Global Alliance
for Genomics & Health

GA4GH Discovery Work Stream Meeting

Minutes and Actions 2023-2024

Table of Contents

Overview
Start-Up Guide to Discovery Work Stream and Discovery Products
Links to Archived Meeting Minutes
Links to Sub-Group Minutes and Actions Documents
Meeting Protocols
List of Meetings
- 2024-04-10
- 2024-03-13
- 2024-02-14
- 2024-01-10
- 2023-12-13 - CANCELLED
- 2023-11-15
- 2023-10-11
- 2023-09-13
- 2023-08-09
- 2023-07-12
- 2023-06-14 - CANCELLED
- 2023-05-10
- 2023-04-12
- 2023-03-08
- 2023-02-08
- 2023-01-10

Overview

These are the minutes and actions for the Discovery Working subgroup.

The GA4GH Discovery Work Stream builds and coordinates standards for federated, secured networks of data and services, to form an “Internet of Genomics” for enabling data discovery and retrieval in health research and clinical genomics.

The Discovery Work Stream is led by Nara Sobreira and Michael Baudis. For details on how this Work Stream operates please read the Discovery Work Stream Organizational Structure & Vision document.

Global Alliance
for Genomics & Health

This group meets at a high-level monthly. In addition, the sub-groups listed below meet on their own schedules. Participation in these groups is open but requires adherence to the GA4GH Standards for Professional Conduct.

Start-Up Guide to Discovery Work Stream and Discovery Products

Discovery
- Repository

Beacon
- Roadmap and information
- Specification
- Repository

Data Connect (formerly Search API)
- Repository
- Use cases

Networks (Service Registry/Service Info)
- Specification
- Repository

Links to Archived Meeting Minutes

1. Minutes and Actions 2017-2018
2. Minutes and Actions 2019-2020
3. Minutes and Actions 2021-2022

Links to Sub-Group Minutes and Actions Documents

This Work Stream has meetings for the following sub groups.

1. Beacon API - 2017 - 2018 - 2019 - 2020 - 2021
2. Discovery Search - 2018 - 2019 - 2020 - 2021
3. Discovery Networks - 2018 - 2019
4. Case Discovery - 2019

Global Alliance for Genomics & Health

Meeting Protocols

- Please note that by participating in meetings, attendees agree to adhere to the GA4GH Standards of Professional Conduct
- Meetings may be recorded for note-taking purposes. Recordings will be deleted within three months of the meeting taking place.

List of Meetings

https://www.ga4gh.org/what-we-do/calendar/

2024-05-06

Chair(s): Nara Sobreira, Michael Baudis
Attendees (Affiliation):
Apologies: Beatrice Amos

	Agenda Item	Person/Time

Link to meeting recording:

Meeting summary:

2024-03-13

Chair(s): Michael Baudis
Attendees (Affiliation): Maria Yazykova

	Agenda Item	Person/Time
	Discovery-related sessions at Connect	

Meeting summary:

Global Alliance for Genomics & Health

Meeting was cancelled due to limited attendance (maybe the usual DST shift gap …), but a fruitful discussion occurred: Maria & Michael chatted a bit about general development directions for Beacon filters & complex variant queries …

2024-02-14

Chair(s): Michael Baudis, Nara Sobreira
Attendees (Affiliation): Beatrice Amos, Jordi Rambla, Ian Fore (NIH), David Bujold, Arthur Ko

	Agenda Item	Person/Time
	• Welcome, Housekeeping	

	• Connect meeting in April - planning
	  ○ current session proposals?
	  ○ option: variant queries - Beacon & beyond
	  ○ option: future of service-info
	  ○ Beacon use - networks & aggregators	

Link to meeting recording:
https://us02web.zoom.us/rec/share/WhHcscNnHnOk6CkX8L8RBtfGUK95J7UU4qG5g3iiurv-YX8uO9Vija7NmNmf2UoV.htrdMXZtbDRph3yy

