

# The GA4GH Beacon Protocol

## A Standardized Format for Federated Genomics Data Exchange



Michael Baudis | BC2 2021GA4GH Session | 2021-09-13

# The Global Alliance for Genomics and Health

## Making genomic data accessible for research and health

- January 2013 - 50 participants from eight countries
- June 2013 - White Paper, over next year signed by 70 “founding” member institutions (e.g. SIB, UZH)
- March 2014 - Working group meeting in Hinxton & 1st plenary in London
- October 2014 - Plenary meeting, San Diego; interaction with ASHG meeting
- June 2015 - 3rd Plenary meeting, Leiden
- September 2015 - GA4GH at ASHG, Baltimore
- October 2015 - DWG / New York Genome Centre
- April 2016 - Global Workshop @ ICHG 2016, Kyoto
- October 2016 - 4th Plenary Meeting, Vancouver
- May 2017 - Strategy retreat, Hinxton
- October 2017 - 5th plenary, Orlando
- May 2018 - Vancouver
- October 2018 - 6th plenary, Basel
- May 2019 - GA4GH Connect, Hinxton
- October 2019 - 7th Plenary, Boston
- October 2020 - Virtual Plenary, June 2021 - Virtual Connect ...
- October 2021 - Virtual Plenary ...

GENOMICS

*A federated ecosystem for sharing genomic, clinical data*

Silos of genome data collection are being transformed into seamlessly connected, independent systems

The Global Alliance for Genomics  
and Health\*

SCIENCE 10 JUNE 2016 • VOL 352 ISSUE 6291



Global Alliance  
for Genomics & Health

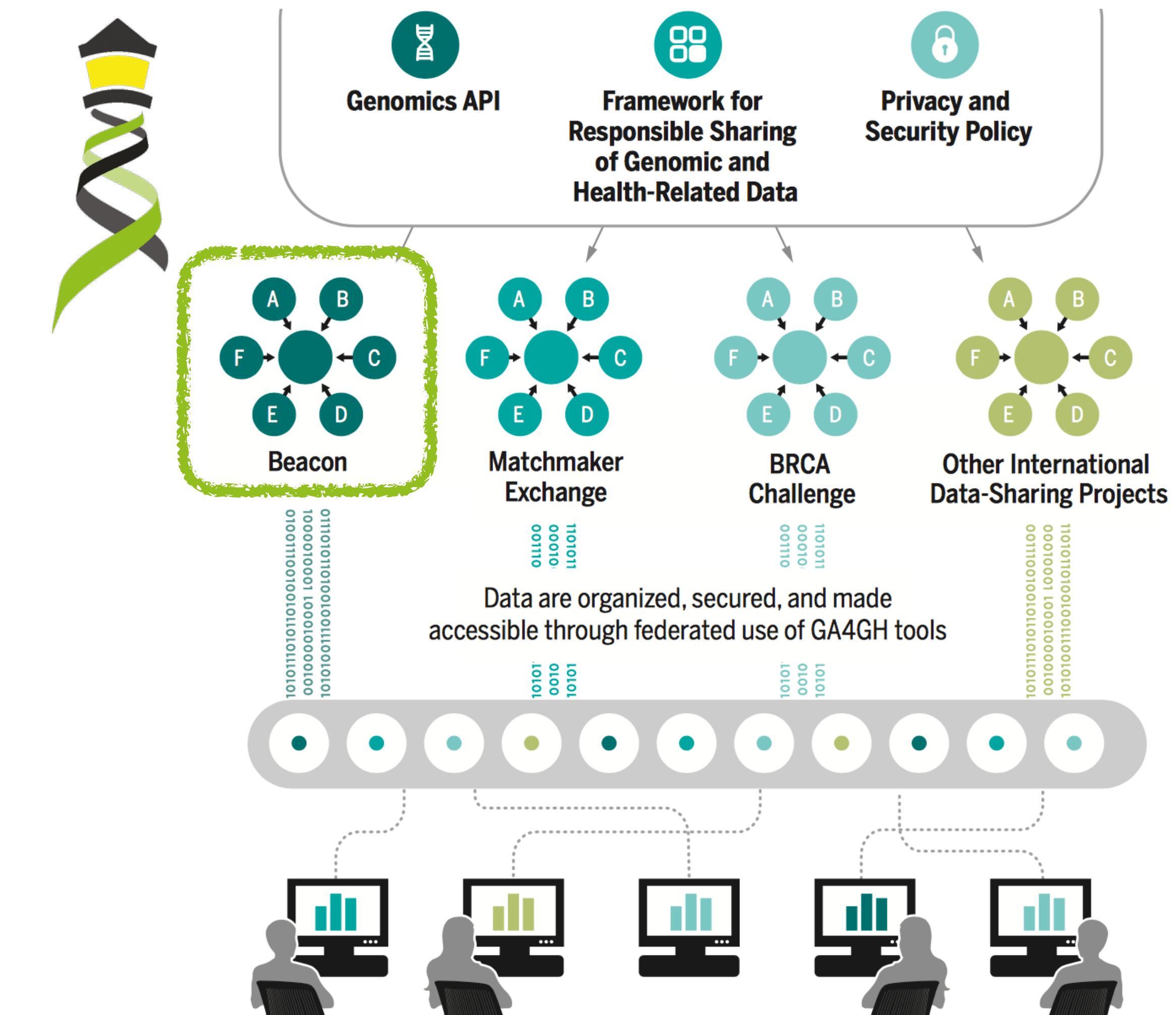


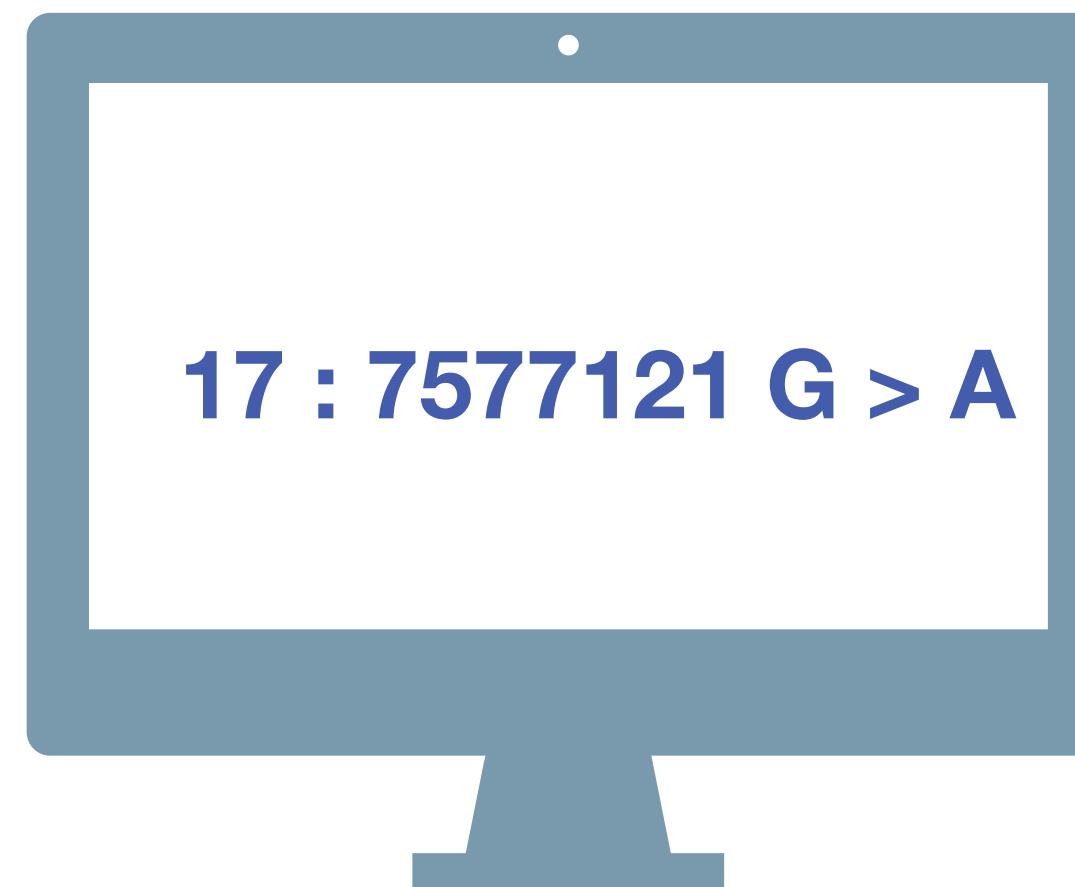
## GENOMICS

# *A federated ecosystem for sharing genomic, clinical data*

Silos of genome data collection are being transformed into seamlessly connected, independent systems

