<div align="center">

2020 GA4GH Connect Virtual Meeting

FHIR Implementation Guide Breakout Agenda

Details subject to change.

</div>

Meeting Goals: Discuss Phenopackets-FHIR mappings

Relevant Work Streams: Clin/Pheno

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

Chair: Moni Munoz-Torres

Notetaker: Lindsay Smith

Wednesday, March 25, 2020

Start Time	Discussion Topic	Related Materials	Speakers
20:55	GA4GH Housekeeping Introduction to breakout	Slides	Lindsay Smith (LS) Moni Munoz-Torres (MMT)
21:00	Phenopackets on FHIR with Kids First data	Slides	Natasha Singh (NS) Allison Heath (AH)
21:15	Discussing a Public Test Server		All
21:25	Phenopackets, RDF and FHIR	Slides	Harold Solbrig (HS)
21:40	FHIR Mappings	- Mappings Spreadsheet - Github Space - Draft IG	Moni, Peter, Alejandro, Jules
21:50	End	--	--

Minutes: (transcript)

Introduction

- MMT: To learn about Phenopackets, visit phenopackets.org The focus of this subgroup is to generate a FHIR Implementation Guide (IG) based on the Phenopackets standard. An IG is a set of rules about how FHIR resources are used (or should be used) to solve a particular problem, with associated documentation to support and clarify the usage. The Phenopackets-FHIR IG will increase the availability of high-quality standardized phenotypic information for genomic research and genomic medicine.

Phenopackets on FHIR with Kids First Data

- AH: Have been testing out the implementation guide that has been set up, part of that has been setting up an entire infrastructure for FHIR. Have a lot of feedback in terms of phenopackets and more so in terms of overlap between phenopackets and FHIR.
- NS: What we've done over the last few months is set up a GIthub repo to contain our FHIR model. Developed tools to validate pieces of the model. Has to pass validation. Deployed dev server. Loaded all resources that are part of FHIR phenopackets onto our server. Ingested two datasets into the server. Have circled back to the beginning and looking at building a model from scratch. We began mapping data from the clean data warehouse. Once we have the spreadsheet, developed python modules that would produce the different phenopackets resource from a row of data in our warehouse. Couldn't do all of the different top level elements but could use most as is. Did make a modification to disease, because is was bound to SNOMED.
- We built the dashboard to see what the search API could do with the data. Wanted to start browsing the data in our server. Organized like how FHIR organizes.
- Can see we have 2600 disease resources loaded. Disease profile comes from phenopackets. Takes some time to load, has to search through the FHIR API.
- Can display attributes as a bar chart. New attribute called "term", bound MONDO and NCIt to it.
- AH: This is highlighting the FHIR complexity, and trying to layer phenopackets on gets tricky. Trying to make it useful has been something this dashboard highlights.
- NS: using phenopackets as a starting point was very useful. Could use a lot as it, but got to points where we didn't know how to populate the more complicated entities. Thinking about developing a model from scratch.
- AH: not fully from scratch, leveraging FHIR base models. Should we bring extensions in the FHIR model, or just say "this is where we're storing the data".
- NS: What's going to end up happening is go through the exercise developing a Kids First FHIR model, and be able to circle back to phenopackets with more feedback. Then start to collab on making changes. Developed our own FHIR 101 - up on Github for those that want to view.
- AM: When talked about extension needed for disease. Disease a profile based on condition, and it is not bound to SNOMED. Has an example binding, but not mandatory. Only would be necessary if the binding was required. For Phenopackets on FHIR, few bindings are required. Added something for code is not something I would encourage.
- NS: Resources that we did use in phenopackets we didn't have to change much.
- MMT: All of the work being done should be checked with members of FHIR working group(s). We need to make sure we are using the model and the language appropriately. Whichever route we choose for mapping, it must allow non-lossy exchange of clinical information between Phenopackets and FHIR structures. This point helps us to start our discussion about the public test server.

Discussion about a Public Test Server

- AH: This is practical experience, and the earlier narrative is that FHIR is an extensive platform. Once you get the data in there there's lots of interesting things. Want to think about a public test server with data people can use. Can have different versions of an IG. Are others interested and what would that look like?
  PR: Very interested. Need to go through each of the elements one-by-one and document the reasons for choosing them (ie. why 1 of the 5 evidence resources). Interesting to explore the extent of round trip. Great way to make unit tests and explore edge cases. Not going to be possible to translate FHIR messages because seom FHIR messages will have more detail. Might be a good idea to have an official library for this that would implement IG and provide documentation of what we expect to happen.