Meeting summary:
Pre-meeting:

- quite a bit

The group talked about the development and maintenance of Beacon networks, specifically about the transition from Beacon Version 1 to Version 2 and the implications for users and implementations. There are concerns about resource allocation and the need for continued support to ensure the functionality and relevance of Beacon networks, with emphasis on involving users and Driver Implementations. Jordi mentions ongoing projects, such as the development of Beacon Version 2 by ELIXIR and the need for coordination between different Beacon networks. Michael and Nara made a call for more active engagement from users to drive the development of aggregator services that can integrate data from multiple sources, including Beacon and Matchmaker Exchange nodes.

Global Alliance for Genomics & Health

2024-01-10

Chair(s): Michael Baudis, Nara Sobreira
Attendees (Affiliation): Beatrice Amos, Ian Fore (NIH), Jonathan Fuerth (DNAstack), Miro Cupak (DNAstack), Heather Creasy (NCI CRDC), Angela Page, David Bujold, Arthur Ko, Ian Fore, KathyReinold

Agenda

	Agenda Item	Person/Time
	• Welcome, Housekeeping: HAPPY NEW YEAR!!!! • Please introduce yourself in the chat • Connect meeting in April - planning (3 meetings remaining) • Upcoming steering committee meeting in January	
	Study group updates Working group updates	
	Follow-up from last meeting	
	Connect 2024 attendance, pre-planning	
	Discovery Strategic Roadmap 2022/2023	

Link to meeting recording:
https://us02web.zoom.us/rec/share/UtG4CnJQ2twTlR6Ol1RDLndkCUcKWoiPnzDepLgtH0dyKeu1cXS4D2hcUov6_fiD.FsCikuB609WpG_1L

Action items:

- possible Connect proposal: Session "Future of Service Info - Expansion & Adoption"
  - Miro attending? Other stakeholders?
  - too narrow/included in something else?
- Connect proposal: "*Genomic Variant Formats: From Cytobands and VCF to Beacon Queries*" <= better name?

Meeting summary:

Ian has suggested that the service info should be used more effectively. Michael and Ian talked about the ownership and usage of service info across different areas.

Global Alliance for Genomics & Health

Is TASC a suitable home for Service Info?
The group discusses the importance of Service Info as a foundational element across various areas and products within GA4GH.

Other Discovery WS sessions that overlap:

- 2 DaMaSC sessions - Schema Registry and either Style sheet for data model or Landscape. To be fleshed out in the DaMaSC session.
- Cohort building

There was a call to revitalise Service Info as a more comprehensive and inclusive tool that could cater to different use cases within GA4GH. Suggestions were made to hold a session at the upcoming meeting to brainstorm and shape its future without necessarily making it a TASC issue. Service Info had relevance across different work streams within GA4GH, maybe a session to bring interested groups together to ensure a cross-work stream understanding.

Data Connect:

Discussion on the connection between Data Connect and Elwazi project. Taku/Ian talked about his practical example - discussed the use of phenotype data and its connection to DRS IDs in data queries.
Ian/Taku discussed identifying specific issues, looking for ways to link various data points, and considering a standardised format to enhance data retrieval efficiency between different projects. He'll be presenting in the next Data Connect meeting.

2023-12-13 - CANCELLED

2023-11-15

Chair(s): Michael Baudis, Nara Sobreira
Attendees (Affiliation): Beatrice Amos, Miro Cupack, Jonathan Fuerth, Howard Lander, Kayla Socarras

Agenda

	Agenda Item	Person/Time
	Welcome, Housekeeping	
	Follow-up from last meeting	
	Rolled over items… Service-Info work -> interest? VariantMatcher <-> Beacon working group / planning meeting?	