**A federated data ecosystem.** To share genomic data globally, this approach furthers medical research without requiring compatible data sets or compromising patient identity.





# Beacon

A **Beacon** answers a query for a specific genome variant against individual or aggregate genome collections

**YES | NO | \0**

## Introduction

... I proposed a challenge application for all those wishing to seriously engage in *international* data sharing for human genomics. ...

1. Provide a **public web service**
2. Which accepts a query of the form “Do you have **any** genomes with an “**A**” at position **100,735** on chromosome **3**?”
3. And responds with one of “**Yes**” or “**No**” ...

“Beacon” because ... people have been scanning the universe of human research for *signs of willing participants in far reaching data sharing*, but ... it has remained a **dark** and **quiet** place. The hope of this challenge is to 1) *trigger the issues* blocking groups ... in way that isn’t masked by the ... complexities of the science, fully functional interfaces, and real issues of privacy, and to 2) *in short order* ... see *real beacons of measurable signal* ... from *at least some sites* ... Once your “GABeacon” is shining, you can start to take the *next steps to add functionality* to it, and *finding the other groups* ... following their GABeacons.

## Utility

Some have argued that this simple example is not “useful” so nobody would build it. Of course it is not the first priority for this application to be scientifically useful. ...intended to provide a *low bar for the first step of real ... engagement*. ... there is some utility in ...locating a rare allele in your data, ... not zero.

A number of more useful first versions have been suggested.

1. Provide *frequencies of all alleles* at that point
2. Ask for all alleles seen in a gene *region* (and more elaborate versions of this)
3. Other more complicated queries



**"I would personally recommend all those be held for version 2, when the beacon becomes a service."**  
**Jim Ostell, 2014**

## Implementation

1. Specifying the chromosome ... The interface needs to specify the *accession.version* of a chromosome, or *build number*...
2. Return values ... right to *refuse* to answer without it being an error ... DOS attack ... or because ...especially *sensitive*...
3. Real time response ... Some sites suggest that it would be necessary to have a “*phone home*” response ...



# Beacon Project in 2016

An open web service that tests the willingness of international sites to share genetic data.



