Global Alliance for Genomics & Health

[ARCHIVED]

New meeting notes here: **Discovery_subgroup-Beacon_API_Meeting_Minutes_2025**

GA4GH Beacon API

Discovery Work Stream

2017 | 2018 | 2019 | 2019/2020 | 2021 | 2022 | 2023 | 2024 | 2025

Minutes and Actions 2024

These are the minutes for the Beacon API Sub Group of the GA4GH Discovery Work Stream. For further information, please visit the GA4GH Work Stream page. This call happens every second Tuesday of the month.

Table of Contents:

2025-01-14
2024-12-10
2024-10-08: Connect review and Best Practices
2024-09-10
2024-07-09
2024-06-11
2024-05-14: Connect review
2024-04-09
2024-03-12
2024-02-13
2024-01-09
2023-12-12
2023-11-14: Beacon planning meeting
2023-10-10: Cancelled
2023-09-12: Connect Planning

Global Alliance for Genomics & Health

2023-08-08: Connect Planning
2023-07-11 - Small group meeting
2023-06-13: 6th meeting 2023
2023-05-09: 5th meeting 2023
2023-04-11: 4th meeting 2023
2023-03-14: 3rd Meeting 2023
2023-02-14: 2nd Meeting 2023
2023-01-10: 1st Meeting 2023

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.
- Dates should be specified in the ISO format yyyy-mm-dd

Meeting Details:

To join the call please use Zoomlink:
https://zoom.us/j/8713805924

GA4GH have added password protection:
Zoom Passcode = GA4GH

- Mailing list ga4gh-beacon@ga4gh.org (To join this group please contact liina.nagirnaja@crg.eu or beatrice.amos@ga4gh.org)

About Beacon Scouts

Beacon has several subgroups, working to solve specific problems within Beacon v2. These include:

Beacon Scout groups
Beacon Aggregation and Cohorts: aims to gather requirements for aggregated requests and responses.

Beacon Filters: aims to simplify the addition of the AND and OR boolean operator and enhance the current filter solution for improved functionality.
Beacon Variants: aims to expand its repertoire with more SNP/SV cases, propose specification extensions to accommodate new use cases, and introduce typed queries such as deletions, fusions, and translocations.
Beacon Dev: aims to stabilise the API by addressing bugs, optimising code, and refining the specification.
Beacon matchmaking: is gathering requirements for matchmaking use cases, and proposing Beacon API extensions.

Previous Meeting Minutes

This was previously represented as the Beacon Strategic Group

1. Beacon API - 2017 - 2018 - 2019 - 2019/2020 - 2021 - 2022

List of Meetings

2025-01-14

New meeting notes here:

Discovery_subgroup-Beacon_API_Meeting_Minutes_2025

2024-12-10

Chair: Tony Brookes, Michael Baudis

Attendees (affiliation): Liina Nagirnaja, Beatrice Amos, Sergi Aguiló Castillo, Dmitry Repchevsky, Oriol Lopez-Doriga, David Bujold, Alessandro Sulis, Gordon Krieger, Kayla Socarras, Maria Yazykova, Miguel Santos, Natassa, Pamela Heredia, Vittorio Meloni

Apologies: Jordi Rambla

	Agenda Item	Person/Time

1.	Welcome and introductions:
	- Connect meeting in Boston 1-4th April 2025
	  - Session proposals due Jan 6th: please suggest proposal ideas here: https://easyretro.io/publicboard/mjJSZh2ANUR0qe1or9gIcY2Dt3z2/5d11bd74-eb92-4d5c-9101-cdd86c8cf000
	- Beacon meetings to accommodate Eastern timezones to begin in Jan
	- Beacon Implementers Forum starting in Jan/Feb
	- Work on Beacon 2025 roadmap:
	  - Please continue to fill out Beacon Community Survey: https://docs.google.com/forms/d/1HH89cyK6jXSvrL36uFdA-0FagCZf64nFlQ9r2tliLZE to help shape the direction and priorities
		Bea 2’

2.	Exploring the Sample Locator, Samply.Beam, and new Stratification capabilities	Tobias Kussel 15-30 mins
3.	Updates on the Scout groups: summary document	Liina

Link to meeting recording:
https://us02web.zoom.us/rec/share/xEv7d549joaoPoqZcRv9k2ng9U9g6dCLWOkhGf3uMK8HM1uBPq2XRUaJhrYlLJVN.j9P5640T9862BGxa

Meeting Summary:
Intro:
Tobias Kussel will provide an in-depth overview of the BBMRI Sample Locator platform's current functionalities and tech stack. He will cover pre-stratification mechanisms, anonymized data aggregation, and secure data-sharing processes. He will also introduce "beam-based transport tunnelling" via Samply.Beam, a middleware developed to navigate networking complexities like firewalls and proxies.
We'll also discuss the recent GDI pilot, utilising Samply.Lens and Samply.Beam for Beacon query visualisation, along with possible Beacon API extensions in EUCAIM to support compound queries and expanded stratification options.

—--

Intro:

Tony emphasized the importance of consolidating progress with various scouts and addressing outstanding issues in the Beacon framework. He acknowledged some pending fixes for bugs and urged clarity on the status of scouts and bug resolution.
Tony also **proposed exploring Beacon's potential for metadata discovery in 2025**, highlighting interest from UK health projects in using Beacon to transition from central to federated catalogs. He suggested this as a focus for future development and encouraged Beatrice to note it for planning.

Beacon Scouts summary:
Liina shared a document summarising self-reflections from each of the Beacon scouts groups, outlining their current statuses, goals, and future directions.

1. Filter Scouts:

They have both short-term goals, such as recommendations for ontology filters and filtering term extensions, and long-term goals, including logic for combined filter queries. They are preparing proposals for community review.

2. Cohort Scouts:

Have progressed with a specific proposal presented in their last meeting and are exploring ideas like federated cohort building. Discussions have touched on philosophical aspects of what Beacon's role should be, emphasizing less analysis-oriented approaches.

3. Variant Scouts:

Led by Michael, are addressing breaking and non-breaking changes and working to align with the latest VRS schema. While no specific deadlines are set, Michael is actively suggesting and refining ideas.

4. Dev Scouts:

Focus on maintaining and improving the GitHub repositories, addressing governance, feature deprecation, branch management protocols, and improving visibility of information. Efforts also include clarifying the version release process and cleaning up code.

Liina noted the need for clarity about decision-making processes within the Beacon community and suggested forming a Beacon network implementation group or channel for better collaboration. The summary also included tagged issues and an overview of how the scouts envision their direction for 2025.

—

Tobias Kussel presentation:
Tobias from the German Cancer Research Center presented tools and technologies developed for federated data exploration within the **BBMRI-ERIC** ecosystem. He showcased the **BBMRI**

Locator, a search interface that queries patient-level data across sites in real-time while ensuring anonymity through differential privacy methods. The infrastructure relies on modular components, such as a deployment vehicle ("Bridgehead") and an abstract query translation system, enabling interoperability with various standards like FHIR and OMOP. A key innovation, **SimplyBeam**, facilitates secure and efficient data transmission through restricted network zones, overcoming challenges with firewalls and proxies. Kussel discussed the application of these tools in scenarios involving Beacon endpoints and explored integration methods to enhance secure data sharing and querying capabilities. The open-source nature of these tools invites collaboration and supports diverse research needs across federated networks.

—

Dmitry asked about the types of data used in the system, particularly whether it stores only medical records or also includes genomic data.
Tobias clarified that BBMRI has limited genomic data integration, mostly molecular markers, and aims to enhance genomic queries using the Beacon protocol.
Dmitry expressed concerns about data storage formats, such as using MongoDB with the Beacon format, and queried the communication protocols used for querying data.
Tobias explained that their system employs standard HTTPS for secure and firewall-compatible communication and uses generalised query grammars, which can be adapted to site-specific data formats.

Tony expressed interest in aligning projects and leveraging the discussed locator system in larger initiatives. He emphasised the **need for increased collaboration and international funding** to support the work, noting the challenge of relying solely on volunteer efforts.

Michael commented on the integration of Beacon for data discovery in biobanking and its potential as a lightweight protocol for sample discovery and data profiling. He highlighted the alignment between Beacon and other query languages, while Tobias acknowledged Beacon's utility but noted that their explorer tool offers capabilities beyond Beacon's current scope, such as analysis tasks formulated directly as queries.
There was agreement on **the need to extend Beacon for profiling aggregate data and enabling discovery across diverse resources**.

Beatrice summarised her 3 points:

1. managing heterogeneous data sources and ensuring interoperability through standards like FHIR, OMOP, and Beacon;
2. addressing scalability challenges as sites and data grow; and
3. exploring what contributions or collaborations the open-source community might bring.

Tobias Kussel responded by emphasising the importance of understanding use cases and engaging with various communities to enhance interoperability and refine tools. He suggested

exploring solutions such as **deploying Beacons in secure networks, enhancing privacy-preserving data sharing, and improving exploration tools**.
Tobias also acknowledged the need for deeper engagement with the Beacon community and expressed interest in collaboration.

David mentioned ongoing efforts within aggregation and cohort scouting to build meaningful interfaces directly utilizing Beacon without alternate mechanisms.

Dmitry proposed using JSON Path for filter specifications in Beacon protocols, citing its alignment with the Beacon schema, and also discussed exploring advanced security protocols, including encryption standards like E4G, for enhanced end-to-end security.
Tobias agreed these suggestions were valuable, sharing that Beam currently uses end-to-end encryption with application-level security but welcomed further exploration of Dmitry's recommendations.

2024-10-08: Connect review and Best Practices

Chair: Jordi Rambla, Tony Brookes
Attendees (affiliation): Liina Nagirnaja, Beatrice Amos, Sergi Aguiló Castillo, Dmitry Repchevsky, Oriol Lopez-Doriga, Kayla Socarras, Maria Yazykova, Deepthi, Miguel Santos, David Bujold, JB Poline, Mamana Mbiyavanga, Miguel Santos, Sebastian Urchs, Gordon Krieger, David Croft, Kayla Socarras
Apologies: Michael Baudis

	Agenda Item	Person/Time
1.	Welcome and introductions	Bea 2'
2.	Beacon at Connect - key takeaways and highlights	Jordi / Sergi
3.	Beacon Best Practices	Kayla Socarras / 15'
4.	Beacon Scouts - status updates	
5.	Update on the product approval process (URGENT!)	Beatrice

Link to meeting recording:
https://us02web.zoom.us/rec/share/4GwGRZTuG7qjkMR4paoF4RqB36CpDDawVq4BE0D9rKlD5DI7bQCbvmpPEHkdpAto._lHtuJQuFd3mrOKD?startTime=1728395813000

Meeting Summary:

Global Alliance
for Genomics & Health

Jordi has some pull requests open that require modifications or clarifications (not urgent) - some of the PRs from July need slight adjustments that Maria and Daniel can help with.

Tony raised concerns about the **difficulty in narrowing down key issues** and suggested **focusing on a few main priorities with clear timelines**.

What is the current objective of the Beacon standard?
Tony suggests that it seems to lack a clear definition compared to version 1. He asked if the goal was to:
a) fix issues with version 1,
b) deploy it in more settings, or
c) create a universal Beacon network,
noting that the group cannot tackle everything at once.

Jordi responded by acknowledging the **need to stabilise and solidify version 2** of Beacon, as projects have only recently begun implementing it and are encountering issues. He noted that while there are many requests for extending the Beacon model (such as for transcriptomics, images, etc.), these extensions may not be the remit of this group. Regarding a global network, Jordi agreed it is a goal of GA4GH but not directly the responsibility of the working group, which focuses on specifications. Ultimately, much of the group's direction depends on feedback from Driver Projects and community engagement.

Beatrice suggested a multi-pronged approach for the Beacon project to gather feedback for the Beacon roadmap to guide Beacon development. The Beacon community survey, which she plans to first review with Liina to refine the questions, aims to gather meaningful feedback from the Beacon community. It's important to listen to the community and implementers, particularly Driver projects, which use and test the standards. The goal is to create a coordinated approach that incorporates all perspectives and helps establish clear priorities. She mentioned a leadership meeting to decide on these priorities, which will shape the roadmap and key deliverables moving forward. Beatrice acknowledged that the current process feels chaotic due to the many ongoing activities, but reassured the group that efforts are being made to stay on course.

Connect meeting review:
Jordi summarised key takeaways from Connect. He highlighted two main points:

1) the discovery of several Beacon implementations, including some networks that were not following any specific standards due to the lack of formal guidance, and
2) a productive conversation around the benefits of different models for genomic data exchange.

He noted that there was strong support for a hierarchical network model, akin to the Beacon network approach, over a peer-to-peer model like Matchmaker Exchange.

Jordi proposed two major outcomes:

a) the formation of a Beacon Implementation Group to create a more organised community around Beacon efforts

   Tony Brookes (in chat):

> Development & Beacon User Group (DEBUG) !?

   , and

b) moving forward with the idea of an "Internet of Genomics" using the Beacon network on a larger scale, potentially under GA4GH's leadership.

He also mentioned minor points, such as aligning Beacon with the VRS project to increase its adoption.

Sergi added that time zone differences, particularly with Australia, made it difficult to engage globally. Action: create an alternating meeting for the Eastern hemisphere, perhaps every 6 months.
David noted that the Cohort team had recently solidified use cases for Beacon v2 and were working on a proposal for cohort aggregation. He emphasised the importance of leveraging GIF projects for practical implementations across Driver Projects.

Jordi concluded that while multiple, varied implementations of the spec are inevitable, as long as the process remains under control, it could still be considered a success.

Kayla presentation [link]:
Kayla introduced the work being done by the Variant Level Matching (VLM) platform group, which aims to connect various databases using the Beacon network v2 API for rare disease research. The group focuses on one-sided variant matching to gather data for solving unsolved disease cases. She outlined a theoretical use case involving querying databases to find evidence for variants of interest and discussed the complexity of matching to individuals with consistent or inconsistent phenotypes. She provided an overview of the team involved, including connections to other platforms like Seqr and Variant Matcher, and mentioned ongoing efforts to bring Seqr online by early 2025. She introduced the Best Practices document, started in April, to track decisions and justifications for the platform. Although written in a layperson-friendly format, she's collaborating with software engineers to include more technical details to help both non-technical and experienced users establish their own nodes and join the VLM network.

Jordi noted that while Beacon flexibility can be beneficial, many projects prefer more structure. He likened the situation to Google Docs, where users can customise formatting but may need a style guide to ensure consistency. Similarly, the VLM platform should follow certain best practices and branding guidelines to maintain coherence. He suggested that in the future, it may

be necessary to establish rules or templates on top of the specification to validate not only the technical syntax of responses but also adherence to project-specific rules.

Dmitry: *What does "similarity" mean in the context of genomics? Is it based on mutations, variants, haplotypes, or population frequency? How should this concept be understood, especially by implementers who may interpret it differently?*

Tony explains that the similarity being described refers to relationships between different ontologies, such as HPO terms, SNOMED terms, and others, using a graph-based approach. This pre-computed graph allows for querying across these ontologies, enabling users to move from one concept to another (e.g., from genes to pathways to diseases). While it's not focused solely on "similarity," it allows for powerful querying and discovery of relationships across multiple concepts.

Beatrice provided an update on the product approval process for Beacon v2.1.0. She explained that GA4GH has introduced a new process requiring all updates to be reviewed and approved by the Product Review Committee (PRC), which consists of three members, which have just been confirmed. **Please note that development can continue in parallel. This approval is considered more of a formality and should not affect ongoing work.**
Jordi expressed concerns about sharing the unapproved version of Beacon, but Beatrice and Tony clarified that it is possible to share both the approved and unapproved versions, leaving the choice to users. Beatrice confirmed she will coordinate with Andy Yates for a final confirmation. There was also a discussion about breaking changes already introduced into Beacon v2. Liina suggests streamlining the PRC process for minor bug fixes to avoid unnecessary delays.