	Experiments Metadata -> news? Beacon plans?	
	Connect 2024 attendance, pre-planning	

- re variant matching in Beacon: Beacon v2 options are shown at https://beaconplus.progenetix.org/ (always developing…) and discussed in Beacon docs

Link to meeting recording:
https://us02web.zoom.us/rec/share/mlrnzCskgk5x8dIojo9YcwKWJzjH10hO9n2HAiDBpK-aat8ruwtUVFVZhwTjlwIJ.WVELAshhZOWuiehb?startTime=1700061962000
Passcode: KSy7&J+s

Main action items:

1. Follow up to get representatives from the missing groups (DaMaSC, Beacon and Experimental Metadata) involved in future meetings.
2. ✔ BA - send the notes to Michael so he can review them.
3. Follow up on David Bujold, Experiments Metadata to see how their work aligns with the Beacon data model and what concepts may be missing or represented differently.

Main decisions:

1. Making these monthly meetings an hour long instead of half an hour to allow more time for discussion.
2. Exploring using service info to provide more information about what individual nodes/services can provide, such as geographic location data.
3. Starting to document use cases and common patterns for service info to potentially update the standard.

Meeting summary:
Summary of the previous meeting's objectives.

Aims is to **gather leads** from the Discovery groups to discuss the roadmap and the future of the Discovery working group.
Missing representation from Discovery groups like Experimental metadata, Beacon and DaMaSC.
Update on the connections between databases like Variant Matcher. Kayla suggests using Beacon to improve matching, citing structural issues with Data connect.
Kayla will update in the upcoming Beacon meeting.

Need to clarify misconceptions and address communication gaps. Michael mentions the importance for Beacon framework to address issues around lacking information despite the protocol's openness.

Global Alliance
for Genomics & Health

Nara mentions existing capabilities within the Variant Matcher and identified gaps like variant effects, thinking about ways to incorporate these missing elements into the protocol through iterative development. Nara proposes adding a query parameter for variant effects, allowing users to filter and index Beacon data more effectively.

Maybe we need to explore functionalities without feeling blocked by external discussions. Also showcasing the functionalities of connected nodes like Variant Matcher in upcoming meetings to offer a better understanding to the entire group and to outline the technical details and parameters required for requests and responses within Beacon.
Suggest a focused meeting to delve into these discussions and avoid diverging into other topics prematurely.
Service info: need to provide more information about service info and its role in the Beacon protocol.
Miro suggests **creating a database of groups using Beacon** metadata to facilitate data integration and schema integration across different GA4GH products.

2023-10-11

Chair(s): Nara Sobreira
Attendees (Affiliation): Beatrice Amos (GA4GH), Jonathan Fuerth, Ian Fore
Apologies: Michael Baudis

Agenda

	Agenda Item	Person/Time
	Welcome, Housekeeping - Communications - Connect review, recap	
	-	
	Project Updates, and any previous action items: ● Beacon API v2 ● DaMaSC ● Data Connect API	All
	* Variant Matcher & Beacon - How to make this happen? (& Data Connect?)	Nara

	* Re-activating work on service-info (expanding - e.g. geographic data, ... - w/o breaking) with concomitant implementation (Beacon dev. branch?) * Re-activating SchemaBlocks for the parts of DaMasC which fit there and review of structure, supported by Data Connect requirement examples # * Change in Discovery profile, responsibilities w/ emphasis (product, documentation?) on data service usage cataloguing	
	AOB	

2023-09-13

Chair(s): No leads present
Attendees (Affiliation): Beatrice Amos (GA4GH), Jordi Rambla, Ian Fore, Wolfgang Pernice, Isaac Odiase, Miro Cupak, David Bujold, Mike Conway (NIEHS), Jonathan Fuerth

Apologies: Michael Baudis, Nara Sobreira

Agenda

	Agenda Item	Person/Time
	Welcome, Housekeeping - GA4GH Connect 2023: San Francisco -	
	Discovery sessions at Connect/Plenary September 2023: - Beacon Extensions - Beacon Filters - Phenopacket Beacon Connect - cancelled - Data Models and schema consensus - (Federated variant matching)	
	Project Updates, and any previous action items: ● Beacon API v2 ● DaMaSC ● Data Connect API	