Beacon Network

Search Beacons

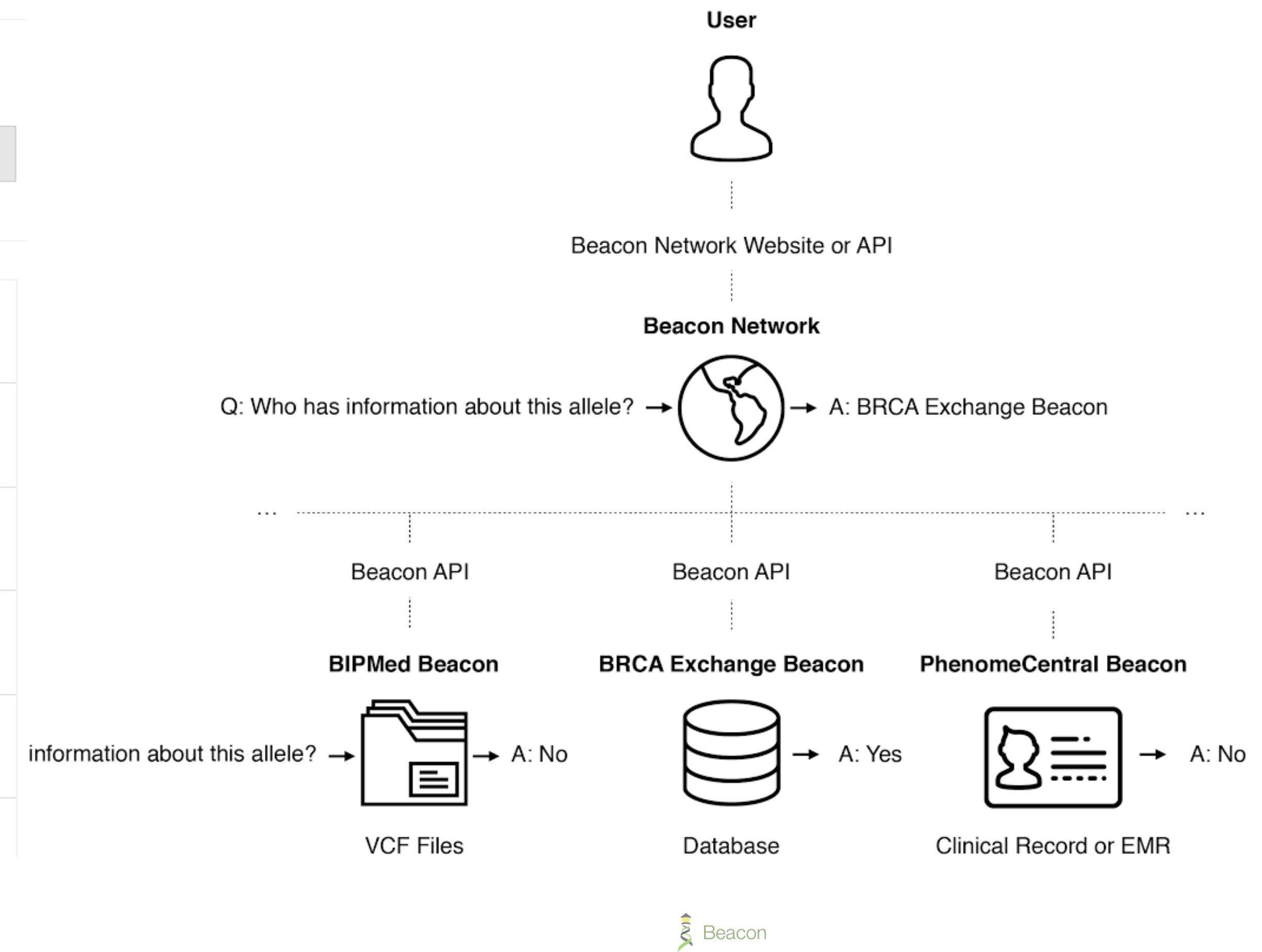
Search all beacons for allele

GRCh37 ▾ 10:118969015 C / CT Search

Response All None  
 Found 16  
 Not Found 27  
 Not Applicable 22

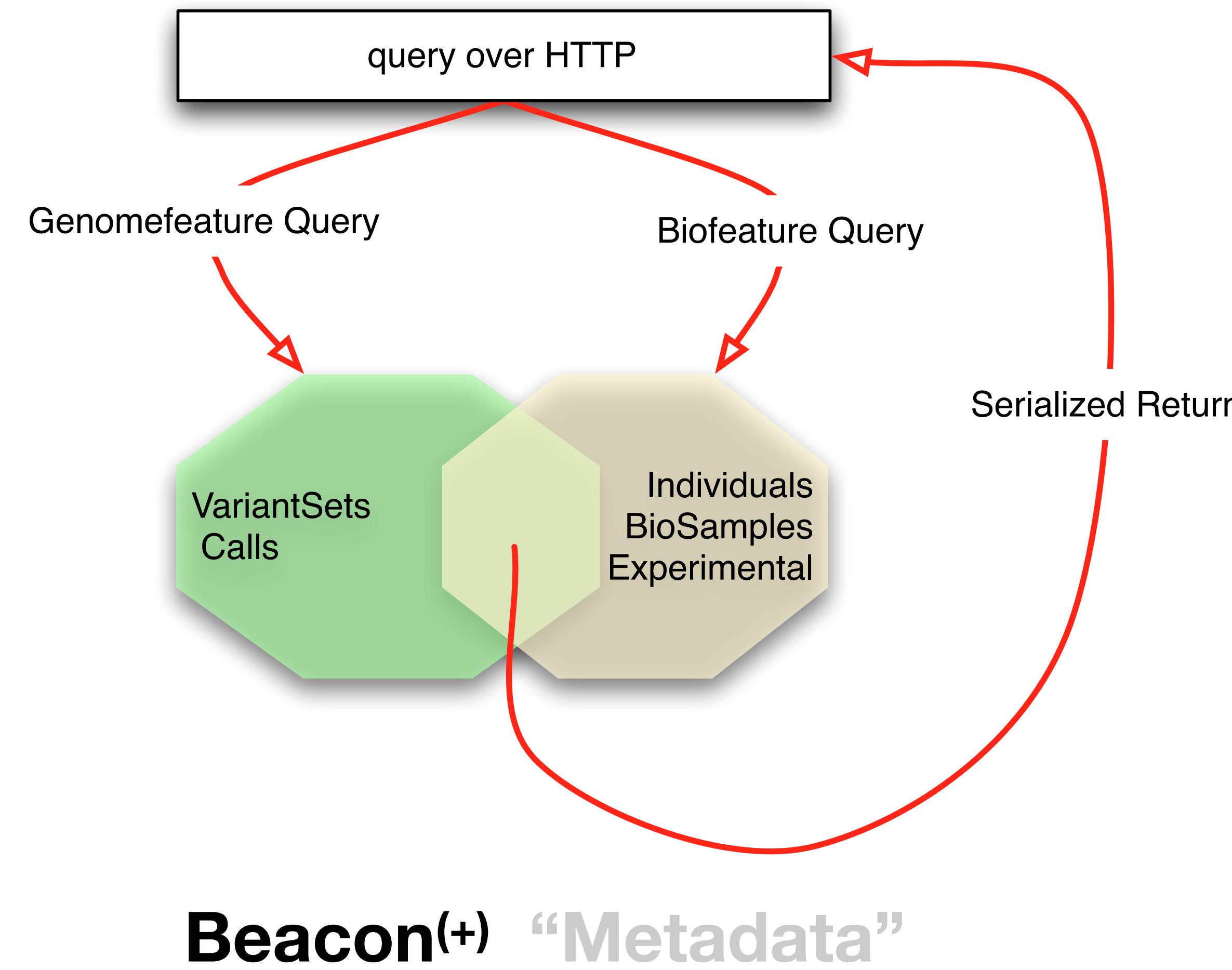
Organization All None  
 AMPLab, UC Berkeley  
 BGI  
 BioReference Laborato...  
 Brazilian Initiative on ...  
 BRCA Exchange  
 Broad Institute  
 Centre for Genomic R...  
 Centro Nacional de A...  
 Curoverse  
 EMBL European Bio...  
 Global Alliance for G...  
 Google  
 Institute for Systems ...  
 Instituto Nacional de ...

BioReference	Hosted by BioReference Laboratories	Found
Catalogue of Somatic Mutations in Cancer	Hosted by Wellcome Trust Sanger Institute	Found
Cell Lines	Hosted by Wellcome Trust Sanger Institute	Found
Conglomerate	Hosted by Global Alliance for Genomics and Health	Found
COSMIC	Hosted by Wellcome Trust Sanger Institute	Found
dbGaP: Combined GRU Catalog and NHLBI Exome Seq...		Found



Date	Tag	Title
2018-01-24	v0.4.0	Beacon
2016-05-31	v0.3.0	Beacon

# Minimal GA4GH query API structure



# ELIXIR - Making Beacons Biomedical



- Authentication to enable non-aggregate, patient derived datasets
  - ELIXIR AAI with compatibility to other providers (OAuth...)
- Scoping queries through "biodata" parameters
- Extending the queries towards clinically ubiquitous variant formats
  - cytogenetic annotations, named variants, variant effects
- Beacons as part of local, secure environments
  - local EGA ...
- Beacon queries as entry for **data delivery**
  - handover to stream and download using htsget, VCF, EHRs
- Interacting with EHR standards
  - FHIR translations for queries and handover ...

# ELIXIR Beacon Project

- Driver project on GA4GH roadmap
- aligns with Discovery Work Stream
- strong impact on GA4GH developments as a concrete, funded project

The screenshot shows two cards. The left card is titled 'Driver Projects' and contains text about real-world genomic data initiatives. The right card is titled 'ELIXIR Beacon' and provides links to its implementation studies and champions.

**Driver Projects**  
GA4GH Driver Projects are real-world genomic data initiatives that help guide our development efforts and pilot our tools. Stakeholders around the globe advocate, mandate, implement, and use our frameworks and standards in their local contexts.