- NS: Don't get the full understanding until you do the modelling. If we can more people involved in end-to-end and looking at a FHIR model phenopackets and see how it applies to dataset. WIll find more feedback and modifications here. Hard until you get into loading data.
- GW: Is the data source for phenopackets always going to be an EMR?
- NS: One of the datasets are EMR based. Lots of data through redcap. Some coming out of a data warehouse.
- GW: If that is the case going forward, the strategy is to create phenopackets as a FHIR profile. Get it as part of major FHIR release, then in the future have FHIR vendors supporting the profile.
- AM: Long-term plan. Take the FHIR IG through the GA4GH process and the HL7 process. Also discuss how to do this as a collaborative effort.
- GW: Doing some pilots now would also be fantastic.

Phenopackets, FHIR and RDF

- HS: I use phenopackets as an example for working with JSON-LD and FHIR. Some of you are probably aware that FHIR has three representation forms. RDF problem is that spec is in text and takes quite a bit of programming to make a server. Hard to get uptake. Recently, the JSON-LD community have taken issues and created a new spec that gives us a mapping description to go from JSON to RDF and back. One of our experiments was to take phenopackets Individual and put a taxonomy on it. Can convert FHIR JSON into RDF automatically, also putting together JSON-LD contexts. Created a mapping from phenopackets individual to FHIR as well. I mapped them together and merged the RDF. Also tweaked the identifier of the individual. Got an aggregation of the phenopackets and FHIR patient information. One thing we get right out of the box are FHIR mapping points. Ie. got FHIR individual and patient. Able to put phenopackets namespace taxonomy in here. JSON-LD is bi-directional. Have to do a bit more work to get RDF back to JSON. Took phenopackets in native form, put together with FHIR. GO from FHIR patient, merge phenopackets, and when I get back I required the birthdate.
- JSON-LD playground: supply a context to native JSON. What phenopackets calls ID, I'm going to call the patient. DOB, I transfer into FHIR patient DOB and add a data type on it. On taxonomy I give it a subject. RDF Ntriples come out the bottom. What I'm doing is real-time editing and saying what the meaning of a little bit of JSON is in RDF. The RDF JSON-LD, RDF Turtle and RDF Ntriples are all identical. Gone from complete set of phenopackets to RDF and back out. I can use one context going in, and use FHIR going out (still working on this). Could put in url for context and start with set of things from phenopackets, and get out FHIR RDF. Done a fairly rapid round trip.
- Can take plain old JSON and add context - context are the easiest things to author to get into RDF. Key is that they can be completely separate. The JSON-LD tooling takes context as a parameter. In context, can add the semantics.
- MH: Impressed with this work as a way to validate the mappings and roundtripping. As far as terminology picker, have talked about modifying EBI lookup tool. That tool gives different display options about tree views.
- JJ: Q/ Can the JSON-LD context generate new elements? Ie. Phenopackets has a metadata element.
- HS: It's mostly no. The purpose of JSON-LD is assigning somatics or transforming existing structures to RDF. In framing language, have supplied ability to default things if not in RDF. Still only gives part of what you're asking for.
- JJ: stumbling block will be elements that don't map directly to FHIR. Do have a mapping, but not in same area.
- HS: My guess is that you're trying to put system field into coding entry.
- JJ: System is part of resource.
- HS: JSON-LD works great from URIs to CURIEs. If you have to add a field that isn't present, it needs to be a structural transform. Useful to separate the structural transformations from semantic transformations. Two-step process, but would be more clearly documented and easily implemented.

Phenopackets-FHIR Mappings

- MMT: How do we move this discussion in a concerted way?
- NS: If we could do a parallel effort to our team. Pick a test dataset, test server. Map test dataset to phenopackets. Going through an exercise would be useful.
- PR: One issue is that rest of world cannot share CHOP data. If we want to do this as a group, need fake but realistic data. One worry is that if the model is changed and if we don't get wind of it the models will diverge.
- NS: Main goal of starting from scratch is more for learning purposes.
- AH: How do we get data to do this. Some cancer data is open access out there. How to make sure we bring the feedback back as we remodel and readjust and stay in alignment.
- MH: Suggest to try and see if we can utilize the RDF approach and tooling. Have a small group that can work on going through mapping and adjusting those things. Might learn from recommendations to the FHIR community.
- AM: Main issue is that the experts from the FHIR side of things haven't been contributing. Regardless of tech, we need participation of the experts. Needs to verify that mappings are the correct ones.
- MH: Have Bob and Harold. Reality is that the representation from the interpretation packet doesn't exist in EHR. Need a process to validate content in existing way.
- AH: Not just mappings, but approaches to mappings. Been trying to engage FHIR experts but it's been going back to build from the base and engage people there.
- BW: What extensions could we agree upon at a basic level to describe that basic structure of a research project. See a lot on how the patient centered data, observations, treatment etc. are in great detail. Ideally anyone trying to implement in FHIR wouldn't have to do it themselves.
- AH: Agreed.
- JJ: Phenopacket biosample sounds like something you're talking about.
- BW: Will take a look