	Revisit Roadmap Discovery Strategic Roadmap 2022/2023 (Keep on Agenda)	
	AOB	

Link to meeting recording:
https://us02web.zoom.us/rec/share/pHR_WqFHHjChMKfX2O6G04xrDhNafIZCO2RrGst9vQ-jdVmhpwyZkj6X6z8YXi7M.K0rBwXUN_tVm30SD?startTime=1694614923000
Passcode: 4Ctye=9M

2023-08-09

Chair(s): Michael Baudis
Attendees (Affiliation): Beatrice Amos (GA4GH),

Jordi Rambla, Heather Creasy, Ian Fore, Emily Kefferson, Isaac Odiase, Kathy Reinold, Tony Brookes, Lauren Fromont, Miro Cupak

Apologies: Nara Sobreira

Agenda

	Agenda Item	Person/Time
	Welcome, Housekeeping - GA4GH Connect 2023: San Francisco - Submit a poster abstract: form - Register for the conference: https://broadinstitute.swoogo.com/ga4gh11thplenary - Hotel booking link for Plenary >> https://book.passkey.com/gt/219333164?gtid=6a35e33de59169ab53000d64ed38ab96	
	Project Updates, and any previous action items: ● Beacon API v2 ● DaMaSC ● Data Connect API	
	Discovery sessions at Connect/Plenary September 2023:	

	- Beacon Extensions - Beacon Filters - Phenopacket Beacon Connect - Data Models and schema consensus - (Federated variant matching)	
	Revisit Roadmap Discovery Strategic Roadmap 2022/2023 (Keep on Agenda)	
	AOB	

2023-07-12

Chair(s): Beatrice Amos (GA4GH)
Attendees (Affiliation): Nara Sobreira, Jordi Rambla, Heather Creasy, Ian Fore, Emily Kefferson, Isaac Odiase, Kathy Reinold
Apologies: Michael Baudis, Tony Brookes, Lauren Fromont, Miro Cupak

Agenda

	Agenda Item	Person/Time
1	Welcome And a warm welcome to Discovery's new co-lead Nara Sobreira	Beatrice
2	Project Updates, and any previous action items: ● Beacon API v2 ● DaMaSC ● Data Connect API	Jordi Rambla Kathy Reinold
3	Discovery sessions at Connect/Plenary September 2023	
4	Steering committee meeting updates?	
5	(Meeting frequency/ days / times still appropriate? Communication preferences / frequency?)	
8	Revisit Roadmap Discovery Strategic Roadmap 2022/2023? (Keep on Agenda)	
9	AOB	

Link to meeting recording:
https://us02web.zoom.us/rec/share/LqYqDHtilpWyL-TrGGhu4B2i9npUOq-LZcDAHVyCVVTwVuJCySPHIva52pKYoMt8.elqZcrDsDwp5aq3h?startTime=1689171927000

Global Alliance
for Genomics & Health

Passcode: !K7M%*0W

Minutes:
Warm welcome to Nara Sobreira, who is joining as our new Discovery Workstream lead. We're all really excited that she has joined.

NS: has been part of the Global Alliance. But more on the a search API part of things and sometimes connected to the Beacon API also.

Please review the DaMaSC Connect session proposal **here**. Lots of information will be provided here.

Beacon update:
JR: discussion on Connect session in September.
Suggesting 2 sessions:
1) Beacon extensions or add-ons (Beacon for images. That would start with radiology.) Another is plugging into Beacon v2 model on top of OMOP (common data model, sponsored by Odyssey, receiving a lot of traction in Europe - observational tool used predominantly in hospitals for secondary analysis..
2) Conversation round requirements for filtering and boolean logic

JR: Suggested GA4GH to connect with Odyssey to explore ways of working together. See recent Beacon meeting minutes and recording.