**ELIXIR Beacon**  
<https://www.elixir-europe.org/about/implementation-studies/beacons>

Europe  
**Champions:** Jordi Rambla, Juha Tornroos, Gary Saunders

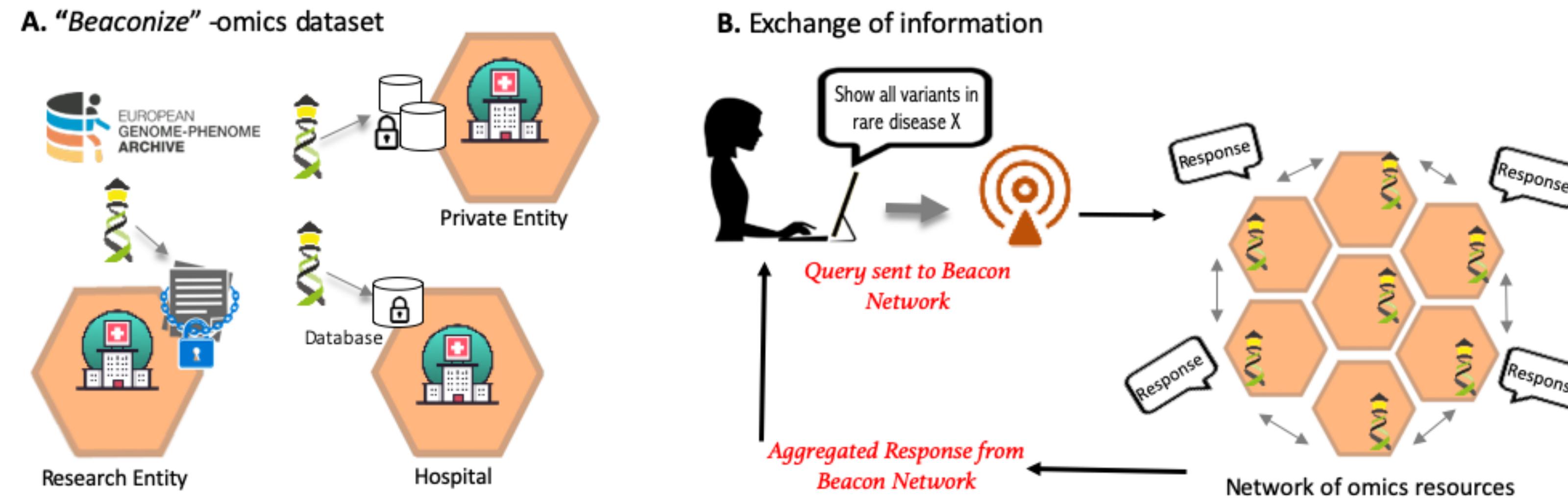
## v1.1 and roadmap

- structural variations** (DUP, DEL) in addition to SNV
- ... more structural queries (translocations/fusions...)
- Beacon queries as entry for **data handover** (outside Beacon protocol)
- layered authentication system using **ELIXIR AAI**
- v2** **filters** for phenotypic & technical metadata
- v2** Extended quantitative responses
  - Ubiquitous **deployment** (e.g. throughout ELIXIR network)



# Beacon v2

Enabling "omics" data access and retrieval from a variety of resources



A schematic representation of how Beacon works. (A) Beacon API implementation and (B) A Beacon query and aggregated response

# Beacon v2 - Clinical Beacon requirements

Authors: Jordi Rambla, Michael Baudis, Anthony J Brookes, Lauren Fromont, Claudia Vasallo, Aina Jené

The original GA4GH Beacon implementation (up to v0.3) was conceived as a protocol for sharing the presence/absence of a given, specific, genomic mutation in a set of data (from patients of a given disease or from the population in general). Although with some potential benefit, e.g. in the area of rare disease diagnostics, it was *not* designed for clinical use but chiefly to foster data sharing by triggering the inquisitiveness of researchers once some data of interest is discovered in another institution. While later extensions of the protocol (v1.0 - v1.n) expanded the query and response options, this did not deviate from the general "existence of variants in resource X" paradigm.

The simplicity and success of the concept has generated the request of making it more powerful, more useful in healthcare environments. The requests include:

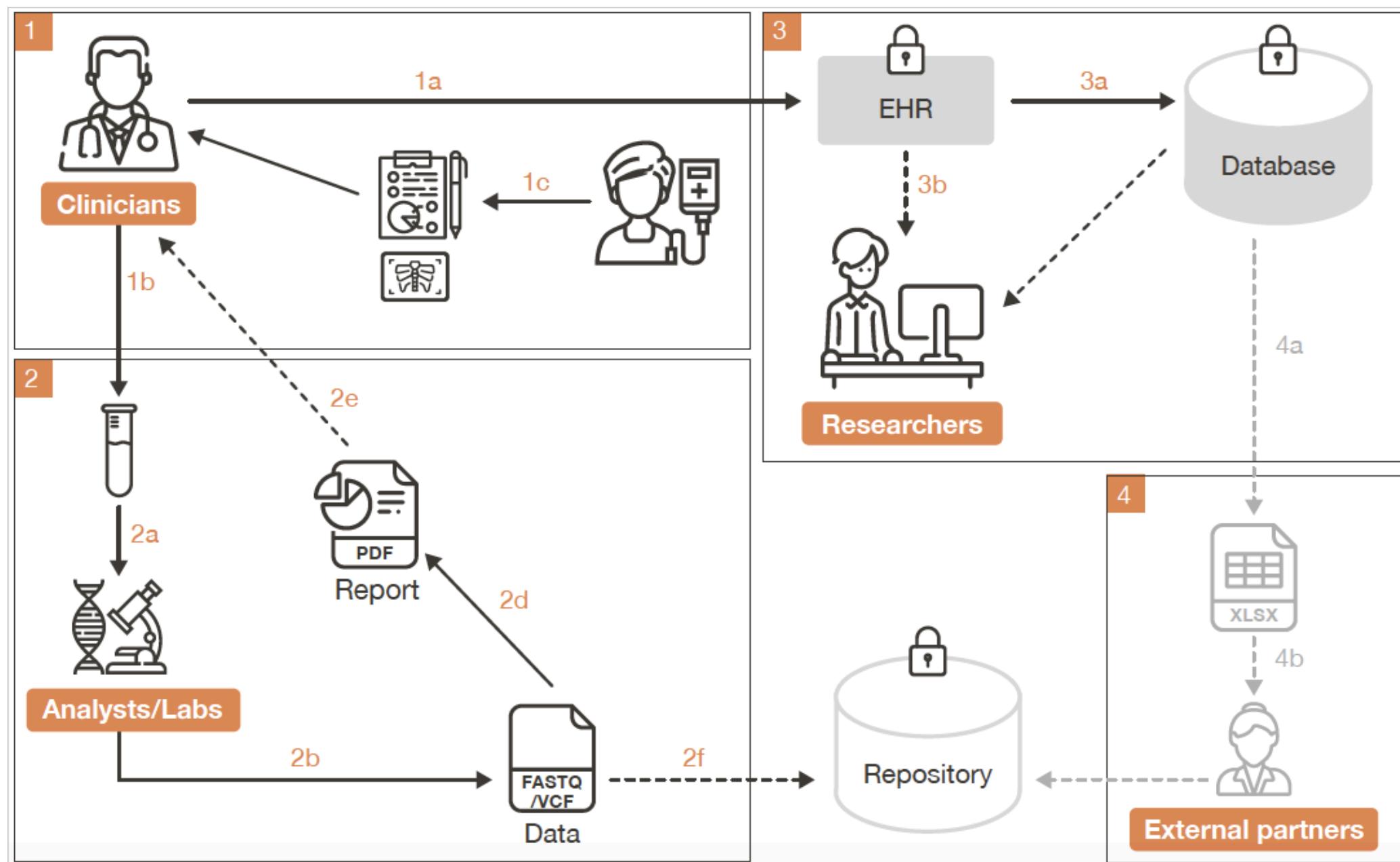
- Allowing more informative queries, like filtering by gender or age
- Allowing to trigger the next step in the data access process, e.g. who to contact or which are the data use conditions
- Jumping to another system where the data could be accessed, e.g. if the Beacon is internal to a hospital, to provide the Id of the EHR of the patients having the mutation of interest.
- Including annotations about the variants found, among which the expert/clinician conclusion about the pathogenicity of a given mutation in a given individual or its role in producing a given phenotype.

## The process

The GA4GH Beacon group started a set of meetings and interviews with GA4GH Driver Projects and with ELIXIR partners in order to determine the scope of the next generation Beacon. The goal was to be useful without breaking the simplicity that made Beacon version 1 successful.

