2020 GA4GH Connect Virtual Meeting

Pedigree Breakout Agenda

Details subject to change.

Meeting Goals: Minimum Core Dataset

Relevant Work Streams: Clin/Pheno, LSG

Zoom Link: https://zoom.us/j/4426465151

Chairs: Orion Buske & Grant Wood

Notetaker: Lindsay Smith

Wednesday, March 25, 2020

Start Time	Discussion Topic	Related Materials	Speakers
20:00	Pedigree Update/Overview	Slides - Use Case Collection - Gap Spreadsheet - Minimal Dataset Doc	Grant & Orion
20:10	A Universal Use Case and the Minimum Core Dataset	Slides	Grant
20:20	How a Consent Engine for Family Health History Might Work	Slides	Grant
20:30	VCF and pedigrees	Slides	Luis & Yossi
20:40	Priorities and action items	Slides	Orion

Minutes: (transcript)

Overview

OB: Sent out a driver project survey to look at gaps. Nine driver projects responded, 23 responses. Has helped to make a case for this working group. Have use cases from driver projects and the wider community. Grant has been leading a minimum dataset effort. The next 6 months: have a draft to present at plenary at the end of september. Have the draft be implemented in the REDcap plugin, phenopackets. Test it out here before a larger interoperability pilot. Have a blog post out around that time. We need to have a draft data model spec in JSON or technical way that people can sink teeth into. Over the next two months, get a small group to dive into work and make design choices.

Q/ What about use cases, where does that fit in? We should identify the actions to do with the data, what questions to answer. Come up with some basic questions to ask of the data.
A/ That is now asked in the use case requirement collection document. Would be great if someone from the VCF team can contribute an use case into that document.

Universal Use Case & Minimum Core Dataset

GW: 5 years ago, had an ehealth task team focused on health history. Tools inventory in the GA4GH toolkit. From the gap analysis, some comments on Clin/Pheno - need to include companies, can't just focus on rare disease and cancer, data needs to be integrated with tools.
HL7 has a history developing family health history standards. We don't want to recreate the wheel, leverage other standards out there. HL7 created an IG, not a lot of implementation but hoping for more adoption. In 2007, defined the minimal core dataset that every family health history tool should collect. The contents of the document was published and can search in pubmed. Have revisited that work and are building a new table (link in agenda above). Result is to come up with the elements and description. Trying to map to what is currently in ped and FHIR.
This morning with the gap analysis, one question was about the organization you worked for. Majority was from academic research and only 4 from healthcare. This diagram tries to capture the complexity of what happens. There are different stages of the workflow, data capture in different applications. Idea is that from that data we can generate new knowledge and put into database, use that data in clinical decisions. Want to highlight that family history is part of a complex workflow.
Share the idea of universal FHH use case. We have the EMR, outside of the EMR we might have a repository with genomic and clinical data. There is a window for patient-facing applications. Describe different types of family health history apps on the left side. List research, genealogy, etc. The idea of trying to show that when we do this pedigree minimum core dataset, if we put FHH in the center, and the different stakeholders shown around. The min core dataset sets a bar that fits for all of these different stakeholders. As we do this work, need to think in terms of all of the different stakeholders.

How a Consent Engine Might Work

GW: Assumptions - create a full pedigree, coded disease, coded AOD, researchers as part of study or trial, available to healthcare consumers and providers, researchers. Shown is a screenshot of a tool. If we put a consent menu bar: If I am an user, I want to decide how that data is shared. Can select only certain conditions to be shared. Can select whether I want to share with certain relatives.

Pedigree in VCF

LB: VCF is a way of storing variant calls from genomic data from individuals or groups. Added the ability to write ped file directly in the header. Can represent trios as child, mother, father. Can define family trees from this. Can also represent asexual relationship like clonal cell lines or tumors. We were missing the ability to describe twins and triplets. Also working out issues on how to describe family trees when there are unknown members of the family - ie. no IDs. How to make it clear what ID to use. Things that are not included: no good way to describe complex consanguinity, no phenotype information (but can include in file), provenance information on how relationships were defined, no way to capture uncertainty. Not a ton of people use the pedigree line as far as we know (don't get feature requests). Would be useful to hear if people are interested in putting more information in, or whether we should store separately.
OB: Great use case to consider.
JJ: With regards to last bit about preferences to separate or not. I think it makes sense to delegate specialist formats. For EXOMISER, use vcf but ped for family history information. From the phenopackest perspective, have a balancing act. Have some representation of variant information but don't want to represent the entire VCF. Point to a VCF file.
LB: If no one adopts the pedigree information, then it ends up stuck there. WHen you have simple information, we have VCF that is 100 and most are trios, people just want to know what the trios are.

Priorities/Next Steps

OB: Distill down what the highest priorities seem to be. What I want to pose to the group: should any of the items shown here not be a priority? Twins, pregnancies/miscarriages, adoption, social relationships, donors, infertility, transgender. All of these things have standardized visual representations in pedigrees, but not in data models.

YF: Infertility seems strange. It sounds like a phenotype, or medical condition possibly unrelated to relationships between people involved.
OB: Can infertility in a person or as a non-productive relationship.
ZS: Maybe representing a relationship that did not produce children.
DT: Can have three parents, can have mitochondrial donor, biological donor. Can have 5 or 6 people involved. Could affect maternal-fetal relationships.
OB: Defining the relationship as non-productive because infertile. Property of relationship rather than the person, but I can understand why it is seen as a phenotype. Can make this more explicit.
PR: Hard to standardize that. Worried that might introduce noise into data.
OB: How common is it to record something like that?
DT: In fertility treatments, if going on to a pediatrician, family usually gives the full history. Not common, but will show up.
JJ: The infertility part is hard to define, but in sense you don't need to. Just recording the relationships.
JH: Not a priority.
OB: Priorities, but could be postponed - provenance, visibility and consents, ambiguity.
LB: Would argue that provenance would be important for tooling.
OB: Initial draft for data model

NS (Kids First), JH, JJ (GEL), AGHA in for pedigree,