IF: OMOP and Odyssey focused on the health care world rather than the research world. Problem of trying to connect 2 worlds.
Are you using beacon to make research data sources available? Or Sources and making them available through Beacon.

JR: It's healthcare data in a repository that's meant for secondary research.
IF: explains the difficulties of representing research data in FHIR as it's a system that's meant for healthcare data.

JR: Comes from a Spanish personalized medicine project. Spanish/ European hospitals have agreed to use OMOP as their way of sharing data. Growing adoption. In Europe, a project called Eden is also moving data from many different hospitals into OMOP in the hospitals. In effect creating kind of a federation of hospital data in OMOP format. Starting in Spain, but spreading, with links to Europe too. Odyssey has many tools, but nothing quite like Beacon.

EJ: Would like HDR UK to be involved as a use case for : "Ian, you'd be interested in seeing how well it works when you point it at other things."

EJ: Explains rapid set up of a way of doing cohort discovery, across 50 different data custodians on board. (Scotland and Wales). Have a cohort discovery tool embedded in HDR gateway https://www.healthdatagateway.org/ accessing many different datasets across the UK
Found Odessey tools lacking, so built tool called Carrot https://co-connect.ac.uk/docs/, https://hdruk.github.io/CaRROT-Docs/ (Convenient and Reusable Rapid OMOP Transformer: bringing together health data from across the UK)
Would be interested in open standards, if already done with OMOP and Beacon.

JR: suggests participation at Plenary in San Francisco, as more information will be available.

EJ: (in chat):
https://www.jmir.org/2022/12/e40035/
https://co-connect.ac.uk/

IF: Data Connect (as Miro is absent)
Hopes discussion can continue on how Discovery standards can come together, as some relation to Kathy's discussion on DaMaSC, e.g defining a method that GA4GH dynamically describes data sources, in a way that mappings can be created. Hopes that conversation from London Connect can continue in San Francisco about how the discovery standards come together.

✅ BA: check Nara has a calendar invite to DaMaSC meeting.

NS: would like to propose a session to discuss using the Beacon API to connect variant level databases and the next MME-type project.
KR: One of the DaMaSC use cases is the fact that our GA4GH variant databases have 3 different representations for variants and how do we bring them together. So ties in with what is being done in DaMaSC.

IF: suggests continuing from last year in that computable cohort representation series, which is looking at Beacon and Data Connect and FHIR, and Odyssey in the context of what the problem is and not in the context of let's take a technology. And focus on the technology. Let's focus on the problem.

2023-06-14 - CANCELLED

2023-05-10

Chair: Michael Baudis
Attendees: Kathy Reinold, Miro Cupak, Mike Conway (NIH/NIEHS), David Bujold, Ian Fore, Heather, Amreen, Beatrice Amos

Global Alliance for Genomics & Health

Actions

	Actions Arising	Assigned To	Deadline
1	Recap Connect - Discovery Roadmapping (agenda) - DRS and Data Connect (agenda) - Beacon, Phenopackets, Cohorts… (agenda)		
2	Plans for working groups, schedule		
3	Leadership notes		
4			

Minutes

- Note about DaMaSC
    - Kathy working on soliciting speakers for a Workshop at Sept GA4GH Meeting in San Francisco. More at the next DaMaSC meeting. Question of whether DaMaSC operates under Discovery umbrella or TASC umbrella TBD after Discovery re-org (new co-chair coming).
- Beacon - Phenopackets
    - ICGC Use cases likely to drive this
        - This GIF project is bringing together Phenopackets and Beacon
- DRS & Data Connect session