2024-09-10

Chair: Michael Baudis, Tony Brookes
Attendees (affiliation): Beatrice Amos, Sergi Aguiló, Dmitry Repchevsky, Oriol Lopez-Doriga, Kayla Socarras, Gordon Krieger, Maria Yazykova, Vittorio Meloni, Deepthi, David Croft, Francis Nguyen, Pamela Heredia,
Apologies: Jordi Rambla, Liina Nagirnaja, Tim Beck

	Agenda Item	Person/Time
1.	Welcome and introductions	

2.	Beacon Scouts - status updates ● Variants (Michael)	
3.	v2.1 …	
4.	Website - genomebeacons.org	

Link to meeting recording:
https://us02web.zoom.us/rec/share/NgB73FDRkWHjsDt-6r34uVQzu9TeBK5JTPEbKeOZIckkl-Nbic1QQ7SYHvHXbnvd.zraGvdu2G8QlwrsV

Pre-meeting notes:

Scouts - Variants:

- last meeting 2024-08-19
- **outcome** cytogenetic events:
  - no additional cytogenetic parameters suggested right now; should be implemented in front end
  - following VRS "adjacency" model for fusions
  - keeping discussion of local implementation (how to represent or query fusion events in your database …) separate from the model/query parameter design
- **outcome**: Github tag for variant scouts issues - please follow
- subtopics for separate work (so far no schedule/separate meetings but landing doc)
- need for updated "variation" part of the variant model following VRS v2 (updated VRS to be demo'd at the Australia Connect/plenary?)
  - Michael started an updated schema (proposal)
  - currently very early draft demonstrating the addition of VRS v2 structures to the genomicVariations model
- overall the "VCF-like" use of "prototypical variants" which have then "caseLevelData" occurrences was seen as problematic
  - this is a problem mostly independent of the re-definition of the "variation" part
  - use of per-instance variants ("calls") with id/fingerprint to identify similar events seems worthwhile to consider

—--------

Variant Scout update: Beacon scouts: VARIANTS 2024
Key issues included the need for additional query parameters, specifically querying and representing fusions, aligning with ongoing VRS work. The group agreed not to use cytogenetic parameters for queries but to translate them to sequence-based formats. There was also a focus on handling multi-variant queries, co-occurrence of variants, and updates to the variant model, particularly with the introduction of adjacency models.

Global Alliance
for Genomics & Health

A GitHub tag for variant scouts was introduced, please contribute! The updated Beacon variant model now incorporates elements from VRS2, with ongoing testing and adjustments.

Tony expressed concern about the overwhelming number of questions and possibilities in the Beacon and Scout groups, suggesting that they may not be addressing key issues effectively. He noted that the groups are struggling to focus and prioritise due to the volunteers' limited time and the vast scope of tasks. Tony proposed that instead of trying to address everything at once, **the teams should concentrate on a few critical issues, assign them to the scouts, and set specific timeframes for resolution**. Michael agreed with this approach but highlighted the importance of community involvement in guiding the process and balancing the workload. It may not be possible to resolve all issues immediately, but focusing on manageable tasks and adapting based on community feedback could lead to better outcomes.

Dmitry raised concerns about the current data model, specifically the issue of genomic variants being linked to multiple individuals or samples. He described this situation as problematic for his implementation of Passport visa systems, where such ambiguity complicates filtering and accessing data.
Dmitry suggested that each variant should be tied to a single individual or dataset to avoid inadvertently exposing sensitive information about multiple patients. The current schema lacks proper links between individuals and datasets, leading to potential data modeling issues.
Dmitry proposed addressing this by improving the specification and linking individuals directly to datasets or cohorts.
This issue should be formalised as a topic for further discussion and potentially added to the specification as a new issue.

Sergi mentioned improving the way genomic data is linked to datasets. He highlighted issues with the current system, where variant IDs are not effectively linked to specific datasets or cohorts. He pointed out that the existing method, which involves managing separate files for these links, is prone to mistakes and performance issues. A new approach could involve updating the standard to better handle ID linking directly within the datasets and cohorts. This would address problems with data management and improve clarity in workflows, such as querying for bio samples based on specific criteria. **This issue should be documented and treated as a development priority, with a focus on creating better solutions and improving documentation for these processes**.

Beacon documentation and domain management:
Michael reviewed the current state of the documentation, noting that many updates had been made by himself. The current state of the Beacon documentation is in need of feedback and contributions from the community. The original domain was under Beacon-project.io, but it is now transitioning to genomebeacons.org, which will serve as the primary site for Beacon documentation. This new domain is being updated and will eventually replace the old site.

Global Alliance for Genomics & Health

Michael would like to consolidate and update frequently asked questions (FAQs) and improve the documentation process. Please contribute community FAQs - anyone with questions or information should contribute to the documentation or submit issues on GitHub. Liina and others will be actively involved in restructuring and enhancing the documentation. Increased participation in maintaining and improving the Beacon project's online resources required.

This page is one which needs updates :-) http://docs.genomebeacons.org/contribute/
... probably should also be moved to the main doc site...
http://genomebeacons.org/people/

Michael expressed concerns about the administrative burden of managing these assets and suggested that establishing a non-profit organisation or similar structure might be beneficial for long-term maintenance and governance.

Release policy for Beacon Version 2.1:
We need to ensure that changes align with release policies and these changes should be effectively communicated to the community.

Dimitri asked about the implications of supporting Beacon 2.1 in implementations and whether any network adjustments would be required. He noted that the changes were minor and questioned the necessity of a formal approval process for this version.

Beatrice provided an update on the new product approval process for GA4GH. This process requires minimal work from contributors but involves getting approval from three independent reviewers. The purpose is to ensure that standards remain relevant and implementable on a global scale.

The Beacon 2.1 release was a minor update, and won't need full GA4GH approval. Beacon 2.0 could remain the officially approved version, while 2.1 could serve as a development version. Users can choose to follow 2.1 if they wish, but it was agreed that the more significant updates and approvals would be reserved for future major versions.

2024-07-09

Chair: Jordi Rambla
Attendees (affiliation): Beatrice Amos (GA4GH), Sergi Aguiló (BSC), Dmitry Repchevsky (BSC), Maria Yazykova, Tom Conner (GA4GH, Broad), Oriol Lopez-Doriga Sagales, Nara Sobreira, Daniel Tsirulnikov, Jon Eubank, Miguel Santos (INSA, PT), Kayla Socarras
Apologies: Liina Nagirnaja, Tony Brookes, Michael Baudis

	Agenda Item	Person/Time
1.	- Register for Connect Plenary meeting - *One meeting remaining before Connect Australia!*	Bea / 2 mins

2.	- Demo by Microsoft Research: slides
	  - Beacon API Client auto-generated from standard files
	  - *List of issues to be fixed (with fix proposals) in standard files*
		Daniel Tsirulnikov, Maria Yazykova / 15 mins

3.	● - Scout updates:
	  ○ - Dev
	  ○ - Cohorts
	  ○ - Filters
	  ○ - Variants
	  ○ - Matchmaking
	  ○ - RNA
		Scout leads

Link to meeting recording:
https://us02web.zoom.us/rec/share/gkkvIKJYCHkEZVT-pC6cfsIDTP0S03ixVZTYABqyE8WSrxxknJ52Xdtmh-3Afjec.2fHtaxVMgoi9KmHW

Notes/Minutes:

Daniel's presentation: (Link to slides)
Microsoft Research has been working to implement internal APIs using the Beacon API standards. Firstly, Microsoft Research aimed to develop APIs for their internal needs, but they soon realised the potential of using Beacon API standards to auto-generate code, which would significantly reduce maintenance costs and streamline updates.

Daniel Tsirulnikov (software engineer at Microsoft Research), took the lead on this project, focusing on creating a Python API client for the Beacon API. The goal was to develop standardised code that would ensure consistency and reduce manual coding errors. This approach also included benefits like automated boilerplate code, inline documentation for ease of use, and the potential to extend code generation to other programming languages in the future.

He compared the current manual process of writing requests with URLs, headers, and data to the improved process, which utilises auto-generated objects from the standards. This change simplifies the task for developers by ensuring requests and responses are standardised.

More benefits of this approach include the **ability to generate server code templates**, making it **easier to implement Beacon instances**. The process also **serves as a validator for**

schema files, catching errors early. The team has identified and fixed several issues with the specification files, and they plan to submit these fixes to the main GA4GH Beacon v2 repository.

The workflow involves:

- cloning the main Beacon repo,
- applying necessary fixes,
- preprocessing files to combine the framework and endpoints,
- generating the code,
- fixing any issues,
- running tests, and
- updating the Python Package Index if there are changes to the standards.

The ultimate goal is to use the generated code internally and open source it for external use.

Daniel demo'd the initial setup and execution of their automated code generation process for the Beacon API standards. He highlighted the config files and necessary patches that address existing issues not yet resolved in the main repo. The process involves cloning the Beacon repository, applying patches, creating a virtual environment, and installing dependencies.

He showed how easy it is to run the script from the command line, which automates cloning the repo, applying patches, setting up the environment, and generating code from the standards. The generated code includes both model and API components based on the Beacon schema, significantly streamlining the creation of standardised requests and responses.

He included an example of querying bio-samples using the generated code, demonstrating both simple GET requests and more complex POST requests with filters.

As for next steps, the team plans to submit the identified bug fixes and address an ambiguity in the Beacon v2 repository. He hopes to have these updates reviewed and integrated, aiming to complete and open-source the project by August⭐👏.

—----

Dmitry shared his past experience implementing Java without automated code generation - on the ELIXIR Repository. While Beacon specifications have likely improved since then, he faced challenges due to not using OpenAPI for their project.
https://github.com/elixir-europe/java-beacon-v2.api

He raised concerns about the data model, questioning whether the current implementation generates Python files for beta models like bio-samples or if it only utilises a protocol and part of the response as standard Python.

Response by Daniel - he clarified that the current implementation does not generate specific parameters like "samples" for bio-samples. Instead, it generates all objects specified in the standards. This means that any object defined within the standards will be created automatically by the system.

Usage of OpenAPI and related tools for generating code and specifications for the Beacon API: While Python was the current tooling choice, OpenAPI Generator supports various languages, including Java.

Regarding server-side implementation, it was clarified that OpenAPI Generator primarily generates routing and model structures but doesn't include data fetching or business logic. This is seen as a good starting point for server implementations but requires manual addition of business logic tailored to each instance's requirements.

The approach of using config files to generate OpenAPI specifications dynamically mentioned. This approach aims to consolidate the Beacon API configuration details into a single source, facilitating easier management and ensuring consistency between implementations. This is great for developers that are unfamiliar with OpenAPI - it allows them to modify configurations directly and regenerate specifications as needed.

There was a discussion about the implementation and testing of Beacon API standards using configuration-based approaches. Daniel did experiments to validate the viability of generating server endpoints and running tests. He aims to create standard test examples for Beacon APIs, ensuring consistent implementation across instances. Using a Python client for API calls allows researchers to fetch data directly into their notebooks, facilitating data analysis and aggregation tasks efficiently.

The framework is flexible, where configurations can be adjusted or extended with minimal tweaks, enabling reuse across different implementations. The intention to open-source this framework was mentioned to promote community adoption and lessen the need for constant support from the team.

Jordi suggested potentially dedicating a separate team to testing,but perhaps people are spread too thin?

Looking ahead, the next steps involved Daniel and Maria finalising their code and presenting their findings on issues encountered, along with updates to the code generator. The expectation was to address these issues and refine the tooling further, possibly preparing for a future review or deployment phase.

Daniel talked about his strategy moving forward, outlining two main objectives:

Global Alliance for Genomics & Health

Firstly, he will submit identified issues to the repo for review while continuing his ongoing dev tasks. These tasks included finalising tests and making necessary adjustments to a package they were preparing to open-source. Once everything was validated, he plans to release it on the Python Package Index for broader use.

He asks what is the best approach to submitting fixes for the issues, with considerations around whether to directly submit pull requests or start with issue submissions.
He's leaning towards creating issues first to facilitate pre-pull request discussions and streamline the review process.

Maria shares a link:
fixes from presentation:
https://github.com/ga4gh-beacon/beacon-v2/compare/main...datsirul:beacon-v2:datsirul

Variant Matcher Scout update (Nara):
Nara shared updates regarding progress made on integrating databases and refining APIs for querying. Jessica reported successful setup of Gen2 data to be queried alongside existing databases via the Variant Matcher. This integration aims to streamline data accessibility across platforms, critical for ongoing projects.
She mentioned recent considerations on expanding query capabilities between databases, particularly focusing on querying based on specific genomic features beyond simple variant identification.
She talked about the feasibility and necessity of these enhancements, with ongoing dialogue expected to refine implementation details leveraging the Beacon API effectively. Also mentioned is the security measures for data exchanges, exploring the relevance and effort involved in implementing GA4GH Passports. She questioned the necessity of this security layer given the nature of the queries being performed. These topics are slated for continued discussion to ensure both technical advancement and robust security protocols in future meetings.

Filters Scout update (Maria):
The focus was on reviewing suggestions by Dimitri regarding challenges and resolutions for filtering by age. Several help issues were discussed and addressed based on Dimitri's input. Apart from these updates, no other significant developments were reported during the meeting.

Dev Scout update (Jordi):
No specific issues with significant advancement. Need to review and finalise fixes currently in development and see when to merge them into the main branch. Team need to prioritise tasks and ensure all updates and fixes are thoroughly examined before integration. Liina will update on next meeting date.

Global Alliance for Genomics & Health

2024-06-11

Chair: Jordi Rambla, Michael Baudis
Attendees (affiliation): Beatrice Amos (GA4GH), Liina Nagirnaja, Kayla Socarras, David Bujold, David Salgado, Francis Nguyen, François Ancien, Gordon Krieger, Ian Fore, Maria Yazykova, Oriol Lopez-Doriga Sagales, Nara Sobreira

Apologies: Tony Brookes

	Agenda Item	Person/Time
1.	- Register for Connect Plenary meeting - Connect session agendas - deadline due June 19th - Connect session proposal form - *Two meetings remaining before Connect Australia!*	Bea / 15 mins
2.	VLM status update	Kayla / 10 mins
3.	External meetings - ELIXIR All Hands	AHM participants
4.	Scout updates: - Dev - Cohorts - Filters - Variants	Jordi, Scout leads, session leads, attendees / 15 mins
	AOB	All

Link to meeting recording:
https://us02web.zoom.us/rec/share/wwuJFmcvL0JQsOZ2FEE21hWKG0HOveR5_BQubbG0HFwkETJ8Ij0Y86Q-OtgpU4r8.-oD0fVxh8BBlNB75

Notes/Minutes:

Connect session brainstorming:

Beacon Sessions - Brainstorming

Jordi mentioned potential sessions for the upcoming plenary and preceding Connect meeting. Suggestions included:

- showcasing real-world use cases to cater to an audience unfamiliar with Beacon,
- organizing a coding hackathon to engage a broader audience with practical activities

- have sessions focused on technical extensions and their implications.

Michael proposed with the GA4GH tech team - building a Beacon registry and testing setup using existing Beacon standards to identify gaps. This setup would send queries to Beacons in a test network, helping to profile them and provide feedback on responses, ultimately enhancing the system based on **real feedback**.

Bea - we need to identify suitable topics that would resonate with the local audience. Bea and Liina to flesh out these ideas and potentially propose them formally. The overall goal is to leverage the event's location to engage with local implementations and provide valuable, participant-driven discussions.

Please can everyone suggest ideas. Bea will check with Scout leads!

Kayla and Heidi are interested in organizing a session on variant level matching, incorporating related topics such as phenotype data, variant exchange builds, Passports, and Phenopackets. Kayla mentioned that Heidi was interested in collaborating on this effort due to significant overlap. Nara will send a follow-up via email to finalize these details.

Bea will help identify who will be present, not only from the Beacon community but also from other projects such as the Driver Project and other Australasian initiatives. Bea to check if projects like the 100K Asian Genomes will be present, emphasizing the importance of including relevant local and regional projects, whether or not they are part of GA4GH.

Variant Level Matching (VLM) (Kayla and Nara):

Contact Kayla Socarras: VLM meetings are at 8AM EST for the last Thursday of the month (Email: ksocarra@broadinstitute.org)

Kayla mentioned the development of a best practices document for creating a node in the VLM network, which she plans to share once it's more refined.