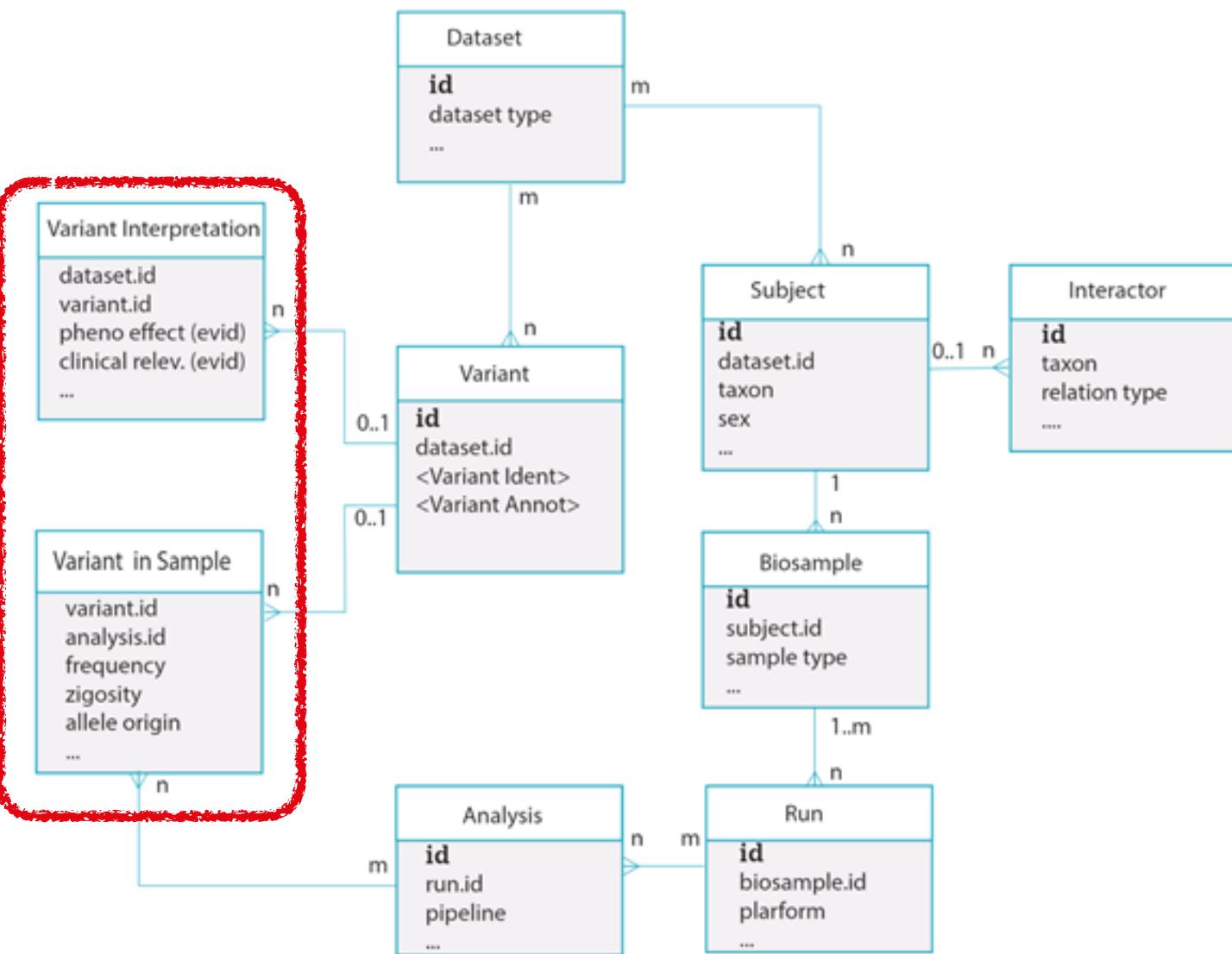
Interviews were conducted with the following GA4GH Driver Projects:

- Autism Speaks
- BRCA Exchange
- CanDIG
- EGA, ENA, EVA
- EuCanCancer
- European Joint Programme - Rare Diseases
- H3Africa
- GEM Japan
- Genomics England
- Matchmaker Exchange
- SVIP /SPHN
- VICC



Data flow and data sharing of genomic and phenotypic data in healthcare

The Beacon v2 draft 2 entities logical model (implementations may vary)



# Beacon v2 Code

## Contribute & Implement

- Github repositories for
  - ▶ beacon v2 framework
  - ▶ beacon v2 models
  - ▶ tools, website, templates ...
- framework with schema definitions for requests and responses, endpoints template examples
- models for request & response implementations, but in principle up to implementers to modify

## beacon-framework-v2

Beacon Framework version 2

### Introduction

The GA4GH Beacon specification is composed by two parts:

- the Beacon Framework (in *this* repo)
- the Beacon Model (in the [Models repo](#))

The **Beacon Framework** is the part that describes the overall structure of the API requests, responses, parameters, teh common components, etc. It could also be referred in this document as simply the *Framework*.

A **Beacon Model** describes the set of concepts included in a Beacon version (e.g. Beacon v2), like *individual* or *biosample*. It could also be referred in this document as simply the *Model*.

The Framework could be considered the *syntax* and the Model as the *semantics*.

Refer to the [Models repo](#) for further information about the Model and how to use it.

The Framework doesn't include anything related to specific entities but only the mechanisms for querying them and parsing the responses. The BF is, therefore, independent from/agnostic to any specific Model. It can be leveraged to describe models from other domains like proteomics, imaging, biobanking, etc.

A **Beacon instance** is just an implementation of a Beacon Model that follows the rules stated by the Beacon Framework.

If you are a Beacon implementer, then, you don't need to clone this (Framework) repo, you only need to **copy (or clone)** the Beacon Model and modify it to your specific instance. You will find plenty of references to the Framework in the Model copy, and you will use the Json schemas in the Framework to validate that both the structure of your requests and responses are compliant with the Beacon Framework. The [Beacon verifier](#) tool would help in such validation.

The Framework repo includes the elements that are common to all Beacons:

1. The configuration files
2. The Json schemas for the requests, the responses, and its respective sections
3. The files of every Beacon root
4. Examples of all the above (using a fake and simple Model)

	jrambla	Jordi Rambla
	mbaudis	Michael Baudis
	Tom-Shorter	Tom Shorter

	MrRobb	Mr.Robb
---	--------	---------

# Beacon v2 Paths

## Progenetix utilizes Beacon v2 REST paths

- Beacon v2 paths are used in the Beacon specification to scope query and delivery
- Progenetix uses a default `/biosamples/` + query path for its front end queries, and then collection specific methods for data retrieval (see next)
- current implementation addresses a core subset of all options, and evaluates some still moving targets

→ variants\_interpretations

→ variant instances versus prototypes

→ ...



### Base `/biosamples`

#### `/biosamples/` + query

- `/biosamples/?filters=cellosaurus:CVCL_0004`
  - this example retrieves all biosamples having an annotation for the Cellosaurus CVCL\_0004 identifier (K562)

#### `/biosamples/{id}/`

- `/biosamples/pgxbs-kftva5c9/`
  - retrieval of a single biosample

#### `/biosamples/{id}/variants/` & `/biosamples/{id}/variants_in_sample/`

- `/biosamples/pgxbs-kftva5c9/variants/`
- `/biosamples/pgxbs-kftva5c9/variants_in_sample/`
  - retrieval of all variants from a single biosample
  - currently - and especially since for a mostly CNV containing resource - `variants` means "variant instances" (or as in the early v2 draft `variantsInSample`)

### Base `/variants`

There is currently (April 2021) still some discussion about the implementation and naming of the different types of genomic variant endpoints. Since the Progenetix collections follow a "variant observations" principle all variant requests are directed against the local `variants` collection.

If using `g_variants` or `variants_in_sample`, those will be treated as aliases.

#### `/variants/` + query

- `/variants/?assemblyId=GRCh38&referenceName=17&variantType=DEL&filterLogic=AND&start=7500000&start=7676592&end=7669607&end=7800000`
  - This is an example for a Beacon "Bracket Query" which will return focal deletions in the TP53 locus (by position).