Kathy - should DaMaSC be part of Discovery or TASC? Planning a session at Connect in September.
Michael - Connect review - anything interesting or outstanding? Anything that affects this group?
David - unfinished discussion on Beacon/Phenopackets and how they interact. Do we want to beagle to query in Phenopackets?
MB - how can we align them better? Beacon does not want to do schema definitions and models where unnecessary. One of the questions is how can we make Phenopackets as a data model directly supported as a standard Beacon?
Many discussions round implementations - driven by use cases, some driven by Melanie Courtout etc.
Ian - need more representation for Discovery. Feeling that a lot of discussions are theoretical or abstract. Need more implementations. Disconnect. Suggests should involve a practical side - more GIF projects that will use Discovery products, e.g. ARGO projects.
MB - new co-lead nomination in progress. Discovery is currently in a transition phase, but in the future the focus will be more on the Discovery. Where are resources? We now have products that have reached a level of maturity.
Miro - Data Connect session at Connect meeting - Discussion was round how to obtain metadata representing compound objects. 2 approaches: one being bundles in DRS and the

other Data Connect. There was agreement to provide best practices. No tradeoffs for adopting one specific approach. Will need to continue this discussion.
Mike Conway: is working on Gen3 and FASP. Fwiw that's our interest, through the lens of Gen3 data meshes which also intersect with DRS. (Comment) Building a data catalog that brings together dynamically data from many data sources. Trying to understand data mesh standards.
MB: Just to iterate: The main strength of Beacon is on the discovery of data, which doesn't have to be variants - cohorts features, files ... And access to content can be provided in various ways; Beacon default model, alternative models, data handovers (e.g. file links ...). All used in implementations. (Comment)
MB - need to separate the framework from the data model. The framework describes how should query and response be shaped?

Meeting recording:
https://us02web.zoom.us/rec/share/_j3OYikgctZsb1tGau6WuLi8w0d00HID_4mfXloqtS52KRblFweWpSBoLu_fDgtK.CnVnUzcbzCsZyd22?startTime=1683728945000
Passcode: $MPX4&Vs

2023-04-12

Attendees (Affiliation): Michael Baudis, Kathy Reinold, Ian Fore, Jordi Rambla, Jonathan Fuerth, Jorge Oliveira

Actions

	Actions Arising	Assigned To	Deadline
1	Review Connect Materials - Discovery Roadmapping (agenda) - DRS and Data Connect (agenda) - Beacon, Phenopackets, Cohorts… (agenda) - April Connect Schedule - for reference		
2	Connect attendance?		
3	Ideas/plans for roadmapping…		
4			

Agenda

	Agenda Item	Person/Time
1	Discovery @ Connect	
2		

3		

Minutes

Connect attendance (p or v):

- Michael (p)
- Jordi (p)
- Kathy (v)
- Ian (p)
- Jonathan (v)
- Miro (v)

2023-03-08

Attendees (Affiliation): Michael Baudis, Ian Fore, David Bujold, Amreen, Kathy Reinold

Actions

	Actions Arising	Assigned To	Deadline
1	Propose Roadmap Session for April Connect - Be sure it's not just theoretical	Amreen	
2			

Agenda

	Agenda Item	Person/Time
1	Connect planning (last call)	
2		

Minutes

- Q (MB): Connect agenda?
  - Reviewed agenda here: https://docs.google.com/spreadsheets/d/1PSYI-bPwn41ziX-voxSVQAxxJjS0YWUA7jB9lMKblvQ/edit#gid=630249860
- Q (MB): GA4GH website update status?
  - Anybody interested in doing at least some general refresh on the https://ga4gh-discovery.github.io/ site? Michael could do this _but doesn't want to be all by mimself_. But there is an argument for it w/ the lengthy cycles at ga4gh.org …

- Agreed that the Schema Registry implementation is under Data Connect but there is considerable overlap with an Information Model Standard and thus both teams should come together in DaMaSC meetings (maybe also in Data Connect meetings?)