Nara added that Beacon has been used to connect the first two nodes, with Variant Matcher now linked to Franklin. These databases allow researchers, clinicians, and patients to connect based on specific variants, providing information on the presence of variants in other databases and their zygosity. The next step involves exchanging contact information, which may require minor adjustments to the Beacon API, and addressing IRB issues related to user consent for sharing information.

Nara also mentioned ongoing discussions about the necessary permissions for sharing phenotype data and efforts to maximise the amount of information shared without requiring extensive permissions. The third node, managed by Tudor and involving the University of

Washington's Gene 2 MP database, has been connected, and testing is underway. Will be live soon.

Michael raised the idea of including contact information in Beacon responses, suggesting this be handled through a standardised handover object. He emphasised the need to file this as an issue on GitHub to develop a schema for embedding or linking contact information in Beacon responses.

Nara explained that for gene-level exchanges, they currently provide the name and institution of the contact. Michael agreed and suggested that defining this process as part of the schema would standardise it, allowing for secure and consistent sharing of contact information. He noted that while various implementation details would need to be discussed, filing it as an issue on GitHub would be the appropriate next step.

Jordi asks about the ethics committee or Institutional Review Boards (IRBs). Specifically, there were concerns about the protocols for sharing clinician contact information when a match is identified and the limitations on how much phenotype information can be shared between clinicians. Do these limitations apply to both clinicians involved or just one?

Nara mentioned two main issues:
The first issue pertained to data sharing consent for users providing information to Franklin. Users are aware that Franklin aggregates their data and shares findings, but they were not specifically informed that their contact information might be shared with other users interested in their data. Franklin is now establishing a protocol to ensure users explicitly consent to this before their contact information is shared. Each database may have its own procedures regarding this issue.

The second issue involved the extent to which phenotyping data is identifiable and can be shared. There is a debate on how much phenotyping information makes an individual identifiable, particularly for those with rare diseases. In Cleveland's variant database, there are no significant limitations on sharing phenotypic information. However, the Cavatica database, following the NIH-funded "All of Us" project rules, does not share phenotype information if a variant appears in fewer than ten individuals, hindering rare disease research. The team, including Dr. Thomas and Heidi, discussed this issue and concluded that there should be no strict limitations on sharing phenotype information unless it involves very small and isolated populations. They decided to adopt a policy similar to Cleveland's, allowing as much information sharing as possible, provided that individual sites ensure proper consent is obtained.

Maria addressed concerns about privacy and security, emphasizing that many reported issues are corner cases. Jordi would like to hear more practical, real-world concerns from DIY projects rather than theoretical ones. Nara mentioned that some don't see risks in sharing phenotypes or allowing phenotype queries, suggesting a preference for focusing on concrete issues rather

than statistical exercises or hypothetical scenarios. Having a person's genotype often provides sufficient information without needing to consult additional resources like Beacon.

There was a discussion about the intersection between variant discovery and representation vs annotation. It was noted that while gene-specific queries are possible, there is a lack of documentation for using variant effect filtering terms.

The suggestion was made to review and potentially update the standards to explicitly define access control mechanisms appropriate for different use cases, such as data discovery, variant matching, and browsing aggregated data. This would ensure that privacy and security considerations are adequately addressed across all functionalities. The importance of this review was highlighted, given the diverse roles and access requirements that Beacon now supports.

Beacon Variant Scouts (Michael):
This group will be determining how to represent variants, define query types, and decide which parameters and filters to use. Will aim to achieve agreement and create documentation based on provided use cases.

Beacon Cohorts (David):
The group is focusing on the aggregation scale and had their first meeting since Connect. They agreed on a primary use case for a proposal to the main Beacon group. This use case involves improving the aggregation of data within a cohort, specifically to facilitate displaying distributions of values (e.g., through histograms or pie charts). While the main Beacon specification allows for this, it currently requires a large number of queries. Will propose either an additional endpoint or augmenting existing endpoints to streamline this process - will ensure that any changes will not break the existing specifications.

Beacon Filters (Maria):
In last week's Scout meeting, the group summarised their discussions and takeaways from the Connect meeting. It was their first meeting since Connect, and they focused on defining criteria and procedures for evaluating and deciding on proposal solutions. Multiple solutions had been proposed and agreed on a procedure to manage this through GitHub issues - will allow for discussion and decision-making over time. Once comments are resolved, Maria says they plan to close the discussion and finalize the solution for the next Beacon API update. A proposal for filter changes was also posted on GitHub for further discussion [link].

Michael said he encountered complexities with a seemingly simple issue regarding a proposal he had made earlier. The main challenge was dealing with more complex query structures, which could address many issues but also add significant complexity. The group debated whether to pursue simpler solutions in parallel or tackle the larger, more complex problem all at once. **Questions raised about the administrative process for integrating proposals from**

the Scouts teams into the Beacon code. The current procedure involves documenting the solution, reviewing it by the API team (including Michael and others), and pushing it forward after multiple review rounds. However, a formal step-by-step process is still lacking.

Jordi on the process for proposing and integrating solutions:
He prefers to avoid excessive formalization and suggested an organic approach. The guidance was that philosophical or non-concrete discussions should go into GitHub discussions, while more concrete issues with potential code suggestions could proceed to GitHub issues. Once a proposal is sufficiently developed and the Scout team is satisfied, it should be presented at the meeting for review and feedback. After a review period, if there are no major objections, the proposal could be integrated into the main project, potentially leading to an update in the beacon version.

Michael agreed with this approach, emphasizing the importance of following up on GitHub discussions to ensure progress isn't stalled due to lack of feedback. He noted that sometimes issues might seem obvious but still require active engagement. An example mentioned was the "excluded attribute" issue, which highlighted the complexities and dependencies involved in seemingly straightforward problems.

Jordi feedback from ELIXIR/ GDI meeting:
Multiple countries are collaborating to establish a Beacon network. He said that this collaborative effort has raised unforeseen issues, prompting a need for more prescriptive guidance from the Beacon team. There's a growing preference for clear recommendations on network coherence, ontologies, and other technical aspects. He suggests that the community might expect more direction from the Beacon team rather than purely democratic decision-making. He proposed revisiting the level of guidance provided in the Beacon V2 specification, suggesting a more descriptive approach, even if it adds complexity.

Jordi says clarity outweighs complexity in this context, as it benefits the community in the long run.
Michael says **it's important to document use cases and query scenarios** to focus on relevant aspects and avoid unnecessary complexity. We need to document specific features, which can also serve as test scenarios and aid in understanding the system.

David says while a more prescriptive approach might seem less democratic, it could actually simplify processes, especially considering projects like the Pan-Canadian Genomics Library and collaborations with the GDI.

Global Alliance for Genomics & Health

2024-05-14: Connect review

Chair: Jordi Rambla, Tony Brookes
Attendees (affiliation): Beatrice Amos (GA4GH), Ramón Benítez Brito, Ivo Fokkema, Hangjia Zhao, Sara Sant'Ana, Sergi Aguiló, Dmitry Repchevsky, Tim Beck, David Bujold, Oriol Lopez-Doriga Sagales, Alex Tsai, Gordon Krieger, Jon Eubank, Liina Nagirnaja, Ian Fore, Maria Yazykiva, Matthew Baker, Olesya Melnichenko, François Ancien, Kayla Socarras, David Croft

Apologies: Michael Baudis, Tony Brookes

	Action Items	Assigned to
1.	Please can all session leads review session meeting notes, and add key takeaways	All session leads and contributors

	Agenda Item	Person/Time
1.	Welcome, reminders: - Plenary in Melbourne, Australia	Beatrice 2 min

2.	Review of Beacon satellite meeting in Zurich
	Review of Connect meeting in Ascona
	- Decisions made, action items, key takeaways
	- Filters
	  - Clear request for the OR functionality
	  - Considerations for GET & POST support
	- Variants
	- Around the world in one query
	  - Several ideas with VRS, normalization…to explore
	  - How to declare the supported notations?
	- Beacon resolved
	  - beaconId in the Beacon response, is it actually needed?
	  - Should we consider decentralized networks?
	  - Discuss with TASC if service-info and service-registry should be revived
	  - OpenAPI: Make developers generating the OpenAPI from their implementation
	- Cohorts
	  - ICGC ARGO lighting a Beacon on top of Arranger
	  - Bento needs aggregations to facilitate UI display
		Jordi, Scout leads, session leads, attendees

3.	Beacon Scouts updates and next steps: - Filters	Scout leads

	- Variants - Dev - Cohorts - Matchmaking	
4.	Status of previous action items: - March 12th: Work on pathogens extension for Beacon (Jordi) - March 12th: Consider adding "feature" as new entry type to Beacon (Jordi) - proposed collaboration between Beacon team, ELIXIR structural variations community (a.k.a. hCNV), GKS, other interested parties	
	AOB	

Link to meeting recording:
https://us02web.zoom.us/rec/share/pMpdIA43ph4Frh2ghlOFxHjnMkcYCH4FtcuNCazgoDgizjFDVCcdJzrud4iyhjxN.Gm-hPUCURmHtELXz

Notes/Minutes:
Jordi - Beacon meeting was positive. Noted a preference for meetings with pertinent sessions. Finds more productive.

Liina recently joined the team (replaces Lauren). Liina in research for ~20 years, most recently focusing on genetic and genomic determinants of male infertility.

Connect review:
Satellite meeting in Zurich, focused on Beacon-related conversations. Discussions were insightful and David said it was good for understanding Beacon philosophies etc.
Jordi mentioned that there were productive discussions regarding the integration of Beacon network requirements, with various ideas proposed. Topics included the explicit inclusion of the Beacon network in the specification and the feasibility of centralized networks. Tony Brookes had suggested exploring the support for centralized networks and utilizing service info registry and service register to meet additional needs.

Open API:
Discussion on facilitating developers to create an Open API from their own implementation rather than providing a predefined API interface. This approach aims to simplify the process and address potential complexities associated with Open API integration. Need to explore tooling to facilitate this process was proposed to overcome existing challenges.

Global Alliance
for Genomics & Health

During the cohorts meeting, Jordi mentioned Argo's plans for a significant upgrade and the importance of aggregations for Bento's UI charts. Aggregating results to facilitate chart plotting is a priority, with the vendor team taking the lead.

Maria said it's important to implement filters in standardization to avoid complications in Beacon networking and federated search. She focused on reviewing open GitHub issues related to filters, with plans for the Filters Scout to address filter querying, display methods, logic, and implementation strategies.
Tim - upcoming Filters Scout meeting on June 3rd.
Maria says there is a need proactive implementation of features by various parties, underscoring the urgency for standardization decisions to prevent proliferation of different implementations.

Maria highlighted the importance of focusing on query formats and validation strategies for filters. Will be discussed in the upcoming meeting in June. She said some users are employing Beacon beyond its intended purpose of discovery, particularly in querying data, which necessitates further understanding and potential adjustments. Feedback from projects deploying Beacon implementations is beginning to emerge, revealing the **need for more sophisticated solutions than what the current version of Beacon offers**.
The balance between complexity and ease of use is crucial, and community feedback will guide future developments. The Filters team is building a library of filter queries, including those supported by the spec and those desired by implementers, to test future updates effectively. This initiative aims to ensure consistent support and compatibility in future spec revisions.

Jordi talked about the use of synthetic versus real-world data for testing and validating Beacon implementations. While synthetic data is commonly used, real-world data presents more complexities and is increasingly necessary for comprehensive testing. The **focus should be on the queries being sent to Beacon rather than the nature of the underlying data**. As Beacon development progresses, more thorough testing, including unit tests and validation of logic, becomes crucial. Do we need a data schema to understand where the data originates?

<u>Previous action items:</u>
Potential extensions for Beacon: one proposed extension involves incorporating data related to infectious diseases, such as COVID-19, into Beacon datasets. What specific elements should this extension include? Pathogen description and genome data? What about epidemiology-related information? Does this need further discussion?

Addition of a feature endpoint to Beacon to allow for the inclusion of genomic transcript data: It’s an important feature, especially with the emergence of RNA-related extensions. The broader applicability of this feature beyond RNA data, considering its relevance to various experimental types like Chip-seq is mentioned.

There is clear emphasis on the evolving needs of the Beacon specification to accommodate emerging data types and use cases. **We need to ensure Beacon continues to be relevant and useful within the genomics community.**

Bea - we need to ensure effective communication and coordination within the Beacon community to manage the growing interest and enthusiasm for Beacon initiatives. It's becoming increasingly important to organise discussions into definable action items and milestones to track progress effectively. Also in managing Beacon Scouts/projects, and maintaining effective communication between different groups.
Jordi and Oriol ended by emphasizing the importance of aligning efforts with user requirements, improving machine readability, and ensuring compatibility with the Beacon specification.

2024-04-09

Chair: Jordi Rambla, Lauren Fromont, Tony Brookes
Attendees (affiliation): Sergi Aguiló, Dmitry Repchevski, Tim Beck, Francis Nguyen, Sara Sant'Ana, David Bujold, Oriol Lopez-Doriga Sagales,
Apologies: Beatrice Amos, Michael Baudis

	Action Items	Assigned to
1.	- Filters: call for all DPs to contribute some real world queries!	All

	Agenda Item	Person/Time
1.	Welcome, reminders: - Connect meeting in Ascona, April 21st-24th - **Prepare your slides**! - Pre-Connect Beacon working meeting: **Please contribute to @** Zurich Beacon Implementers Workshop 2024-04-19	Lauren - 5 mins
2.	Invited presentation: ● Beacon re-identification	Sara de Amaral - 25 mins
3.	Beacon scouts: - Dev	Lauren and scout leads - 20 mins

	- ~~Variants~~ - Cohorts - Filters: call for all DPs to contribute some real world queries!	
4.	AOB: - Generics presentation by Dmitry - Plenary in Melbourne, Australia	

Link to meeting recording:
https://us02web.zoom.us/rec/share/ddN85bvJwerS7A1LsHGOas3wWdL6zKJyDoVmR9yw0CpboZvTN9nG4lxY3EADGk_j.AfFF2g4trDnl9AHJ

Notes/Minutes:

Brief summary:
Last meeting for Lauren and Dmitry - thank you for all your contributions to Beacon and GA4GH! Best wishes as you began your new chapter.

Sara (masters student from GDI, Portugal) presented on Beacon risks. There were updates from each of the Beacon Scout leads on their current work. Dimitri presented on using more advanced JSON schema features for Beacon specifications. Aggregation and cohorts, filters, and development topics were also discussed. Preparations were made for sessions at the upcoming Beacon Connect meeting to gather input from other projects.

Beacon and re-identification
Suggestion: instead of changing the budget, develop a tool to check for this > informing data owner about the actual sensitivity of the data
Some re-identification is just a question of probability

Dev scout Dev
Agreed on what is in scope > everything that does not fit elsewhere ;)
Will work in sprints of two weeks: discussions first and solutions to the next sprint

Connect: OpenAPI discussion + harmonization between beacon and BN responses

Cohorts/aggregation **scout**
Define the scope of the group
Cohorts: overlap with computable cohorts group (invited them to Connect)
Aggregation: multiple beacon nodes, display summary of the cohorts across the network
Focus on the UCs, use the Connect to attract more DPs
[JR] cohorts endpoint / aggregator not an endpoint but trying to cover dividing counts *by* a specific attribute e.g. disease

Global Alliance for Genomics & Health

Filters scouts
Mission:

1. close and resolve all outstanding issues
2. Most outstanding: AND/OR issue > Supporting any more expression?

Need to collect real world examples: call for DPs!! > collect test cases
> MME, any other?