#### `/variants/{id}/` or `/variants_in_sample/{id}` or `/g_variants/{id}/`

- `/variants/5f5a35586b8c1d6d377b77f6/`
- `/variants_in_sample/5f5a35586b8c1d6d377b77f6/`

#### `/variants/{id}/biosamples/` & `variants_in_sample/{id}/biosamples/`

- `/variants/5f5a35586b8c1d6d377b77f6/biosamples/`
- `/variants_in_sample/5f5a35586b8c1d6d377b77f6/biosamples/`

# Beacon v2 Requests

## POSTing Queries

- Beacon v2 supports a mix of dedicated endpoints with REST paths
- POST requests using JSON query documents
- final syntax for core parameters still in testing stages

```
{  
  "$schema": "beaconRequestBody.json",  
  "meta": {  
    "apiVersion": "2.0",  
    "requestedSchemas": [  
      {  
        "entityType": "individual",  
        "schema": "https://progenetix.org/services/schemas/Phenopacket/"  
      }  
    ],  
    "query": {  
      "requestParameters": {  
        "datasets": {  
          "datasetIds": ["progenetix"]  
        }  
      },  
      "filterLogic": "OR"  
    },  
    "pagination": {  
      "skip": 0,  
      "limit": 10  
    },  
    "filters": [  
      { "id": "NCIT:C4536" },  
      { "id": "NCIT:C95597" },  
      { "id": "NCIT:C7712" }  
    ]  
  }  
}
```



# Beacon v2 Filters

# **Example: Use of hierarchical classification systems (here NCI I neoplasm core)**

- Beacon v2 "filters" assumes inclusion of child terms when using hierarchical classifications



- implementation of hierarchical annotations  
overcomes some limitations of "fuzzy"  
disease annotations

# progenet x

## Beacon+ specific: Multiple term selection with OR logic

<input checked="" type="checkbox"/>	> <a href="#">NCIT:C4914: Skin Carcinoma</a>	213
<input type="checkbox"/>	> <a href="#">NCIT:C4475: Dermal Neoplasm</a>	109
<input checked="" type="checkbox"/>	> <a href="#">NCIT:C45240: Cutaneous Hematopoietic and Lymphoid Cell Neoplasm</a>	310

**Filters:** NCIT:C4914, NCIT:C4819, NCIT:C9231, NCIT:C2921, NCIT:C45240, NCIT:C6858, NCIT:C3467, NCIT:C45340, NCIT:C7195, NCIT:C3246, NCIT:C7217



progenetix

Variants: 0    *f*alleles: 0    Callsets Variants ↗    UCSC region ↗    Calls: 0    Legacy Interface ↗    Samples: 523    [Show JSON Response](#)

Results    **Biosamples**

Id	Description	Classifications	Identifiers	DEL	DUP	CNV
PGX_AM_BS_MCC01	Merkel cell carcinoma	<a href="#">icdot-C44.9</a> Skin, NOS <a href="#">icdom-82473</a> Merkel cell carcinoma <a href="#">NCIT:C9231</a> Merkel Cell Carcinoma	<a href="#">PMID:9537255</a>	0.116	0.104	0.22
PGX_AM_BS_MCC02	Merkel cell carcinoma	<a href="#">icdot-C44.9</a> Skin, NOS <a href="#">icdom-82473</a> Merkel cell carcinoma <a href="#">NCIT:C9231</a> Merkel Cell Carcinoma	<a href="#">PMID:9537255</a>	0.154	0.056	0.21
PGX_AM_BS_MCC03	Merkel cell carcinoma	<a href="#">icdot-C44.9</a> Skin, NOS <a href="#">icdom-82473</a> Merkel cell carcinoma <a href="#">NCIT:C9231</a> Merkel Cell Carcinoma	<a href="#">PMID:9537255</a>	0.137	0.21	0.347
PGX_AM_BS_MCC04	Merkel cell carcinoma	<a href="#">icdot-C44.9</a> Skin, NOS <a href="#">icdom-82473</a> Merkel cell carcinoma <a href="#">NCIT:C9231</a> Merkel Cell Carcinoma	<a href="#">PMID:9537255</a>	0.158	0.056	0.214
PGX_AM_BS_MCC05	Merkel cell carcinoma	<a href="#">icdot-C44.9</a> Skin, NOS <a href="#">icdom-82473</a> Merkel cell carcinoma <a href="#">NCIT:C9231</a> Merkel Cell Carcinoma	<a href="#">PMID:9537255</a>	0.107	0.327	0.434



# Beacon Scouts: Structural Variants

## Re-defining & scoping variant queries

- contributors from different "stakeholder" areas
  - ▶ clinical genomics / rare diseases
  - ▶ variant repository (Ensembl)
  - ▶ cancer research resource
  - ▶ cancer variant annotation repositories
- close integration with ELIXIR h-CNV group
- process involved discussions about semantics of variant types, e.g.
  - ▶ DUP as CNV or in place
  - ▶ DEL as CNV from which size
- general attempt to use Sequence Ontology classes as guidance, but no still ambiguities / lack of terms

## Beacon Scouts: Structural Variants Use Cases & Examples

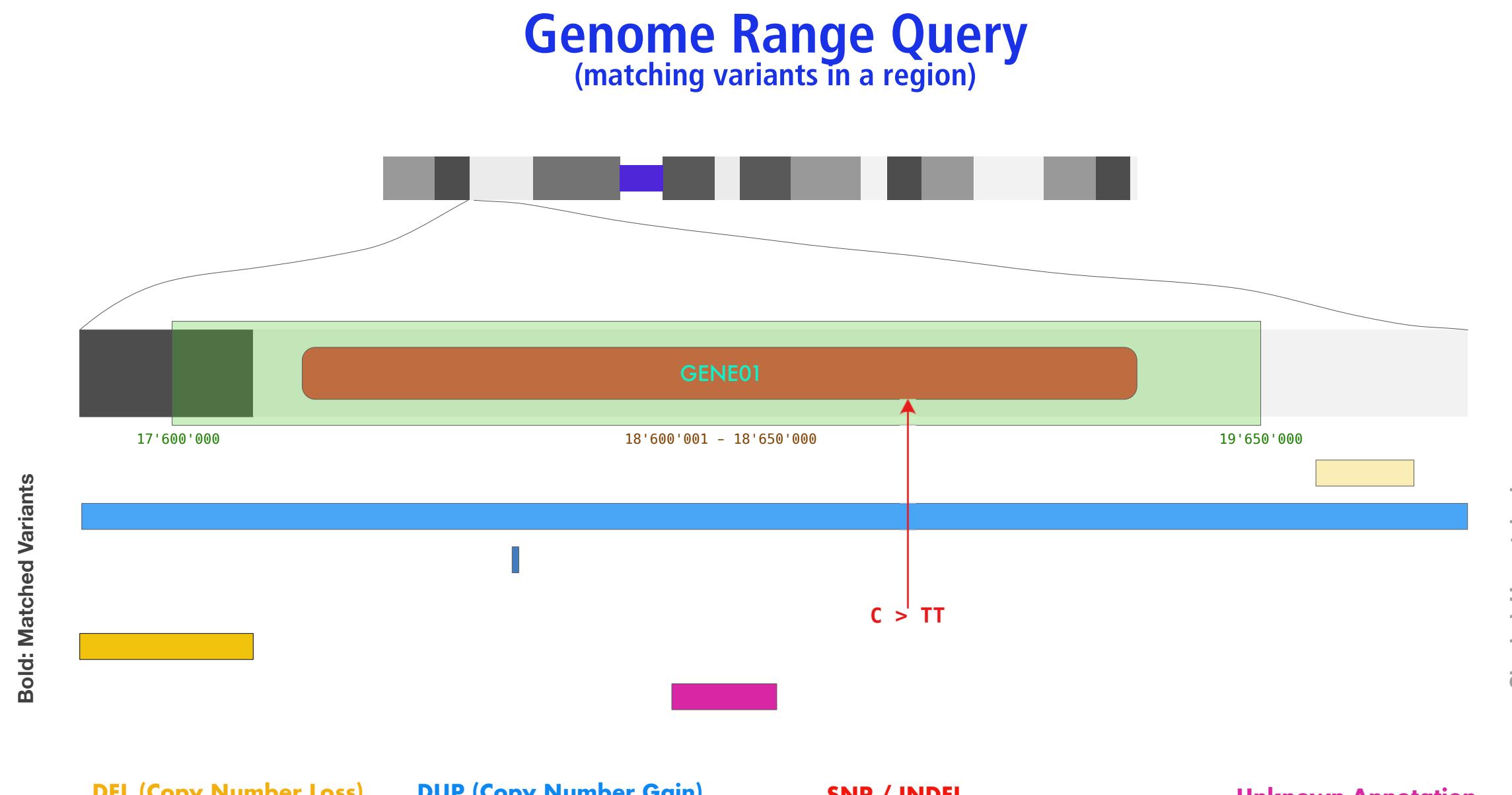
This document develops a set of structural variant types and associated query formats which will be supported by the Beacon protocol. The focus of the initial development is on the possibly limited, but unambiguous definition of query formats, driven and documented through real-world use cases.