2023-02-08

Attendees (Affiliation): Michael Baudis, Jonathan Fuerth, Jeremy Adams, Ian Fore (NIH)
Apologies: Kathy Reinold

Actions

	Actions Arising	Assigned To	Deadline
1	Discovery-related project assessment		
2	People involvement and interests for point 1		

Agenda

	Agenda Item	Person/Time
1	Beacon	
2	Data Connect	
3	Networks	
4	A.O.B	

Minutes

Michael:

- reassessment of Discovery portfolio

Ian:

2023-01-10

Attendees (Affiliation): Ian Fore, Jonathan Fuerth, Fabio Liberante, Jeremy Adams
Apologies: Michael Baudis, Kathy Reinold

Agenda

	Agenda Item	Person/Time

1	Schema Registry Service Standard Proposal	
2	DaMaSC - Next steps? - Meeting time	
3	Beacon	
4	Data Connect	
5	Networks	
6	A.O.B	

Minutes

DaMaSC workshop at April Connect - in London. Ian and maybe Kathy

Schema Registry Service Standard Proposal
JA: ASI DP interest? Who would be the DP interested in this?
FL: Wait until new DP mixer sessions?
+1

IF: DaMaSC. Data Model/Schema Registry (DMR/SR)
Data Connect - as a registry of models/schema. Two tables that use the same model/schema. Reference a common model.

JA: DaMaSC - consensus across GA4GH. DMR/SR - also outwardly looking. An API that any organisation can develop, including GA4GH. Data Connect builds off DMR/SR - search action.

IF: Tiering. DaMaSC - GA4GH is the custodian of the models. Data Connect instance operator is the custodian of the models, layering of different models.

IF: dbGaP data dictionaries. How many data models are there in dbGaP? Quick Python script. 965, 54% of studies have the same pattern model. What does EGA look like from that point of view?
IF: Can create data dictionary from TSV pretty quickly.

JF: Opportunity to make DC feel more relevant, by building a schema that covers that 80% or so of dbGaP studies. Where can we find a good schema and exchange to JSON.

IF: The JSON schema DC is using. JSON schema doesn't tell us what unit it is. Dimensionless is still a unit.