2024-03-12

Chair: Jordi Rambla, Lauren Fromont, Michael Baudis, ~~Tony Brookes~~
Attendees (affiliation): Beatrice Amos, Nara Sobreira, Maria Yazykova, Sergi Aguilo Castillo, Miguel Santos, Dmitry Repchevsky, Oriol Lopez-Doriga Sagales, Deepthi, David Croft, Rodrigo Martin, Sara Sant'Ana, Valentin Groues, Francis Nguyen, Ian Fore, Kayla Socarras, Taku

	Action Items	Assigned to
1.	Prepare topics for pre-Connect session on Friday before Connect: Draft document @ Zurich Beacon Implementers Workshop 2024-04-19	Michael
2.	Finalize session proposals for GA4GH Connect. Deadline is March 20th	Lauren and Bea
3.	Need a transition plan for Beacon scouts before Lauren leaves the role	Lauren
4.	Work on pathogens extension for Beacon	Jordi
5.	Consider adding "feature" as new entry type to Beacon	Jordi
6.	Beacon Scouts: Kickoff meetings planned for Variants scout (Michael), Cohorts scout (David), Developers scout (Jordi)	Scout leads

	Agenda Item	Person/Time
1.	Welcome, reminders: - Connect meeting in Ascona, April 21st-24th - Confirm agenda and schedule	Beatrice - 5 mins

	- See list of relevant meetings below - Plenary in Melbourne, Australia - Pre-Connect Beacon working meeting: Draft document @ https://docs.google.com/document/d/1P97-waXxCjar-XQOrEKbec2d5ElHv8q67pU6aYyB9j8/edit	
2.	Scout updates (in chronological order of first KO): - Filters KO 2024-02-28 - Variants KO 2024-03-18 - Dev KO 2024-03-21 - Cohorts KO 2024-04-02 - Matchmaking (KO tbd with Nara)	Lauren - 10 mins
3.	Discuss Driver Project contacts Driver Project matrix: GA4GH Driver Project Engagement Matrix	Jordi - 10 mins
4.	AOB: - Lauren's last day with the Beacon team: April 12th 🥲	

Link to meeting recording:
https://us02web.zoom.us/rec/share/bE_oX3B1owpIzIcr4CjZcSmF2rBoW34YNKKEWg0jZI06kuoZCuO5crQofQOOB3QQ.gRN65wkEHrbKd1d-

Notes/Minutes:

1. Events

Connect in Ascona, list of relevant sessions (* indicates sessions we we directly participate)
* Beacon variants: Sunday April 21 at 10:30 - 12:00 CEST
* The Beacon Filter solutions: Sunday, April 21 Time TBD
DaMaSC Session I: Sunday April 21 at 15:00 - 16:30 CEST

* Phenopackets for querying Beacon: Monday April 22 at 10:30 - 12:00 CEST
Experiments Metadata Standard: Monday April 22 at 13:00 - 14:30 CEST
* Data Discovery DP Workshop: Monday April 22 at 15:00 - 16:30 CEST

* Around the world in one query: Tuesday April 23 at 9:00 - 10:30 CEST
* Beacon cohorts and aggregator: Tuesday April 23 at 10:30 - 12:00 CEST
DaMaSC Session II: Tuesday April 23 at 13:00 - 14:30 CEST

* Implementation of Beacon in Cancer Use Cases: Wednesday April 24 at 9:00 - 10:30 CEST
* Beacon resolved: a session for development: Wednesday April 24 at 10:30 - 12:00 CEST

Meeting summary:

Global Alliance
for Genomics & Health

Review of the Beacon session agenda: focus of Connect is on Driver Projects and their implementations.

Beacon Scouts updates, and their progress and plans for future meetings. Each scout project's objectives and upcoming activities were outlined, along with the responsible leads for each project. Please contribute and provide feedback on the scout projects - Lauren: lauren.fromont@crg.eu .

Jordi - presentation on use cases related to cancer research and healthcare queries, highlighting the importance of aligning questions with available data and tools: **Meeting Recording**

Discussion of various aspects related to data queries and structures within the context of Beacon. There are challenges and complexities of handling requests that are not directly linked to specific entities or are of a diffuse nature. Need to mention the importance of setting clear expectations and not discouraging people despite the complexities involved.

Examples mentioned of queries related to genomic variations and structural variations - need clear terms and documentation across different use cases. What is the role of ontologies and databases like Clinvar in addressing queries related to variant classification and annotation? Jordi mentions the compatibility between Beacon and Clinvar in mapping data and responding to queries - need proper data mapping for effective retrieval of information.

Jordi and Michael talk about the need to address technical challenges to ensure interoperability between different systems or databases to facilitate efficient data querying and retrieval processes.

In chat:
Dmitry Repchevsky (BSC / ELIXIR.ES) to Everyone:
Phenopackets's serialization is not schema-friendly.
Beacon's biosamples is an intention to create a logic schema for the data found in Phenopackets..

What about the inclusion of variant data from ClinVar into the Beacon model, do we need to query by position due to limitations in the query model? There's a focus on addressing issues related to conflicting classifications of variants and the potential for enhanced annotations within Beacon to provide more comprehensive information.

Also considered is the idea of expanding the interface of knowledge bases like ClinVar to enable real-time queries, Jordi mentions the transition from Beacon v1 to v2 for various databases. The potential value of ClinVar becoming Beacon-compatible is explored, along with the need for standardized APIs to facilitate interoperability.

Need to address conflicting classifications of variants and the need to convey this complexity to users. Maybe important to provide tools and explanations to help users understand the data's context, especially concerning conflicting interpretations. There's consideration of how to handle such complexities within Beacon without overly complicating the querying process, balancing engineering solutions with usability for implementers.

Discussion on improving Beacon's functionality, including the need for better documentation and vocabulary for variant queries. Jordi mentions an upcoming extension of Beacon for RNA expression analysis. There are also plans for adding a new entry type to Beacon for querying by gene or transcript - rnaget. Jordi also mentions the development of a pathogens extension for Beacon, highlighting the diversity of approaches among different Beacon projects.

Lauren announces her departure in April for a new position. 💔

2024-02-13

Chair: Jordi Rambla, Lauren Fromont, Michael Baudis, Tony Brookes
Attendees (affiliation): Beatrice Amos, Nara Sobreira, Ronnie Loosen, David Croft (DKFZ), François Ancien (Université du Luxembourg), Olesya Melnichenko, Maria Yazykova, Gordon Krieger, Dmitry Repchevsky, Sergi Aguilo,Kayla Socarras, Mamana Mbiyavanga, Takudzwa N Musarurwa, Michael Duyzend, Heidi Rehm, David Bujold, Andrew Gearhart, Jonathan Fuerth, Oriol Lopez-Doriga Sagales,

	Agenda Item	Person/Time
1.	Welcome, reminders: - Connect meeting in Ascona, April 21st-24th - **2 Beacon meetings remaining** (*April 9th Connect Planning meeting*) - **Submission form deadline February 16th**	Beatrice - 2 mins
2.	Introduction and Beacon Scouts information and reminder ● Scout enlistment form ● Members and leads listed in the minutes below	Lauren 5 mins
3.	Federated Variant Matching Project	Kayla M. Socarras (Nara Sobreira, Heidi Rehm) -

		15mins
4.	Driver Project: Repository of the International Fetal Genomics Consortium	Ronald Loosen - 15mins
5.	David Croft will be presenting his ideas on the Stratification of Beacon (Link to slides)	David Croft - 15mins

Link to meeting recording:
https://us02web.zoom.us/rec/share/HFiyGdWQdtzjUmUIqwTCVfbqII8DSvhRSKdMIY9NQ7cGTYm6mfLqAIW5x0WGd6LD.ZiMqbJqPyRizhE4T

Notes/Minutes:

2. **Beacon scouts**
   a. **Dev scout**

Name	Email	Role
Jordi Rambla	jordi.rambla@crg.eu	Lead
Michael Baudis	michael@baud.is	Co-lead
Gordon Krieger	gordon.krieger@mcgill.ca	Member
Dmitry	redmitry@list.ru	Member
Mamana Mbiyavanga	Mamana.Mbiyavanga@uct.ac.za	Member
Oriol López-Doriga	oriol.lopezdoriga@crg.eu	Member
Sergi Aguiló	sergi.aguilo@bsc.es	Member

   b. **Filters scout**

Name	Email	Role
Tim Beck	tim.beck@nottingham.ac.uk	Co-lead
Maria Yazykova	mariayaz@microsoft.com	Co-lead
Francis Nguyen	Francis.Nguyen@uhn.ca	Member
Gordon Krieger	gordon.krieger@mcgill.ca	Member
David Bujold	david.bujold@mail.mcgill.ca	Member
Dmitry	redmitry@list.ru	Member
Mamana Mbiyavanga	Mamana.Mbiyavanga@uct.ac.za	Member
Michael Baudis	michael@baud.is	Member

Oriol López-Doriga	oriol.lopezdoriga@crg.eu	Member
Gemma	g.milla.martin@gmail.com	Member
Olesya Melnichenko	olesya.melnichenko@microsoft.com	Member

c. Variants

Name	Email	Role
Michael Baudis	michael@baud.is	Lead
Oriol López-Doriga	oriol.lopezdoriga@crg.eu	Member
Gemma	g.milla.martin@gmail.com	Member

d. Matchmaking

Name	Email	Role
Manuel Rueda	manuel.rueda@cnag.eu	Member
David Bujold	david.bujold@mail.mcgill.ca	Member
Heidi Rehm	hrehm@mgh.harvard.edu	Member
Dmitry	redmitry@list.ru	Member
Michael Baudis	michael@baud.is	Member

e. Cohorts/Aggregation

Name	Email	Role
Jordi Rambla	jordi.rambla@crg.eu	Member
Francis Nguyen	Francis.Nguyen@uhn.ca	Member
Gordon Krieger	gordon.krieger@mcgill.ca	Member
David Bujold	david.bujold@mail.mcgill.ca	Lead
Dmitry	redmitry@list.ru	Member
David Croft	d.croft@dkfz-heidelberg.de	Member
Tim Beck	tim.beck@nottingham.ac.uk	Member
Maria Yazykova	mariayaz@microsoft.com	Member
Mamana Mbiyavanga	Mamana.Mbiyavanga@uct.ac.za	Member
Gemma	g.milla.martin@gmail.com	Member

Olesya Melnichenko	olesya.melnichenko@microsoft.com	Member
Sergi Aguiló	sergi.aguilo@bsc.es	Member

Meeting summary:

Lauren provided updates on the Beacon Scouts, mentioning the various scout teams and the need for leads in some areas ⬆️. Need to clarify the roles of different scout teams to ensure efforts are not duplicated.

Technical aspects regarding variant matching and how different teams approach data sharing and querying were discussed. There was a suggestion to have a joint session maybe at Connect to discuss these issues further and ensure alignment.
There is a need for collaboration on how the various initiatives fit together, especially regarding data sharing and notification mechanisms.

Michael - Driver Project presentation aimed at implementing Beacon networks for sharing genomic and phenotype data, particularly focusing on prenatal genomics within the context of the international fetal genomics consortium. Technical aspects of implementing Beacon was discussed, including storing phenotype data as JSONs in a big query database on the Terra platform and deploying containers to interact with the database. Michael also touched on the importance of expanding Beacon implementation on Terra, integrating beacon with Phenopackets, and collaborating with other institutions to aggregate genomic and phenotype data for broader sharing.

Michael: we need a unified effort to use Beacon as a showcase API for searching over vast genomic and phenotype datasets.
Michael mentions the technical aspects of querying Phenopackets stored in a BigQuery database using fuzzy logic and hierarchical properties. He emphasizes the importance of practical solutions over perfect precision in Beacon implementation. He mentions compatibility challenges between query methods like query by example and query by method, especially in the context of integrating with the Matchmaker Exchange.

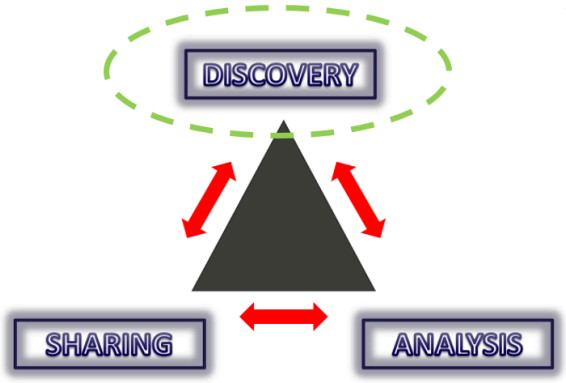
David talks about the concept of stratifiers, which are tools used to summarize data from various sources, typically in a visual format like histograms. Stratifiers help understand data distributions across different categories such as age, gender, diagnosis, and genetic variants. He shares examples from a network called "cool to PPMRI," highlighting the significance of stratifiers in identifying trends and patterns within datasets.

He discusses the technical aspects of implementing stratifiers within beacon systems, mentioning challenges faced when integrating them with queries. While standard filtering works

well with Beacon, stratifiers pose difficulties due to the need for multiple queries for each category, especially when dealing with numerous categories like ethnicity or genetic variants.

David talked about expanding the Beacon API to accommodate the simultaneous querying of multiple filters or codes and receiving back corresponding values in one query. This approach aims to improve efficiency when dealing with large datasets and multiple stratifiers. David suggests a feature request for developers to enhance the beacon API to handle such queries more effectively.
Jordi acknowledges David's feedback as valuable and aligns with the purpose of the "cohort scouts," which aim to address challenges related to stratifiers and querying strategies within the Beacon system.
Also mention of potential solutions for handling complex queries, such as querying for variants in specific genes or combining different types of variants in the same sample.

Michael acknowledges the complexity of these features and the need to balance functionality with ease of adoption. Generally Tony highlights the importance of refining the Beacon API to better accommodate diverse querying needs while ensuring usability and efficiency.



2024-01-09

Chair: Tony Brookes, Jordi Rambla, Lauren Fromont
Attendees (affiliation): Beatrice Amos, Melanie Courtot, David Bujold, Michael Baudis, Oriol Lopez-Doriga Sagales, Jon Eubank, Jorge Oliviera, Kathy Reinold, Kayla Socarras, Miguel Santos, Natassa, Nara Sobreira, Ronnie Loosen, Manuel Rueda, Gordon Krieger, David Bujold, Maria Yazykova, Olesya Melnichenko, Alex Tsai, Ian Fore, Mamana Mbiyavanga, François Ancien (Université du Luxembourg)

Global Alliance for Genomics & Health

	Actions Arising	Assigned To	Deadline
1	**Submission** to Connect	All interested	16/02
2	Follow this up w/ some direct examples which would work for their use cases.	Melanie and David	
3	MME use case presentation by Kayla next month	Kayla	
4	Distribute a document listing the proposed topics for Beacon Scouts/working groups, and people can sign up to lead or contribute to them	Lauren	
5	Will look into the open pull requests and issues on GitHub to address comments and merge requests	Jordi and others	
6	Will start documenting ideas for potential sessions at the Connect meeting in April, with a deadline of February 16th to submit session proposals	Lauren/Beatrice	

	Agenda Item	Person/Time
1.	Welcome and confirm agenda	Beatrice, Lauren
2.	Melanie/Jon/David will be presenting their specific use case for using Beacon for aggregated data (Link to presentation)	Melanie Courtot, David Bujold, Jon Eubank
3.	Roadmap review and refresh for 2024: - 2022/2023 roadmap - Beacon Scouts revival	 Tony / Jordi Michael / Jordi
4.	Action items and next steps for: **Beacon Issues to fix/cleanup - PRIORITY (before anything else)** Beacon Filters: rationalise design of ontology filters & value filters Beacon Filters: Allow arrays in filters (OR logic) Beacon Filters: Multiple filters of same type, AND (not OR) logic Beacon Filters: Allow option of ignoring unsupported filters and giving warning in Results Beacon result options: Results to allow counts or non-integer counts or approximate values or count ranges Beacon result options: Allow results to report on multiple entities, themselves in multiple hierarchical levels, and this implies also adding metadata (at least entity name/ID) to each result entity	Tony / Jordi
5.	GitHub issues & PRs updates	Oriol / Jordi
6.	AOB - Steering committee meeting, Jan 31st - WS updates meeting - Connect meeting in Ascona, April 21st-24th	Beatrice

	Agenda Item	Person/Time
1.	Welcome and confirm agenda	Beatrice, Lauren
2.	Melanie/Jon/David will be presenting their specific use case for using Beacon for aggregated data (Link to presentation)	Melanie Courtot, David Bujold, Jon Eubank
	- **3 Beacon meetings remaining** (*April 9th Connect Planning meeting*) - **Submission form deadline February 16th**	

Ad 5.:

- last version pf Beacon v2 structural variants document
  - https://docs.google.com/document/d/1cwwRQ2PtlN1dBffCugdkbSHWCPmLgLkADd-5mu-rVAw/edit#heading=h.wptwrivzjqha
- proposed collaboration between Beacon team, ELIXIR structural variations community (a.k.a. hCNV), GKS, other interested parties
- previous discussions about the topic at the last hCNV meeting (Michael, Jordi participating); concept for associated "white paper"
- proposal for a Connect session "Genomic Variant Formats: From Cytobands and VCF to Beacon Queries" <- suggestions welcome; idea already announced in the GKS thread
- AI: Bea to engage w/ Giselle & Michael, to create list of participants for new scout team & set up first call

Link to meeting recording:
https://us02web.zoom.us/rec/share/Osl7cAhnMukmlJYQIVf9ByHqFlmf6xpPS-fFFef11lOfgLtj7pwUtUORk-z6nsQk.yshpfMaxnT-CFwf8

Notes/Minutes:

Melanie presentation:
Melanie discusses a couple of use cases related to the aggregation of data. One involves a plan to have a central database storing information about a thousand cancer cases worldwide. Due to various issues, the initial plan of a centralized database has been modified, leading to the need for data discovery across different nodes with varying regulations. We need a dashboard or visual representation of phenotype distribution across the federated network, but there are challenges related to data privacy and granularity.
Another use case involves building a common cancer repository, aiming to exchange and query aggregated cancer data while ensuring controlled access. Melanie mentions the idea of using aggregated data in Phenopackets for machine learning and downstream software applications.