<b>References</b>	<b>1</b>
<b>Conventions Followed in the Document</b>	<b>1</b>
<b>Use of Positional Parameters</b>	<b>2</b>
<b>Variant Types, Documentation and Example Queries</b>	<b>2</b>
INS (Insertion)	2
DEL (Deletion)	3
DUP (Duplication)	5
Amp (DUP more than 2) CN type of approach	8
LOH (loss of representation of second allele, with or without copy number change)	8
INV (inversion)	8
TL (Translocation)	9
Proposal: BRK (Breakpoint)	9
ME (Mobile elements insertion /deletion)	10
CNV - (non directional CNVs) - do we allow cnv queries? / complex CNVs	10
Tandem Duplication	11
<b>Name Based Queries</b>	<b>11</b>
Topics for discussion	11
<b>Technical considerations</b>	<b>12</b>
Format of GET queries	12

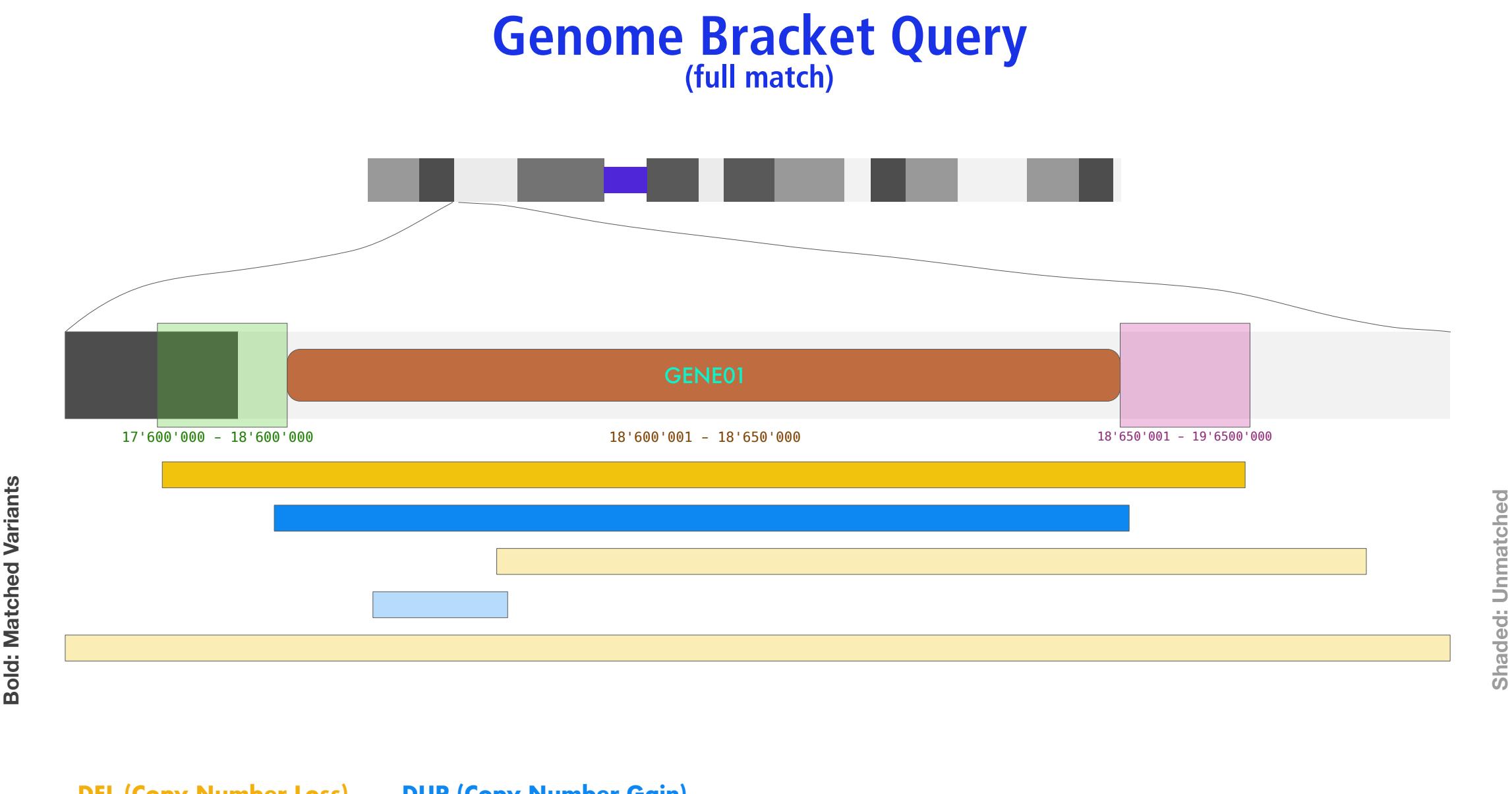


# Beacon v2: Extended Variant Queries

Range and Bracket queries enable positional wildcards and fuzziness



- Genome Range Queries provide a way to "fish" for variants overlapping an indicated region, e.g. the CDR of a gene of interest
- Additional parameters (e.g. variant type, reference or alternate bases) limit the scope of the responses
- new Beacon v2 size parameters to limit structural variants (e.g. "focal" CNVs)



- Genome Bracket Queries allow to search for structural variants with start and end positions falling into defined sequence ranges
- allows to query any contiguous genomic variant (and in principle also can step in for range queries)
- typical use case is e.g. the query for variants such as duplications covering the whole CDR of a gene, while limiting the allowed start or end regions

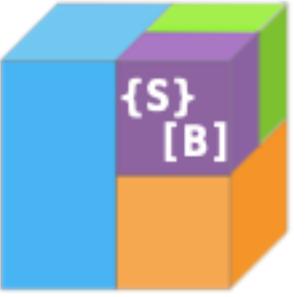
# Beacon & Phenopackets

**Data discovery and delivery using standardized GA4GH formats and schemas**

- Beacon **v2** & Phenopackets **v2** emerging as essential standards for federation and exchange of biomedical data
- hierarchical coding systems and with widespread use of CURIEs
- other formats based on international standards, e.g.
  - ISO (ISO 8601 time & period, ISO 3166 country codes, ...)
  - IETF (GeoJSON ...)
  - W3C (CURIE ...)
- these standards become pervasive throughout GA4GH's ecosystem