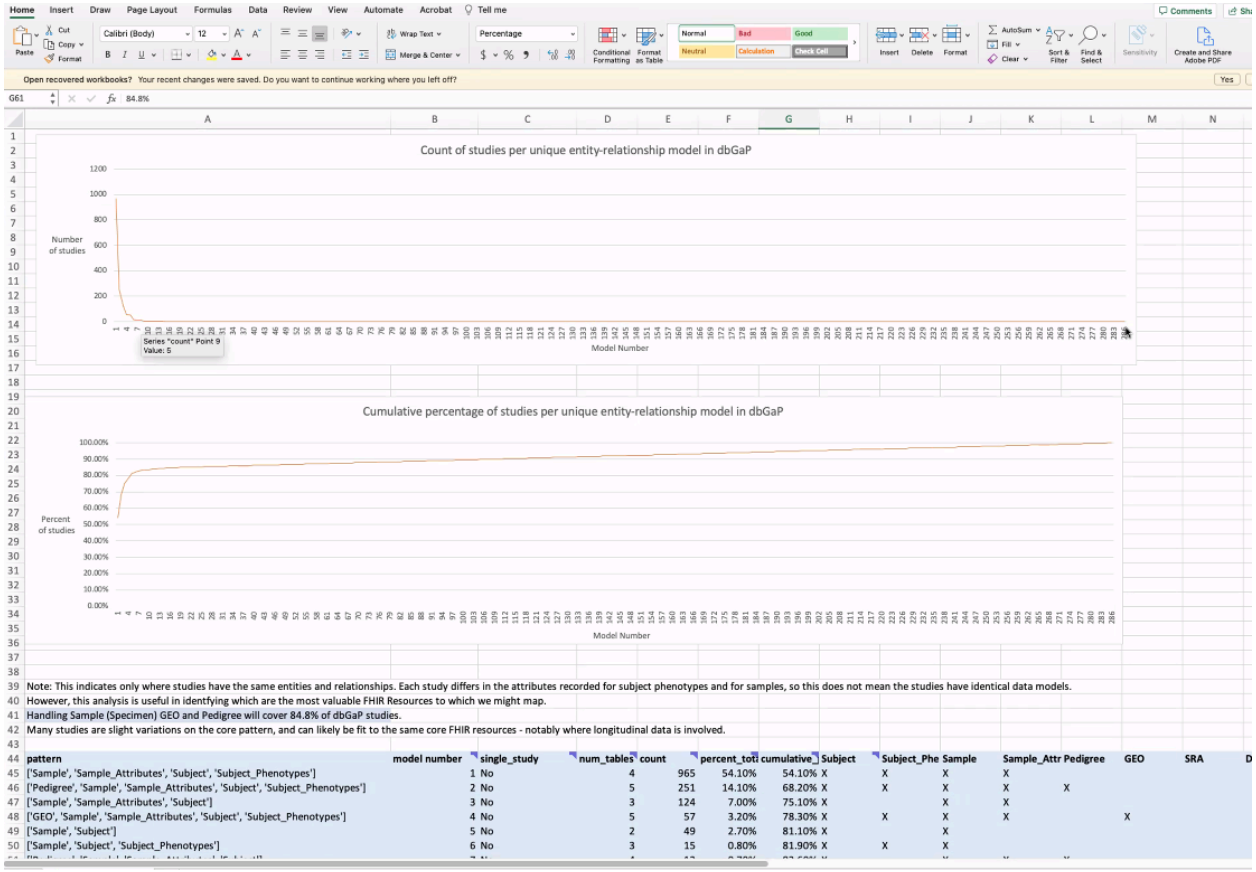
Global Alliance
for Genomics & Health

JA: A one pager on how these components fit together? Who do we need the buy in from? Secretariat, Driver Projects?

IF: 3 tiers. 0 - us. 1 - people that should be on this call - Discovery. Beacon. Need to talk to all of Discovery first. Then 2 - Clin Pheno then 3 - rest of GA4GH.

IF & JA: Push DaMaSC back to TASC. Schema Registry Standard within Discovery.

JA: Phenopackets and VRS etc. Should be a TASC person, Alex or Bob etc.
IF: Need neutrality of Tech Team.



Note: This indicates only where studies have the same entities and relationships. Each study differs in the attributes recorded for subject phenotypes and for samples, so this does not mean the studies have identical data models.
However, this analysis is useful in identfying which are the most valuable FHIR Resources to which we might map.
Handling Sample (Specimen) GEO and Pedigree will cover 84.8% of dbGaP studies.
Many studies are slight variations on the core pattern, and can likely be fit to the same core FHIR resources - notably where longitudinal data is involved.

pattern	model number	single_study	num_tables	count	percent_tot	cumulative	Subject	Subject_Phe	Sample	Sample_Attr	Pedigree	GEO	SRA	D
['Sample', 'Sample_Attributes', 'Subject', 'Subject_Phenotypes']	1	No	4	965	54.10%	54.10%	X	X	X	X				
['Pedigree', 'Sample', 'Sample_Attributes', 'Subject', 'Subject_Phenotypes']	2	No	5	251	14.10%	68.20%	X	X	X	X	X			
['Sample', 'Sample_Attributes', 'Subject']	3	No	3	124	7.00%	75.10%	X		X	X				
['GEO', 'Sample', 'Sample_Attributes', 'Subject', 'Subject_Phenotypes']	4	No	5	57	3.20%	78.30%	X	X	X	X		X		
['Sample', 'Subject']	5	No	2	49	2.70%	81.10%	X		X					
['Sample', 'Subject', 'Subject_Phenotypes']	6	No	3	15	0.80%	81.90%	X	X	X					

○