David presentation:

Global Alliance
for Genomics & Health

Pan-Canadian Genome Library - The primary aim of this project is to establish a federated framework for sharing genomic and clinical data across Canada, focusing on equity, diversity, inclusion, and leveraging data from indigenous and small populations. Their objectives include enhancing healthcare delivery effectiveness, facilitating clinical trials, and positioning Canada as a significant contributor in global research.
The plan involves offering a reference implementation for data nodes, leveraging standards similar to the GDI, with a focus on Phenopackets and Beacon compatibility.
The team is exploring aggregation within the Beacon, seeking to query the Pan-Canadian Genome Library (PCGL) federation to receive aggregate responses from each Beacon node for specific properties.

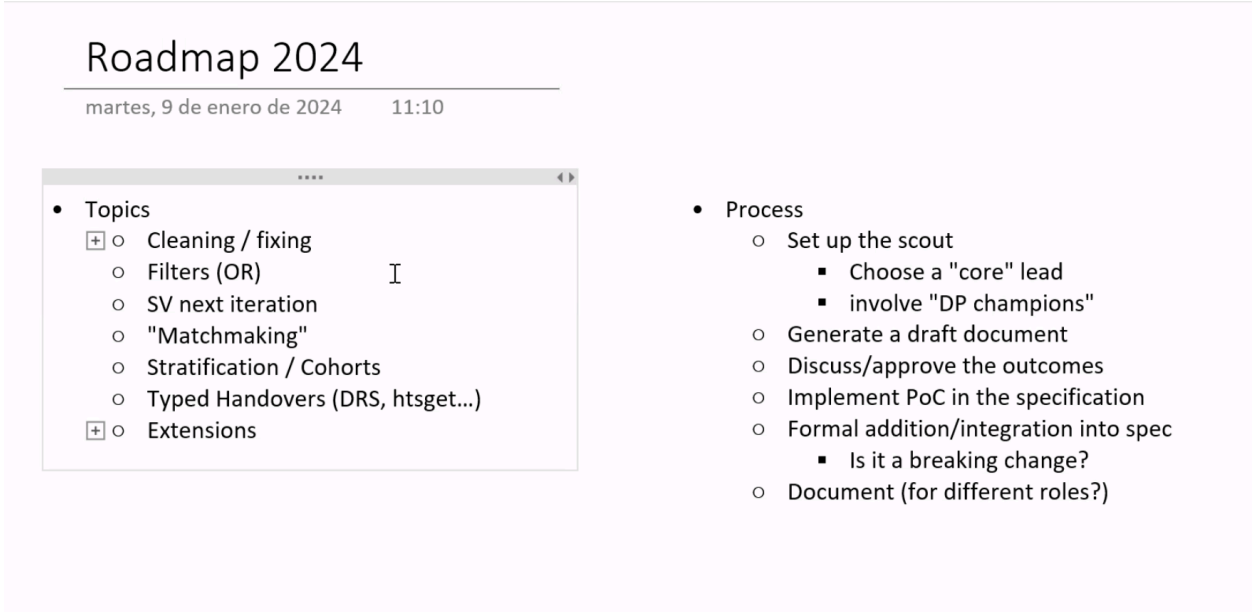
David referenced the CINECA project's intent to provide overviews based on queries and emphasised the desire to enhance compatibility across global initiatives, like the GDI, through standardised data transfer objects, particularly Pheno packets. Overall, the focus lies on enabling federated queries compatible with similar initiatives and potentially advancing interoperability among beacon networks for more insightful analyses.

There were deliberations on the format for aggregation within Phenopackets and its relation to individual-level packets. The focus was on establishing a tabular format for aggregated data that could accommodate queries involving multiple dimensions.
There is a need identified for a standardised format that transcends backend representations, potentially borrowing from existing tools or business intelligence standards.

There are considerations about the granularity of responses based on aggregated data; querying based solely on aggregated data might limit the scope of queries and responses. Perhaps prioritising aggregated data for both query and response could offer flexibility without mandating extensive standardisation. What about the practicality of using aggregated data for queries?

<u>Roadmap discussion:</u>

Global Alliance
for Genomics & Health



Main points / summary of this roadmap section:

- The Beacon group will establish scouts or focused working groups to work on specific topics from the proposed 2024 roadmap in more depth than can be done in monthly meetings. Lauren and Bea will convene.
- We will use the scout model from previous Beacon work, where small teams work independently and report back to this main group, to organise work on the topics. Makes work more manageable.
- We will prioritise cleaning and fixing the specification as the first scout topic, led by Jordi, since it's foundational work
- Lauren will circulate a document listing proposed scout topics for people to volunteer to lead or contribute to them

Jordi talked about the planning and discussing the agenda for the upcoming months, extending through the year 2024.
The initial aspect centred on **cleaning and fixing specifications**, identifying issues, and applying necessary updates, with examples drawn from GitHub discussions and considerations about open APIs.
Jordi mentioned future schema improvements involving filter rebuilds, operator inclusions, and the expansion of document content related to single variations, suitable viruses, and human copy number variations. There was also a mention of collaborating with various communities [what are there?] and individuals interested in these areas.

Also mentioned: matchmaking aspects and certification, alongside minor themes originating from projects like GDI, aiming to enhance Beacon's integration with other protocols from the Global Alliance.

There was a proposal to extend Beacon's model to accommodate partition information, encompassing pathogen metadata and genomics, along with discussions around digital pathology and imaging.
The proposed process for addressing these topics involved forming smaller focused teams: SCOUTS, working on specific elements or issues outside regular Global Alliance meetings. The process includes drafting documents, implementing changes, integrating them into the specification, and finally, documenting the new features for users and developers.
Tony stressed the need to prioritise topics for these scouts due to limited manpower, with suggestions for involvement from both core Beacon working group members and participants from Driver Projects. The goal was to keep the number of active scouts manageable, rotating them over three to four-month periods.
Some topics, like extensions, are maybe less suited for the Scout approach due to their origins from different ongoing projects.

Many concerns and suggestions related to the extensive workload and expansion of the Beacon development group due to increasing popularity. Need to expand the Beacon developer network by involving other groups and entities to distribute tasks and enhance capacity.
Variant scouting and Boolean logic in Beacon queries: There was a proposal to include a subgroup from the Elixir structure variant group to handle the technical implementation aspects. The need for a clear understanding and documentation of the variant description, querying, and reporting processes.

Tasks like developing reference implementations and verifiers for Beacon important. Also the need to secure funding for ongoing projects.
Need to engage key individuals, such as Nicholas Blumberg and Tim Hubbard, who have knowledge and interest in Beacon-related work. Tim Hubbard's focus on trusted research environments (TREs) aligns with the idea of integrating Beacon for enhanced data discovery within these environments.

Meeting chat:
Computable cohort:
Kathy Reinold: Melanie, can you define what you mean by a computable cohort? Perhaps you'll cover it in your presentation.
Melanie Courtot: @Kathy Reinold I mean a representation of a set of individual level data ("virtual cohort") such that it can compute/re compute on the fly (eg if I dive down in phenotypes on the dashboard)
Jordi Rambla: Having a common request/query and a common response format must do the trick, w/o requiring the same format in the backend database
Melanie Courtot: A type of summary-level representation to preserve participant privacy and prevent reidentification while allowing researchers to assess if the virtual cohort is of interest to their research

Kathy Reinold: Does this assume a common data representation/model at the individual level data? (Seems like it's based on phenopackets)
—--------
Aggregated responses:
Michael Baudis: Aggregated responses will have to be provided both at the site of a central aggregator (e.g. a data management center w/ access to multiple protected, potentially international ... subsets) as well as through aggregation by a Beacon aggregator parsing multiple endpoints itself. Fitting our current Beacon network concepts.
Lauren Fromont: Why both? (Genuine question)
Jordi Rambla: I believe "aggregation" here also has the meaning of "not individual/row level" data.
Michael Baudis: Aggregation of multiple endpoints, some of which _only_ can provide aggregated data.
Melanie Courtot: Agreed - we need individual node aggregated data and also cross nodes
Melanie Courtot: I think so - which Is why we think this will go through PXF indeed
Melanie Courtot: But happy to discuss/hear other options
—---------
Tony Brookes: We're tackling pretty much exactly the same challenge in a large federated data platform for dementia. Options we've identified include:
1) creation of standard aggregation stats that are then centralised and/or kept locally, to be used to underpin display interfaces and/or discovery systems (potentially federated, using Beacon)
2) enabling real-time aggregation by processing real/derivative data - using centralised and/or local datasets, to again underpin display interfaces and/or discovery systems
These are not mutually exclusive.
Way forward depends on what 'nodes' are willing/able to do
Option 2 enables for CUSTOM not standardised aggregations!

Michael Baudis: @David actually interesting for cohort comparison etc. .... Obviously dreaming here.

Melanie Courtot: @Tony Brookes we're thinking along the same line - pre generated summaries (that can be reviewed for safety beforehand) as first step
Melanie Courtot: Then ideally something custom/computable
Michael Baudis: Actually +1 to aggregate against the PXF model.
Michael Baudis: Well, PXFshould provide prototype examples. But you can aggregate individual.age_at_sampling etc.
Tony Brookes: That's the same path we intend to follow :-)
I expect the standard aggregations will serve most purposes, but for sample discovery and request, we may need to go further - perhaps even to record level discovery (for which we already have the tooling)

Melanie Courtot: PXF also allows us to run ‘standardized queries’ maybe - eg “how many participants have recurrent cancer within 5 years?’ And you could run that against each cohort to know if it contains the data you are seeking
Melanie Courtot: (Where that specific summary may not be part of the ‘precomputed’ summary or ‘aggregated PXF’)
Melanie Courtot: To me both - while we don’t necessarily need to standardize the first one, it would be nice for comparison. As we seem to be building this anyway across projects, maybe a good opportunity to come together?
Gordon Krieger: I assumed part of the point of standardized aggregation was federating queries, and getting a standardized response

Michael Baudis: I suggest David & Melanie follow this up w/ some direct examples which would work for their use cases.
Melanie Courtot: Thanks for listening - happy to help where we can and/or follow-up offline or in the ClinPheno group)
David Bujold, Mr: @Michael Baudis Yes, happy to do so, thank you!
Lauren Fromont (she/her): The FILTERS topic also echoes Tony’s suggestions in the following agenda point

Michael Baudis: Typed variant queries are missing (or as a subset of the variant scouting ... not only SV).
Lauren Fromont (she/her): Variant scout with subsets could make sense (to avoid too many scouts)
Michael Baudis: Yes.

Lauren Fromont (she/her): Some things we should agree on: 1- which topics _should_ be scouts; 2- what are the priorities

Melanie Courtot: Thanks @Tony Brookes I was waiting for you to give us a spot ;)
Melanie Courtot: (If suitable for this WS)
Tony Brookes: ...and happy to work with you on it :-)

Michael Baudis: Potentially adding "filterGroups" as additional parameter.... keeping simple filters more or less the same.
Lauren Fromont (she/her): … and support to deployers!!

Lauren Fromont (she/her): Side note: documentation would be transversal to all topics/scouts (instead of a separate one)

Global Alliance for Genomics & Health

2023-12-12

Chairs: Anthony Brookes
Attendees (affiliation): Beatrice Amos, Oriol Lopez-Doriga Sagales, Dmitry Repchevsky, Sergi Aguilo, Manuel Rueda, Gordon Krieger, David Bujold, Maria Yazykova, Olesya Melnichenko, Alex Tsai, Ian Fore, Mamana Mbiyavanga, François Ancien (Université du Luxembourg)
Apologies: Lauren Fromont, Jordi Rambla, Michael Baudis, Nara Sobreira

	Actions Arising	Assigned To
1	Beatrice and Lauren to meet and tidy up spreadsheet that lists potential discussion topics and meeting plans for 2024, and share it with the leads (initially)	Beatrice and Lauren
2	Group leads (and Beacon community) to review roadmap at a subsequent meeting to review progress on deliverables and plan for 2024	
3	Explore creation of subgroups within Beacon to delve into topics such as Boolean Logic in Beacon filters (Ask Maria to lead this effort - include Sergi, Dmitry - use comms to circulate new initiative?)	

	Main Decisions
1	Tony has suggested that the Beacon team should do more tidying up and clarification of the Beacon specification before building too much additional functionality on top of it
2	Subgroups should be formed to work on specific topics in more depth, based on interest from the group members, e.g. Boolean logic in Beacon filters
3	Can we identify areas for grants or additional funding - need to articulate these, assign person, timeline.

	Agenda Item	Person/Time
1.	Welcome and confirm agenda: - GA4GH & PHA4GE Infectious Disease Community meeting - feedback? - Steering committee meeting, Jan 31st - WS updates meeting - Connect meeting in Ascona, April 21st-24th	Beatrice

	- **4 Beacon meetings remaining** - Driver Project updates	
2.	Review of previous action items: - Beatrice and Lauren to take the list of topics discussed and put them in the agendas for the next 6 months meetings. - For the next meeting, discuss organizing the GitHub issues and talk about the Beacon scouts meeting on copy number variation - Look into reviving some Beacon scouts to explore topics like metadata discovery in more depth	Beatrice
3.	~~David Croft - presentation: Stratification of Beacon~~	~~David Croft~~
4.	Security in Beacon	
5.	Maria Yazykova (Microsoft Health) **presentation:** a proposal for handling logical expressions within Beacon filters	

Link to meeting recording:
https://us02web.zoom.us/rec/share/6e0rsLdpX9TRJLWGdQYUnTEfLxKiC4T0jHpPFfIxTPaiZabTR76QwGGvD8a6vAvH.6CB1fd6XU2bLNjCB

Meeting summary:
Proposed deliverables shared and topics for future meetings - BA will share the spreadsheet outlining the proposed topics, speaker invites, and discussions for forthcoming sessions.
Tony says there's a need for **engagement strategies for new driver projects**, a **reference implementation** for Beacon, and **improved documentation**. There were mentions of integrating Beacon with FHIR (Dimitry) and the need for support tools for easy implementation.

Tony expressed concerns about the development and support needed for Beacon v2, highlighting its imperfections (how do we solve this?), the specialized implementations around it, and the necessity to address its underlying framework. We need to explore **funding opportunities** to address some of the deliverables listed. Many Beacon participants are over-committed, and there is a risk of losing confidence in Beacon's delivery, if no development. BA will share the documents for review and consolidation to prioritize tasks and gauge community interest and capacity for addressing challenges.

Significant effort is required, and we need a consolidated approach to strategic planning and resource allocation. The scale of work needed and the challenge of balancing aspirations with realistic capacities was mentioned by Tony.

