploads from participating laboratories, their data + consent types will need to be reviewed by an ILDB committee before incorporation of such data into the ILDB.
     1. Who will be in the committee (e.g. patient advocacy, legal, technology experts, etc.)?
     2. What will be its guiding principles?
     3. What will the guidelines for providing/denying such approval be?
  2. Alternately, are we either going to
     1. Externalize responsibility to the data-uploader through a legal disclaimer, or
     2. Programmatically enforce barriers against uploading sensitive data fields by mistake?
  3. What is the list of QC measures to be put in place for maintaining data integrity and consistency across uploads from multiple labs?
  4. Who will resolve any QC related conflicts that arise during upload and what procedure will be followed to resolve them (versioning protocol, etc.)?
  5. Other …

1. **Patient privacy and security frameworks**

*[Sub-group: Snehit Prabhu. Carlos Bustamante? Sam Baxter?]*

*[Snehit: This section partially deals with pragmatic/logistical issues relating to exposing clinical data fields, as well as an open-ended research component to develop privacy enhancing mechanisms for use of sensitive data in research settings]*

* 1. What are the different risks associated with sharing patient health data? List, categorize, describe and define these.
  2. Which data fields from their patient records will labs upload to the ILDB?
  3. What privacy framework will be used to judge the risk associated with sharing a particular field?
  4. Where will the committee’s judgment about each field be recorded? How will it be justified and communicated?
  5. If fields are judged to have differing levels of privacy related risks, how will this determine their usage in various types of research?
  6. What kinds of read/write/download privileges will be assigned to different categories of users by the ILDB?
  7. What (if any) are the quantitative privacy/utility trade-offs associated with each data field in the ILDB?
  8. Other …

1. **Data usage guidelines and oversight/regulation of users**

*[Sub-group: Snehit Prabhu. Sharon Plon? Heidi Rehm?]*

*[Snehit: This section deals with the criteria for and logistics of approving a data-access request. A decision has been made to provide controlled access to known entities only. There will be NO anonymous/public access for now.]*

* 1. What kinds of entities may request access to and be granted access to the ILDB?
  2. What does a *bona fide* data-access request look like? What materials must the access-requesting entity provide (web-form/write-up) to establish validity of research agenda?
  3. Who will adjudicate over each request (ILDB committee, patient advocacy group, data contributor, etc.)
  4. What criteria of the proposed research agenda will be used to establish whether a request for access warrants approval or denial (e.g. requester has to provide power studies to demonstrate the utility of additional case records, etc.)?
  5. Once access is granted to a subset of data (e.g. Noonan’s syndrome cases), how will we limit the set of queries available to the user?
  6. What mechanisms will be used to enforce the terms of data usage?
  7. What mechanisms for attribution and reporting (to the ILDB and the data sources) will be established?
  8. Other …