Global Alliance
for Genomics & Health

Discussion of the various aspects and challenges regarding Beacon initiatives:
Dmitry mentioned the multitude of topics within Beacon, including specialized beacons, format support like FHIR, and the need for a centralized system for informed decisions.
Tony expressed concern about the current implementation of Beacon, citing compatibility issues arising due to individual platform choices. We need standardized formats for Beacon and address the imperfections within the current framework.

Do we need to explore improvements in specification, data, and implementation within Beacon? Maybe we need the simplicity in the standard and easier data mechanization. Maybe it's important to make Beacon more flexible for easy integration with various platforms.
Oriol mentioned that it's really important to think about verification within Beacon, the need for different plugins for diverse data types, and ongoing projects related to Beacon implementations in various areas like cancer registries and medical data. Security concerns and starting with small data sharing to understand the direction for Beacon was also highlighted by Oriol.

Tony talked about crucial strategic discussions regarding the future of Beacon initiatives. Tony raised the question of whether Beacon aims to be a universally usable standard for global interoperability or a framework more suitable for private consortia to build their own versions. Tony suggests that perhaps we need to look at achieving a singular comprehensive version of Beacon that seamlessly integrates globally might not be feasible. Perhaps we need to think about the idea of multiple implementations catering to different domains or use cases? These implementations would focus on specific areas like metadata discovery, OMOP common data models, or specialized networks for different medical conditions.

Need to collaborate and define these domains naturally as the community evolves. The goal was to identify common interests among groups to ensure compatibility and avoid disjointed solutions within similar domains.
We also need to enhance the current Beacon framework, addressing imperfections, functionality, and documentation. A strategic approach involving leadership discussions, community involvement, and potential grant applications was suggested to propel Beacon's evolution.
Regarding **security concerns**: while ensuring security is crucial, it might not fall solely within Beacon's scope. The governance related to data security and the importance of not complicating implementation with stringent security measures.
While **security is pivotal, making it too stringent within Beacon might deter implementers**. Handling security might involve a separate model or domain of consideration rather than being directly embedded within the Beacon protocol.

What is the role of Beacon and what are its limitations? We need to narrow down the scope of Beacon's responsibilities. Dmitry discussed security and the use of synthetic data, exploring the idea of generating synthetic data as a substitute for real data in certain contexts, though it raised concerns about its reliability for precise analysis.

Global Alliance for Genomics & Health

Maria Yazykova presentation
Maria shared a proposal for handling logical expressions within Beacon filters. Maria presented a proposal to enable logical expressions, particularly "AND" and "OR" operations, within Beacon filters. This proposed change involved creating a new object called "filter in normal form," allowing for more advanced filter logic while maintaining backward compatibility with the old implementation.
The presentation detailed the structure of the proposed changes and examples of how these logical expressions could be utilized within Beacon filters. She mentioned that this proposed alteration wouldn't necessitate substantial modifications in the current implementation and could facilitate more advanced filter logic within Beacon.

There remains an open issue regarding handling complex logic that needs further exploration. Maria proposes providing a direction for potential advancements while considering backward compatibility.

Boolean logic within Beacon filters, and thoughts on its implementation and practicality. Maybe Jordi, Michael et al needs to be involved in this as well. Invite Maria to share again in Jan? Different sub group discussed.

Is Boolean logic essential for Beacon and, if so, to what degree of complexity it should be implemented? Diff opinions on whether Boolean logic is necessary; Tony suggests that a **subgroup should evaluate its necessity and potential implementation** - maybe form a task team or a subgroup to examine Boolean logic across filters, with the understanding that the decision to adopt such logic would depend on the elegance and efficiency of the proposed solution.

2023-11-14: Beacon planning meeting

Chairs: Anthony Brookes, Lauren Fromont, Jordi Rambla
Attendees (affiliation): Beatrice Amos, Marion Shadbolt (CanDIG-UHN), Sergi Aguiló (BSC), Dmitry Repchevsky (BSC, ELIXIR.ES), Nicolás Gaitán, Oriol Lopez-Doriga Sagales, Michael Baudis, Maria Yazykova, Nara Sobreira, François Ancien, Gordon Krieger, Olsya Melnichenko, Evan Christensen, Alex Tsai, Rodrigo Martin

Papers of interest this month:

Global Alliance for Genomics & Health

Ensembl have implemented the GA4GH Beacon v2 (36)
https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkad1049/7416379

	Actions Arising	Assigned To
1	Beatrice and Lauren to take the list of topics discussed and put them in the agendas for the next 6 months meetings.	Beatrice and Lauren
2	For the next meeting, discuss organizing the GitHub issues and talk about the Beacon scouts meeting on copy number variation	
3	Look into reviving some Beacon scouts to explore topics like metadata discovery in more depth	

	Main Decisions
1	Adding new topics like metadata discovery, model extensions, and file entity types to the list for exploration in the near future
2	Considering piggybacking a documentation sprint onto the next Beacon Connect meeting in April to make progress on documentation
3	Discussing reviving Beacon scouts to help explore some topics in more depth than can be done in monthly meetings
4	Prioritizing documentation, cleaning up existing issues, and exploring metadata as things that could potentially be implemented in the next year

	Agenda Item	Person/Time
1.	Welcome and confirm agenda: Reminder: **GA4GH & PHA4GE Infectious Disease Community**: Monday, November 27th, at 8pm UK / 3pm ET Presentation by the **PathsBeacon** team about their implementation of GA4GH standards in an infectious disease context. Zoom link: https://us02web.zoom.us/j/89920537485 Meeting ID: 899 2053 7485	Beatrice
2.	Driver projects update Onboarding of new drivers / interest groups etc. for Beacon engagement: technical metadata has to be looped in, to use & improve the Beacon model in	All

	their discussions	
3.	Priorities for Beacon (will be part of the roadmap discussion): Features: - QBE / Variant matching - OR Expressions - Handover Types - OpenAPI - Beacon Network - Documentation/Cleaning: - /info design (what exactly is and is not meant by "I offer filter X") - House cleaning - [LF] Documentation needs a total review and update (could go with "cleaning", but should be explicit) - [LF] Ensuring scalability for support to beacon deployment / development - Explore: - Aggregated data - Beacons for metadata -> mostly a documentation issue? - Files as entry type for integration for computing - Extensions: - RNAget - Pathogen	
4.	[Technical] Any GH issues needed attention? Tony: the need to 'house clean' based on experiences to date. 33 open GitHub issues: https://github.com/ga4gh-beacon/beacon-v2/issues Suggestion to carve out time during each meeting to tackle one or more GitHub issues (Oriol to spearhead). How to decide? Order by priority? Some may be solvable in GitHub, some may need further discussion.	All
5.	[Documentation] How to move forward - Contentathon? - Involvement of GA4GH staff specialized in technical documentation? - Other suggestions? Documentation & how to get more involvement here (MB made some incremental additions following last week's discussion w/ Gemma & Jordi at the CRG but the main point here is to have this factored out to some "fresh eyes" w/ experience in doing such things	Michael, Lauren

	OMOP Beacon is not a technical target but should be documented/findable as one of the options one has when developing. Similar for other examples => documentation of implementation flavours, instances -> see documentation… GA4GH - Alex Tsai will raise with tech team	
6.	[Documentation] Suggestion ELIXIR Beacon - ideas of how to align V1 vs V2 requirements: Can we work with ELIXIR? UK node? Joint effort on documentation? (Training platform?)	Tony
7.	Roadmap - 2023/24: Review the previous roadmap document. Need to update/refresh this with our 2024 objectives (see priorities above)	

Link to meeting recording:
https://us02web.zoom.us/rec/share/TcRRq5qZ949vZhT-T8swaiBYzoDPFQdpY7DuJ0sAZKp9vh76DCkyV8wYZwPWA-NB.-R9jzd2HHphpHTFE
Passcode: G&VhL*$9

Meeting summary:

Driver Projects:
Beatrice plans to **introduce new driver projects** that mention Beacon, welcome them to the group, and discuss how to support their use of Beacon and plan for the next version over the coming weeks.
Some Driver projects already with involvement. Jordi mentions several and **requests Driver Project presentations,** including IPCHiP, RIFGC and Human Pan Genome Project.
Jordi requests examples of case studies for indexing and filtering, with similar results but not identical.

Improving documentation:
Jordi suggests documenting clarifications and examples to improve the Beacon framework, while Tony emphasizes the **need for better documentation** and improvements to the basic design of the standard.
Michael acknowledges inconsistencies in the documentation. Need a concerted effort from others to fill gaps.
Lauren suggests **involving knowledgeable writers** in technical documentation to improve quality and efficiency.

Global Alliance
for Genomics & Health

Tony mentions **potential collaboration with ELIXIR** to help them understand Beacon v2 and extend their work from the previous Beacon project.
Tony suggests a **3-month project to update documentation and training materials for Beacon**, with potential partners from the UK.
Lauren/Beatrice: would Alex Tsai be willing to put effort into the project? Alex to **bring to Jimmy and the tech team** to add to team sprint.
Tony (in chat): I think it is much MORE than just documentation. If we only do documentation (as Jordi insists), we'll still have defects and critical imperfections in the basic standard.

Prioritizing and addressing open GitHub issues:
Jordi: Need to discuss issues in GitHub and find solutions, with a two-fold strategy of dedicating a section in Global Alliance meetings for minor issues and disentangling them.
Need to prioritize GitHub issues and agree on which ones to address first.
Suggest Oriol? **should lead in prioritizing and wrapping up open GitHub issues**.

OpenAPI, metadata discovery:
Jordi: Need to **address inconsistencies in openAPI**; need for more documentation on openAPI and its potential impact on Beacon itself.
Tony: metadata discovery important in real-world projects, believing it to be a massive topic that will become crucial for finding data via attributes of records.

Tony: important to **provide descriptive metadata and aggregate data profiling** of datasets to facilitate discovery and querying without querying individual records.
Jordi mentions the need for better documentation and cleaning of datasets.
Tony suggests **focusing on basic descriptive metadata**, identifying which models are viable, and sending a political signal to the industry that Beacon is interested in working on metadata discovery.

Beacon enhancements:
Jordi suggests reigniting Beacon Scouts to address structural barriers in documentation.
Explore adding pathogens and file entity types to Beacon v2 model.
Michael proposes a catalog of recommended file types for Beacon.

Data filtering: we should prioritise standards for interoperability in genomics, including numerated data and Beacon network extension.
Potential for the use of filtering terms in a genomic variation context, including the potential for scopes on filters and the importance of clear documentation.

Roadmap and Beacon contributors:
Jordi suggests **adding GitHub issues to the agenda** for the next meeting to organize and prioritize tasks.

Global Alliance for Genomics & Health

Michael suggests that the team should **involve more people** in shaping the Beacon project, rather than relying on a small group of individuals.
Lauren proposes adding an additional half-day to the **next Connect meeting** in Switzerland to allow for more participation in the Beacon project.

2023-10-10: Cancelled

2023-09-12: Connect Planning

Chair: Lauren Fromont (CRG)
Attendees (affiliation): Jordi Rambla, Anthony Brookes, Beatrice Amos, Nara Sobreira, Olesya Melnichenko, Maria Yazykova, Valentin Grouès (elixir-luxembourg), David Bujold, Jordi Rambla, Tim Beck, Gordon, Maria Yazykova, Miguel Santos, Dmitry Repchevsky, Gordon Krieger, Sergi Aguilo Castillo

Apologies:

	Agenda Item	Person/Time
1.	Welcome and confirm agenda	
2.	Preparation and review of Plenary / Connect - Beacon filters - Beacon extensions - Informal PxF and Beacon discussion (continuing the Beacon filters session, as originally planned)	Beatrice, Jordi
3.	[Technical] Any GH issues needed attention?	All
4.	**Olesya Melnichenko** and **Maria Yazykova** (Microsoft) presentation	mariayaz@micr… olesya.melniche… *10mins*
5.	**Nara Sobreira** presentation Matchmaker Exchange, (Data Connect), Variant Matcher, Beacon	nsobrei2@jhmi.… *10mins*
6.	[Documentation] How to move forward - Contentathon? - Involvement of GA4GH staff specialised in technical documentation? - Other suggestions?	Lauren

7.	AOB ● Connect / Plenary **next week** ● Next meeting: October 10th (Lauren missing)	

Link to meeting recording:
https://us02web.zoom.us/rec/share/LjVXAtIBtXTjJ91HX44IMWpn2nRWw5ViSAbdQHnEMhKry78RVCa_1kCslDsmlEwU.h8c24LulL1evX2iG?startTime=1694526645000
Passcode: &W?11hB!

Minutes:

Plenary Connect meeting discussion:
Beacon-related sessions at the meeting: extensions and filters.
The main agenda items included discussions on Beacon extensions, filtering terms, and upcoming meetings.
BA - for next Connect suggest scheduling discussions with presenters before session planning. Conflicts.

Maria Yazykova (MY) presentation:
Maria discussed their exploration into creating a query format for requesting access to biomedical data. Showed a **simplified problem statement**, where a researcher needed access to data held by a custodian, but had limited access initially. Explored the idea of using a natural language approach for creating access requests but realized the need for logic, such as AND, OR, and NOT operators.
MY demonstrated the **importance of limiting variations** in fields and providing a dictionary of concept keys to ensure consistency. The **goal** is to make the results of queries easy to process in a format like JSON and interoperable across languages and technologies.
The presentation described the creation of a query format using a conjunctive normal form (CNF) approach, where include and exclude criteria were combined using AND and NOT operators.
MY **plans** to introduce operations into the criteria and incorporate information from data summaries, analytics, and research projects.

Maria's system is a structured approach to queries, focusing on logic upfront. Question about the complexity introduced by "OR" logic and whether Maria's system could be integrated into the beacon framework or if the beacon should adopt elements of her approach.
Group would like to discuss this further at a subsequent meeting.

Nara Sobreira presentation:

NS demonstrates the evolution **of the Matchmaker Exchange (MME) system** and its transition from a matching system to a query-based system, particularly in the context of genomic variants.
Initially, MME was designed as a matching system where users would submit patient descriptions, and the system would find similar cases. MME is shifting towards **becoming more of a query system**. Users can now express their queries as beacon queries, specifying genes or variants of interest. The system then finds matching accounts or individuals based on these queries.
The key difference is that MME aims to provide contact information for the individuals or accounts that match the query, enabling further investigation or collaboration. However, at present, the system does not match on features or the Human Phenotype Ontology (HPO) tree; it focuses primarily on genomic variants.
TB - question about whether the system considers exact matches only or explores similar terms.
NS - for now, matching is based on exact variant names, and feature information is shared but not used for querying.

2023-08-08: Connect Planning

Chair: Michael Baudis, Tony Brookes

Attendees (affiliation): David Bujold, Beatrice Amos, Alex Tsai, Dmitry Repchevsky, Diana Lemos

Apologies: Jordi Rambla (CRG), Lauren Fromont (CRG), Jordi Rambla

	Actions Arising	Assigned To	Deadline
1	Reviewing of Beacon issues and FAQ -> contributions…		
2	Feedback to Giselle about content of Beacon map		
3	Planning of vertical integration projects for GA4GH tool use…		
4	Beacon forward planning ● schema versions ● core Beacon domains (genome beacons, cohort discovery)		

Global Alliance for Genomics & Health

	Agenda Item	Person/Time
1.	Welcome and confirm agenda	
2.	Review of previous action Items: ● Connect ● Terminology issue	
3.	**Connect Planning**	
4.	Assessment of existing Beacons (version, status…) ● ELIXIR started a map from doc	
5.	[Documentation] Establish a plan moving forward	Lauren and Beatrice

Link to meeting recording:
https://us02web.zoom.us/rec/share/tA6A0btJKEH0Vv8fv5DRNWtxnLj44pzD6AABJMrkL1_veTuRyHeXpO70iJnBTZ9I.ooJdtUe9tt_EH0ai?startTime=1691503036000
Passcode: Y!52J#*s

Minutes (action items highlighted in blue):

- Beacon sessions:
  - Filters session (Tim, Tony); Phenopackets for query
  - Beacon extensions (RNAget, EJPRD, images …)
- Attending:
  - Michael (Tue - Thu)

Integrating Beacon with Matchmaker Exchange: possibility of missed opportunities if these efforts remain separate. Discussion with Nara, when she's back.
Vertical Integration of GFH Standards: Discussion about **coordinating vertical integration of standards**.
Planning for Presentations: Nara will present her ideas at the next Beacon meeting. (MB away)

Query Logic and Filtering:

- introducing query logic for filtering terms and combining variant queries for future Beacon versions (2.n?)
- Mention of ongoing discussions and proposals for introducing filtering logic: agreement that the practice of "implied OR for fiilters against same parameter" (w/o specific syntax) is expected (Tony, Michael…)
  - Followup: Michael updated documentation: http://docs.genomebeacons.org/FAQ/#boolean-logic and http://docs.genomebeacons.org/filters/#using-filters-in-queries

- Microsoft's Interest: Mention of Microsoft's interest in creating a query language with Boolean logic for genomics. Discussion about potentially collaborating with Microsoft on this topic. BA to invite Microsoft team to next Beacon meeting.

List of Existing Beacons: Alex is working on creating a list of existing beacons and their information, potentially integrating geographical information. The community is encouraged to contribute to this effort. Suggestions and contributions should be directed to Alex, Giselle, or Lauren.

Idea of creating a **global genomic beacon network** for improved data sharing and discovery. The conversation includes various participants discussing the challenges and potential of such an endeavor.

There's an emphasis on having a unified mission and coalition of organizations such as Elixir, G4GH, the Wellcome Trust, and NIH to support the development of a global genomic beacon network. **Need a clear mission statement and funding to support this effort**, as well as potential ways to integrate Beacon with other initiatives like FHIR (Fast Healthcare Interoperability Resources) and PACS (Picture Archiving and Communication Systems).

Discussed version compatibility, data format, and the need for better error reporting in aggregating networks. There's a focus on developing infrastructure, including **Beacon 3** (or 2.1?), and exploring the possibility of **publishing a paper** to raise awareness and garner support for the global genomic beacon mission. BA to add to roadmap for autumn.

2023-07-11 - Small group meeting

Attendees (affiliation): Sergi Aguilo Castillo, Dmitry Repchevsky, Jordi Rambla (CRG), Tony Brookes, Oriol Lopez-Doriga Saga

Link to meeting recording:
https://us02web.zoom.us/rec/share/UJ8XCLiL6uARlsYZRh7sP-h4QnhyeZmyCdnBmbxLODXmelpDCh3Fs9LH0E476ZCG.KKz3g5vqsi5S2uaf?startTime=1689084041000
Passcode: KhdF^0@0

Sergei Aguilo Castillo presents Beacon-OMOP:

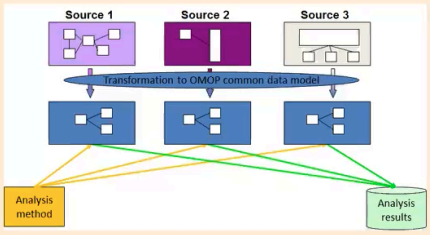
Require beta testers.

Global Alliance
for Genomics & Health

Beacon & OMOP CDM

What is OMOP CDM?

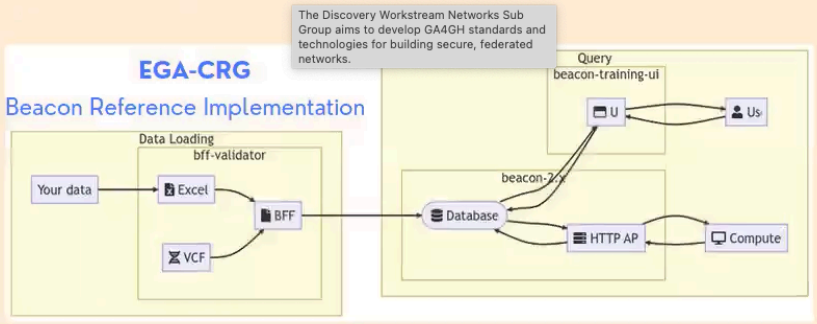
Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM)



- Merge disperse databases into a common format (data model).
- Common representation (terminologies, vocabularies, coding schemes).
- Widely used to share secondary use of research data in hospitals around the world.

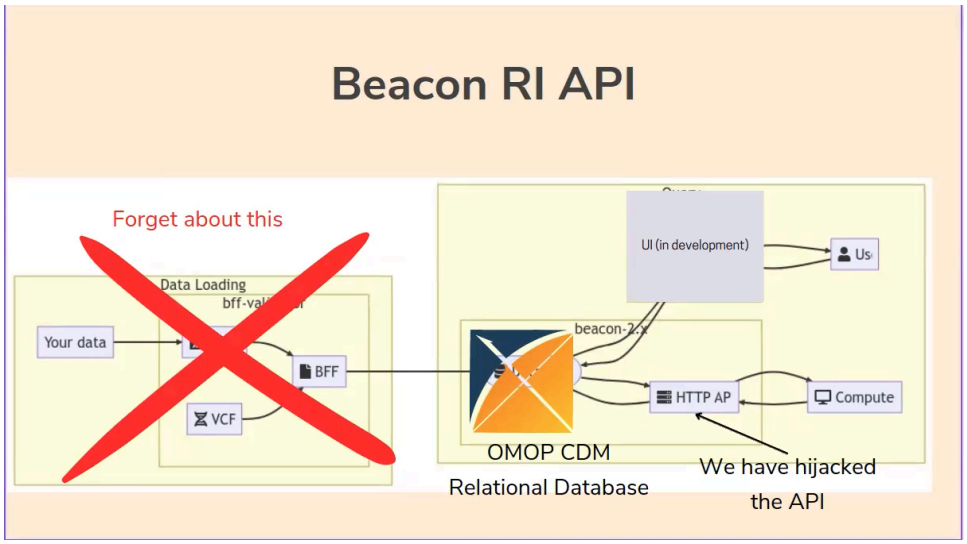
https://github.com/EGA-archive/beacon2-ri-tools

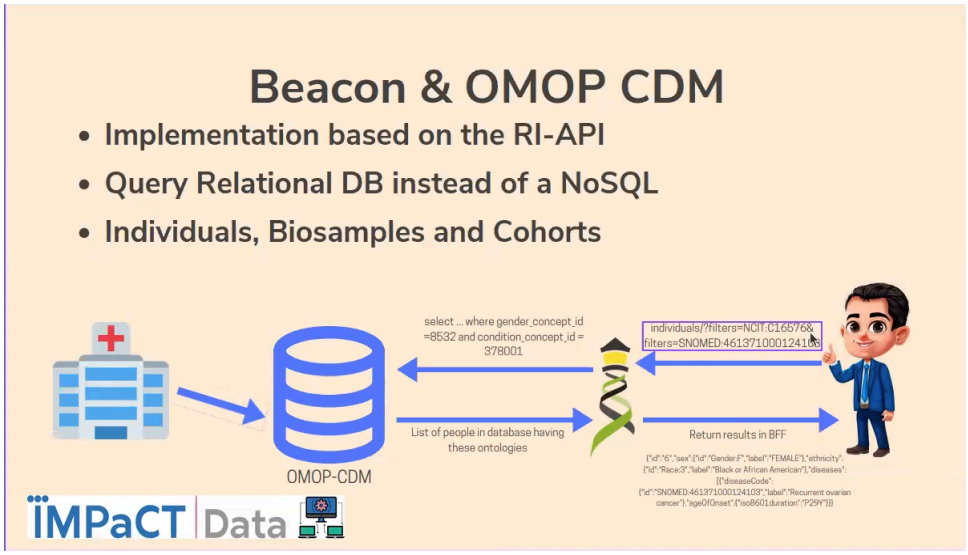
Beacon RI API

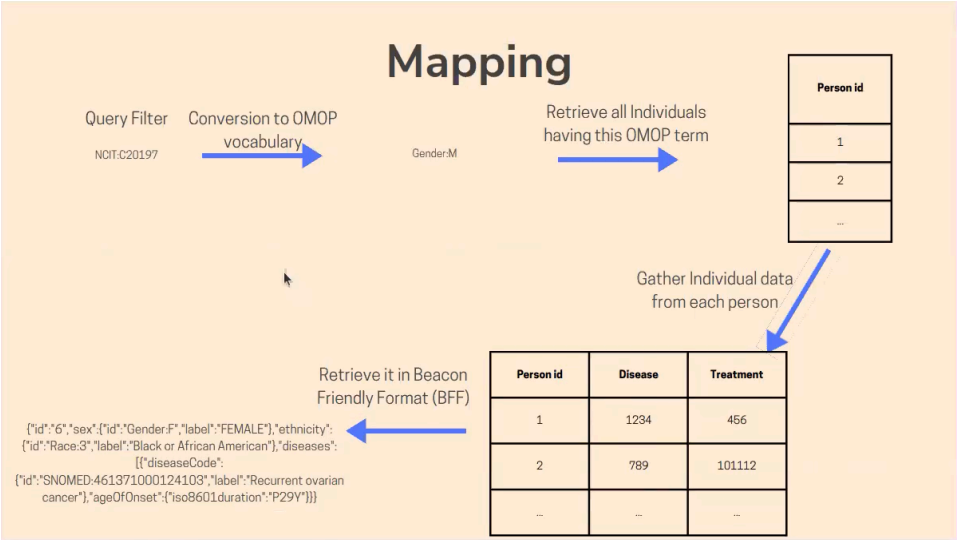


https://github.com/EGA-archive/beacon2-ri-tools

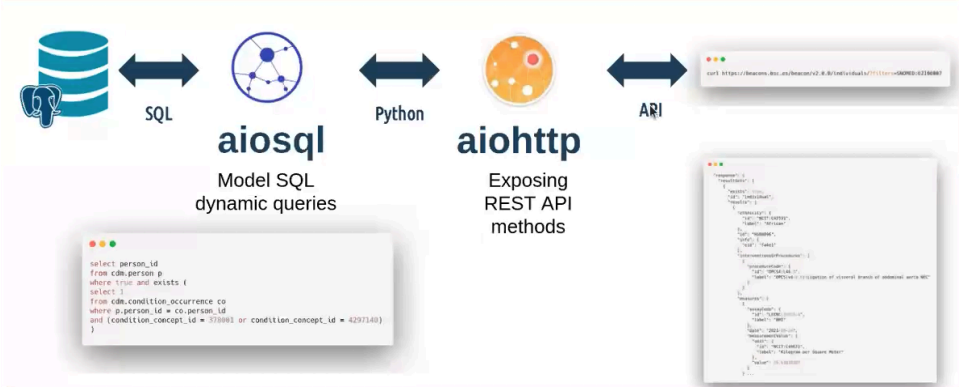
Global Alliance
for Genomics & Health







The filter is queried, converted to OMOP vocabulary, retrieve all Individuals having this OMOP term. Have mapped most of the individual properties from OMOP (including attributes: disease and treatment) and can retrieve in Beacon format.



Demonstration of a typical query, with output json, e.g. dynamic query to database:

```
select count(distinct person_id)
       from cdm.person p
       where true
        and (  ( gender_concept_id = 8532 )  )


               and exists (
                   select 1
                   from cdm.measurement co
                   where p.person_id = co.person_id
                   and (measurement_concept_id = 3012888 and
value_as_number >= 88.0)
           )
```

-End-

TB - OMOP has a lot of momentum - good that this does not use extensions. Universal solution. Advantages of adopting this version.
Currently working to build a data platform for dementia. Who takes responsibility for dealing with harmonization. For Discovery, using a syndication approach. Moving on to record-level discovery. Thinking of using OMOP.
What is the use case for returning data?
JR: for doing any queries, data needs to be returned. Service currently returns all data; subset of data (details) can also be returned if required. Record level data returned as opposed to analysis done on counts.
TB: 3 options if required to further interrogate data

1) permissions to access data will need to be sought.
2) Use of the negotiator tool, where you send a question to the source.
3) Approach to profiling the dataset, then request access to profile of data.

Plan is for datasets, use a data dictionary. For a sample, we want to interrogate at record level.
TB: OHDSI already has a rich set of tools for discovery APIs designed to work with OMOP and software and tools to support that.
JR (keep TB in the loop) will GA4GH follow-up: check if OHDSI (contact = Carlos - project manager from CRG/Eden, Patrick Ryan)

JR: starting to implement Beacon for cancer registries and Beacon for images. Growing adoption of OMOP. Good to keep ahead and with the flow and have Beacon as the discovery tool..
TB: would like to explore RDF for querying health data in the context of FAIR data. (The Swiss personal health network is going that way.)
JR: suggests 2 topics for Connect (please cc TB):

1) Session on filters and boolean logic
2) Beacon add-ons and extensions

2023-06-13: 6th meeting 2023

Chair: Lauren Fromont (CRG)
Attendees (affiliation): Alex Tsai (GA4GH), Beatrice Amos (GA4GH), Gordon Krieger, Miguel Santos, David Salgado, Michael Baudis, Gordon Krieger, Sergi Aguilo Castillo, Manuel Rueda, Dmitry Repchevsky, Andy Yates.

Apologies: Jordi Rambla (CRG)

	Actions Arising	Assigned To	Deadline
1	Connect: call for submissions	All	30-07

	https://docs.google.com/forms/d/e/1FAIpQLScs65R8Bu8_MbD_svlqsTFfmWILLRA7nDW6qpwSoYTQgOj3jA/viewform		
2	Tackle terminology issues — connect idea?		
3			

	Agenda Item	Person/Time
1.	Welcome and confirm agenda	Lauren
2.	[Stakeholders] ELIXIR AHM - Sessions well received including Peter and Angela, - Beacon is part of an explicitly growing collaboration between ELIXIR and GA4GH (standards implementations)	Lauren
3.	[Connect/Plenary] Session Ideas - San Francisco 2023	Beatrice and all
4.	[Doc] Beacon v2 documentation review	Alex (GA4GH) et al
5.	[Technical] Any GH issues arising?	Jordi, Michael?
6.	[Implementation] interpretation of filters	Dmitry, Sergi
	AOB ● Next meeting: July 11th 4pm CEST	

Notes:

Ad 4 (pre-call):

- Michael - I've streamlined & extended the documentation site (combined/updated introduction; categories for "Using …", "Creating…", "Development…" etc.). Please work on the pages (e.g. dev -> branch organisation; "data delivery -> biosamples …" from a user perspective etc.)
- Michael: Regarding the Beacon project site (beacon-project.io) we should
    - trim technical info as much as possible
    - focus on news and publications (Ana does this already)
    - add & constantly update lists of/extended pages about projects and resources w/ beacons; this was one of the "hard to answer questions" at last week's call w/ the NCI childhood cancer project

Minutes:

LF: ELIXIR feedback on Beacon. Beacon/OMOP collaboration planned over summer.

Global Alliance for Genomics & Health

BA: https://forms.gle/qqGRiiz441LgtyQP7 Connect session proposal form. Deadline end July, so BA/LF and team can brainstorm at/for the next meeting (July 11th). Please email/Slack group if you have ideas/suggestions for Connect.
MB please try to avoid presentations virtually as very difficult to hear and engage. How will the development progress over summer? Would be good to hear about implementations. Hot topic is currently filters.
AT: shares experience: created a small scale reference implementation (option 3). Found documentation easy to follow. Attempted option 1 (implementation from scratch) - found documentation was lacking. Hoped to see endpoints he could call. Unclear of how to integrate Beacon into API, and could not progress.
MB: For me a general approach is more about reverse engineering the expected responses. For me Beacon is about the messages; and the OpenAPI part has something to do w/ the technical implementation. (Chat)
LF: We should probably re-invite Tim and try to 1- come to an agreement of “guidelines” and 2- document it (chat)
BA: shared some of AY comments.
MB: writes documentation from the user’s perspective. Does not implement. Has a specific extended use case. Suggests perhaps it should be made more clear how to implement from scratch? Not in his remit to do a complete OpenAPI representation - can we include anyone to take this on? Suggests Beacon is not something you can implement based using OpenAPI endpoints. Difficult question to tackle.
DB: Beacon spec is very broad by design. As Beacon V2 extends, and more projects aim to use in complex situations, perhaps more constraints are needed to remain compatible?
MB: Beacon is about messages: queries and responses. Not about the technical implementation. Need to separate implementation and how messages should be structured. Reference implementation is important for testing datasets, and exploring the possibility of loading own data as a first step.
Confusion arises when taking the step from OpenAPI to query, and then adding filters etc. For this whole process, perhaps the documentation is lacking.
MB has done a lot of work here, and could do with support. BA - seek GA4GH tech support? Has focused on messaging. Combining this with the data model is highly complex.
LF: Beacon coordinators lack technical expertise to work on documentation, but commit time to review? Ana too? Action item: schedule meeting: BA, LF, Ana, AT.
Implementers interpret filters in their own way. No specific guidelines for using filters - Sergei/Dimitry. Tim would like to be involved.
DR: found many differences in the understanding of the specification of filtering, e.g. “types” had variations in definitions. Perhaps should be addressed in Beacon v2.1?
MB: Topic for Connect w/ phenopackets, Clin/pheno: recommendation for terminologies… (Chat)
BA - can we review time of meeting? MB cannot make this time.
Co-lead nominations in progress. Comms will be distributed in due course.

Link to meeting recording:

https://us02web.zoom.us/rec/share/nsLeNJ3zUVE2NxP6a_UAeB40Qm67wUDh2qK_pVO0lTMcRI06dNYcI40e5RbxGKAg.eq1gQJRUvNJ894qP?startTime=1686664601000
Passcode: zk?^5bc0

2023-05-09: 5th meeting 2023

Chair: Lauren
Attendees (affiliation):

	Actions Arising	Assigned To	Deadline
1	Revision discussion on Github		
2	Beacon V2.1 Discussion		
3			

	Agenda Item	Person/Time
1.	Welcome and confirm agenda	Lauren
2.	[GA4GH roadmap] Connect discussions and conclusions - Beacon and Phenopacket - Discovery	Lauren / Michael
3.	[Technical] GitHub issues and PR review	Oriol / Jordi
4.		
5.	AOB: - Next meeting: June 13th	

Notes

- Connect and Beacon roadmap
    - Discovery

[MB] many things (Discovery) are either mature or with unclear maintenance/progression. Beacon is active. Some others did not take quite yet (e.g. some work on Schemablocks). We have to convince entities closer to central management, so Discovery may also be about how

schemas are accessed. Soon to be a new lead in Discovery. Hope that the implementation side will come more and more in focus (e.g. Beacon x PxF; DUOs, etc). Also how to make data accessible (e.g. ICGC, etc).

[JR] many mentions of Beacon. Could be lip service or serious plans, but even if we have half of those, it promises interesting work. Most of the requests we have received are already part of the standard, one way or another. So I think we managed to answer many of their needs. The standard is 1 year old and has been able to progress significantly.

[MB] I have this commitment from Melanie to test some things especially on the PxF side of it. Also there was a discussion about whether we should make Beacon more “Phenopacky” or stick to Beacon model and ensure some level of interoperability (probably best option).

[Anthony Brookes] Actually really impressed on how people have or are adopting Beacon, we are really changing the world. It’s exciting, but also terrifying. If people have problems, they will turn to us for answers. There are several projects that should be considered as specialisation of beacons. This is a challenge that could become massive. Another one is the next version of Beacon, where we must gather requests from the community in order to start prioritising changes for v3. Those two challenges are complementary.

[JR] I believe that at some point we should have a community of beacon users (meaning: deployers?). It would make more sense to have a group when there is a critical mass, in order not to “pollute” the main API group here.

Another group that we are starting is Beacon for images (part of EUCAIM), and Beacon for cancer registries (EOSC4Cancer).

- Specification

[JR] Have been training the EGA team on the specification. Would love to hear thoughts on Spec.

[MB] have been modifying the documentation section in order to have clearer branches.

[Oriol Lopez] we encourage people to contribute to these topic branches (http://docs.genomebeacons.org/contribute/; *“https://github.com/ga4gh-beacon/beacon-v2/blob/website-docs/docs/code-organization.md”* https://docs.genomebeacons.org/code-organization/). From my side, I’d say that we should identify the milestones that would allow for an improved v2.

[JR] Yes, need a v2.1 first

[LF] anything re: network?

[JR] yes, most issues have already been addressed. We don't have endpoint, for aggregators, with the beacons in it

[MB] I think this should be different

[Dmitry Repchevsky] returning set of responses in info for now. Not clear how to return it as a dataset

[MB] should not return the whole info, just enough so people can retrieve them.

[DR] BN does not only aggregate the results, but also the metadata.
[JR] we should look into the service registry, and see what is there. Then we can decide how we want the response to be.

[MB] also a good point about reactivating topics in Discovery. Beacon is now driving the use of Service registry.

[JR] if the registry allows for additional properties, then it could be interesting, but we would need to look at it (not super fresh on that).

- Boolean / count

There is sometimes some inconsistency between boolean and count responses.

2023-04-11: 4th meeting 2023

Chair: Jordi Rambla
Attendees (affiliation): David Salgado, Michael Baudis, Jordi Rambla, Gordon Krieger, Sergi Aguilo Castillo, Manuel Rueda, Dmitry Repchevsky, Oriol, Amreen, Tony Brookes, David Salgado, Diana Lemos, Gordon Krieger

Notes

- Beacon V2 repository: Renaming the branches that structure into main and developed branches and has security features. When you do a pull request, you can see these branches. Started a discussion based on PR requests.
- Michael: Recalled what was being used by the original validator. Document repositories separate outside the Beacon tree. Examples that are outside the model can be moved there.

- Manuel: dereferenced schemas for models are up to date within the “bin/deref_schemast” directory.
- Jordi: You need to change and remove the additional properties in order to make the schema strict. It is a version of the schema. Having this in the central repo should be sufficient.
- Michael: Main branch must be protected, and Michael created a documentation branch as well as updating the active branch - website-docs. When you update the schema, you can use the website doc branch to can push the action to update the website accordingly. Website docs is not a protected page. Website docs are populated from a schema, from forced updates.
- Jordi: we need a protected branch for documentation for synchronicity.
- Michael: The schema document created by schema, with model and framework was created for code particles and inland documentations in the code files in the same repository.
- Jordi: How do we keep the documentation in branch with new versions?
- Michael: The documentation for developers is separate but the website should reflect what occurs as the current version in the main branch. If you update documentation in upstream and main, it should update.
- Manuel: From the moment that main gets updated, there should be a github action that covers. We need to brainstorm how to do this.
- Michael: If anything in schemas and models is updated, it should be pulled across the documentation branch.
- Jordi/Oriol: We will track what happens. We have moved all of the pull requests, but some of them are frozen and outdated. We need to know what to do with these requests. The old pull requests can be accessed by being pushed to the developed branch.
- Oriol: There are six pull requests that open and need to be discussed (we originally had eight, but two were copies). All others have been solved/closed.
- Michael: Everybody who's missing a PR should look in the closed ones & possibly recreate if the issue still has some merit.
- Jordi: Looking into new milestones for V2.1 and V2.2 should be the next priority. We are open to discussion. We are clear on PR requests for now.
- Michael: Look into what future branches we would like to have and combine with release tagging. We need a visible milestone discussion that will be posted on the website, so people can keep track of what is coming up.
- Sessions relevant to this group: Federated Variant Matching, Federated Compute, VA Spec, Discovery Roadmapping,
- There was no developer infrastructure for MME and now we are in a position in which Beacon and Phenopackets are very document focused and MME is not as developed. There is big interest from different groups of technical implementation using Beacon.
- Tony: the main difference in the MME scenario, you are expected to provide details of a real world case and the service decides how to match that and determine how many matches are available. Their system can be adjusted to Beacon.

Global Alliance for Genomics & Health

-

2023-03-14: 3rd Meeting 2023

Attendees (affiliation) *please add yourselves* : Michael, Lauren, Jordi, Amreen, Tony Brookes, Sergi Aguilo Castillo, Diana Lemos, Gordon Kreiger, heimo, Jorge Oliveira, Manuel Rueda, Umar Riaz, Dmitry Repchevski

	Agenda Item	Person/Time
1.	Welcome and confirm agenda	Lauren
2.	[Technical] GitHub organisation	Oriol
3.		

2023-02-14: 2nd Meeting 2023

Attendees (affiliation) *please add yourselves* : Michael, Lauren, Jordi, Amreen, Tony Brookes, Sergi Aguilo Castillo, Diana Lemos, Gordon Kreiger, heimo, Jorge Oliveira, Manuel Rueda, Umar Riaz, Dmitry Repchevski

	Actions Arising	Assigned To	Deadline
1	Annual report to GA4GH (please comment)	everyone	27/02
2	Look into branches and add relevant ones (or branch out)	Jordi and Michael	14/03
3	Trigger discussions in GitHub	Jordi	
4			
5			

	Agenda Item	Person/Time
1.	Welcome and confirm agenda	Lauren
2.	[ELIXIR] Annual report to GA4GH (please comment)	Lauren

3.	[Technical] Status of Beacon API	Jordi
4.	[Coordination] How to keep the discussions offline	Lauren
5.	[GA4GH events] Beacon and PXF alignment at the next Connect	Lauren
6.	Documentation, Implementation guidance … (Galaxy feedback) ● confusion about Beacon v1 / v2 ● which Beacon network -> need one to point at ● clear implementation for reference implementation etc, "from VCF to Bv2" needed	Michael

Minutes

3. Annual report > offline AI
4. Status of Beacon API

[DR] Our implementation is for Beacon v2, but some people started to change the code to align to their data. Also trying to improve the VRS part.
[JR] I'm seeing the situation where we have PRs on specific items (not necessarily minor). I would like to see these as branches in the GH repo instead of PRs. My fear is that otherwise, we will start mixing things and losing coherence. What do you think?
[MB] I think the important thing is that people keep close to the code. (also see 6. below)
[TB] Question to Jordi: how many significant players are taking the core Beacon and trying to implement it? I'm afraid that we're not aware of all of this, and that not everyone is coming to us (this group) to discuss them.
[JR] My suggestion of branches could seem trivial, but was going in that direction. But we're not aware of any "new" player coming to us and implementing from the spec.

Agreement on using branches.
First set of branches:

- Beacon Network
- Privacy features / Thresholds
- Response simplification
- General fixes (incl. Filter use improvements)

Discussion topics: add **Query types / prototypes** + **Clean schemas**

5. Connect + alignment with other GA4GH

More than PXF?

Something with MME?

- [JR] Already presenting at tomorrow's meeting (Tim was listed, but not too many people)
- [TB] EJP + Solve-RD: created a Beacon spec called Vivify (please edit if name is wrong). It is Beacon specified to serve the needs of EJP. Not the same as MME, but of course results can come similarly. Asked Sergi: when does he think we should try to make the bridge? We should wait, get it alive, and then reach out to lots of groups. Tony's preference would be to reach out sooner rather than later (at least for them to know we're preparing something). Maybe we should take the lead to put everyone in the same group. If we're not careful, things could become very messy.
- [MB] used to be that for MME it was very tricky because of limited resources. Heidi pushing that MME is recognised as a standard. Of course this runs in parallel to discussions on how to connect with existing standards (e.g. Beacon and Data Connect).
- [TB] It's not about the API, but about the context/purpose. In MME, the DBs define what constitutes a match (they do not want people to start querying without having submitted a single case). It's very different from Beacon.
- [JR] would like to know more about the things that were difficult to access (because the repo and documentation have not moved for months). If it is not visible enough, then we can work on changing that. If you have more details, we can try to improve it.
- [MB] Another point was the Beacon network (also points to v1). It should not be there or have a big flag to be updated. >> should be picked up at the next BN meeting (or ELIXIR meeting)

6.
[MB] Was talking to Galaxy. They were astonished that v1 is not what Galaxy should use. They did not know where the v2 reference implementation was. People feel lost, and it was also difficult to find for me.

2023-01-10: 1st Meeting 2023

Attendees (affiliation) *please add yourselves* : Michael, Lauren

	Actions Arising	Assigned To	Deadline
1	Discuss how to find funding for Beacon in order to upgrade		

	Agenda Item	Person/Time
1.	Welcome and confirm agenda	
2.	Recap November Connect Session ● Some key points here	?

3.	Connect Session Ideas - London 19-23 April 2023	Fabio
4.	Beacon network ● Possibility to add a branch in Github where everything happens and *then* merge (will be version 2.1, for example), rather than touching the spec directly	Jordi
5.		
6.	AOB ● Doc review ● ELIXIR All hands, deadline for submission: January 23rd ● Creating another minutes document for 2023?	? Lauren Lauren

Minutes

- **Recap Connect**

2 workshops: Migration + Beaconize your data type (summary copied below)

Summary

- Queries to a Beacon rely on IDs - therefore prescriptive recommendations are important e.g. "Use RefSeq for Gene IDs"
- An API specification is a "plan for building a house", you specify which materials you use to build it
- The Beacon Friendly Format is not an official or even standard concept, it reflects an intermediary step for facilitating Beaconization of data
- Differences in approach - some implementations convert all data to BFF once, others have built tooling to do it on the fly
- Harmonizing labelling takes considerable time in Beaconizing data, especially clinical data - people are building tooling to convert/map it into Phenopackets/BFF see https://convert-pheno.readthedocs.io & https://github.com/phenopackets/omop-exporter
- Some members of the community, clinicians especially, feel that providing a specification without a supporting implementation severely limits uptake of Beacon
- While Beacon is very capable, it was envisioned as a *Discovery* tool, not for research. That should usually come later/separately.
- Encouraging a set of standard/minimal terms for data representation would help the growing network harmonize as it grows
- Collecting example data sets and usage would facilitate further uptake
- There already exists a number of backend data to BFF mappings and tooling, but this is not being shared in a centralized location next to Beacon
- The case is similar for documentation, CRG B2RI and Progenetix has some too, but this does not sit with Beacon standard or Starter Kit resources

[JR] There is a lot of information here, perhaps we should have a meeting focusing on those
[FL] Perhaps we can pick the topics (2-3) we would like to discuss in priority

[JO] Will Beacon encompass any authentication mechanism?
[JR] We rely on http for that. It is orthogonal to the Beacon itself. The only that is shown, is that in the configuration, you inform the default (public, registered, controlled). That is for the *specification*. In the RI, it needs to be refreshed.

[JR] being more or less prescriptive is one of the questions being raised. It seems that there is a need for the Spec to be *more* prescriptive.
[AB] Well then we need to know what to be prescriptive about. What is the actual use case we are trying to support then. Or actually, headline use cases
[JR] we can contribute to that from the EGA, as we are implementing Beacon for various domains.

[MB] Here is a project draft for phenopackets/Beacon+. Open to comments and suggestions

Summary

- Beacon can interpret queries in different ways, this means that each query type does not necessarily need its own data model
- We need greater feedback from (would-be) implementers on their use cases
- Beacon can be built in different ways - e.g. PHP on PostgreSQL or Python/Django on MongoDB
- Some groups have begun implementing Beacon's filter capabilities
    - This helps standardize query-response as pairs, but is considerable work
    - Solve-RD working to create "standard" filter set for consortium
- Docker instances have helped implementers understand the specification
- Again Discovery focus of Beacon - it doesn't allow chained Boolean response, so research questions are difficult e.g. cannot ask x<y<z
- No standard way to share "What queries/IDs/filters does this Beacon support?"
    - Info endpoint supports information, as in "Patient age", but not type, e.g. integer
    - Is there a risk of data leak with open sharing of Beacon query structure?
- Variant types is not prescribed, and no way to share this
- The Beacon documentation could be better and resources aggregated
- If a filtering term doesn't exist - a Beacon does not report missing term, only "0" results, so querying across multiple Beacons is impossible if the presence of filters varies
    - Again if you openly report which terms do not exist, this could be a data leak
    - Tony Brookes is collaborating on a potential solution to this
- A "safe" Beacon concept could be created, with input from Legal/Ethics teams
    - Limited queries etc. as *implementation* considerations