➡ Beacon query **filters** correspond well to Phenopackets data

➡ Phenopackets as recommended protocol for Beacon **v2 data delivery**

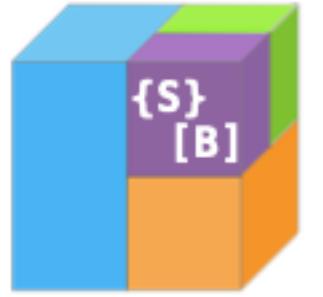


```
"data_use_conditions" : {
    "label" : "no restriction",
    "id" : "DUO:000004"
},
"material" : {
    "id" : "EFO:0009656",
    "label" : "neoplastic sample"
},
{
    "ageAtDiagnosis": "P25Y3M2D"
},
"sampled_tissue" : {
    "id" : "UBERON:0002037",
    "label" : "cerebellum"
},
"histological_diagnosis" : {
    "id" : "NCIT:C3222",
    "label" : "Medulloblastoma"
},
"pathological_tnm_findings" : [
    {
        "id" : "NCIT:C48700",
        "label" : "M1 Stage Finding"
    },
    {
        "id" : "NCIT:C48720",
        "label" : "T1 Stage Finding"
    }
]
```

# GA4GH {S}[B] SchemaBlocks

Standardized formats and data schemas for developing an "Internet of Genomics"

- “cross-workstreams, cross-drivers” initiative to document GA4GH object **standards** and **prototypes** launched in December 2018
- documentation and implementation examples provided by GA4GH members
- not a rigid, complete data schema
- object **vocabulary** and **semantics** for a large range of developments
- ▶ **Beacon** as contributor and user
- ▶ 2021: going forward through integration with GA4GH TASC efforts, towards "standards library"



Biosample sb-phenopackets ↗	
{S}[B] Status [i]	implemented
Provenance	◦ Phenopackets
Used by	◦ Phenopackets
Contributors	◦ GA4GH Data Working Group ◦ Jules Jacobsen ◦ Peter Robinson ◦ Michael Baudis ◦ Melanie Courtot ◦ Isuru Liyanage
Source (v1.0.0)	◦ raw source [JSON] ◦ Github

HtsFile sb-phenopackets ↗	
{S}[B] Status [i]	implemented
Provenance	◦ Phenopackets
Used by	◦ Phenopackets
Contributors	◦ Jules Jacobsen ◦ Peter Robinson
Source (v1.0.0)	◦ raw source [JSON] ◦ Github

Attributes	
Type:	object
Description:	A Biosample refers to a unit of biological material from which the genomic DNA, RNA, proteins) for molecular analyses (e.g. sequencing, array hybridization, spectrometry) are extracted.
Examples:	would be a tissue biopsy, a single cell from a culture or single cell gel fraction from a gradient centrifugation.
Several instances (e.g. technical replicates) or types of experiments (e.g. genome-wide association studies) may refer to the same Biosample.	
FHIR mapping:	Specimen.

Properties	
Property	Type
ageOfIndividualAtCollection	<a href="https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/Age.json">https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/Age.json</a> [SRC] [HTML]
ageRangeOfIndividualAtCollection	<a href="https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/AgeRange.json">https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/AgeRange.json</a> [SRC] [HTML]
description	string
diagnosticMarkers	array of <a href="https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/OntologyClass.json">https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/OntologyClass.json</a> [SRC] [HTML]
histologicalDiagnosis	<a href="https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/OntologyClass.json">https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/OntologyClass.json</a> [SRC] [HTML]
htsFiles	array of <a href="https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/HtsFile.json">https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/HtsFile.json</a> [SRC] [HTML]
id	string
individualId	string
isControlSample	boolean
phenotypicFeature	array of <a href="https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/PhenotypicFeature.json">https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/PhenotypicFeature.json</a> [SRC] [HTML]
procedure	<a href="https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/Procedure.json">https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/Procedure.json</a> [SRC] [HTML]
sampledTissue	<a href="https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/Tissue.json">https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/Tissue.json</a> [SRC] [HTML]

[schemablocks.org](http://schemablocks.org)



# Beacon Project - Partner Engagement & Next Steps

- Working with **partner communities & projects** on *deploying Beacons*
  - ELIXIR hCNV Community
  - European Joint Program on Rare Diseases
  - clinical groups & data initiatives (e.g. Andalucia, Cancer Core Europe, SPHN)
  - variant annotation resources, with optional clinical components (e.g. SVIP-O)
- Improving reference implementation and standards / **compliance** testing
- Beacon **v2** "fast forward" development
- aligning w/ GA4GH standards, through "request & adopt" => SchemaBlocks **{S}[B]**
- networks **throughout & beyond ELIXIR**



# ELIXIR Beacon Network



- developed under lead from ELIXIR Finland
- **authenticated access** w/ ELIXIR AAI
- **incremental extension**, starting with ELIXIR Beacon resources adhering to the **latest specification** (contrast to legacy networks)
- service details provided by individual Beacons, using **GA4GH service-info**
- **registration service**
- **integrator** throughout ELIXIR Human Data
- starting point for "**beyond ELIXIR**" **feature rich** federated Beacon **services**

GRCh38 17 : 7577121 G > A Search

Example variant query Advanced Search

**baudisgroup at UZH and SIB**  
Progenetix Cancer Genomics Beacon+

Beacon+ provides a forward looking implementation of the Beacon API, with focus on structural variants and metadata based on the cancer and reference genome profiling data represented in the Progenetix oncogenomic data resource (<https://progenetix.org>).

[Visit Us](#) · [Beacon API](#) · [Contact Us](#)

**CSC - IT Center for Science**  
Development Beacon

Beacon API Web Server based on the GA4GH Beacon API

[Visit Us](#) · [Beacon API](#) · [Contact Us](#)

**Research Programme on Biomedical Informatics**  
DisGeNET Beacon

Variant-Disease associations collected from curated resources and the literature

[Visit Us](#) · [Beacon API](#) · [Contact Us](#)

**University of Tartu Institute of Genomics, Estonia**  
Beacon at the University of Tartu, Estonia

Beacon API Web Server based on the GA4GH Beacon API

[Visit Us](#) · [Beacon API](#) · [Contact Us](#)

**National Bioinformatics Infrastructure Sweden**  
SweFreq Beacon

Beacon API Web Server based on the GA4GH Beacon API

[Visit Us](#) · [Beacon API](#) · [Contact Us](#)

**LCSB at University of Luxembourg**  
ELIXIR.LU Beacon

ELIXIR.LU Beacon

[Visit Us](#) · [Beacon API](#) · [Contact Us](#)

**European Genome-Phenome Archive (EGA)**  
EGA Beacon

This Beacon is based on the GA4GH Beacon API v1.1.0

[Visit Us](#) · [Beacon API](#) · [Contact Us](#)

**CSC - IT Center for Science**  
Production Beacon

Beacon API Web Server based on the GA4GH Beacon API

[Visit Us](#) · [Beacon API](#) · [Contact Us](#)

# hCNV Implementation Studies 2021-2023 No. 2



## Beacon and beyond – Implementation-driven standards and protocols for CNV discovery and data exchange

- reinforce work on priority areas established in the current hCNV Implementation Study
- extend collaborations with the Rare Diseases and Galaxy Communities, EJP-RD and GA4GH
- Expected outcomes
  - ➔ shared CNV resources testing advanced versions of the Beacon protocol
  - ➔ integration of GA4GH standards such as Phenopackets in such resources
  - ➔ tools for data ingestion and export for standard formats (e.g. VCF, Phenopackets) and CNV-specific improvements of such standards
- connecting to international partners, e.g. Cancer Genomics Consortium (U.S.)

- ▶ WP1 - hCNV community reference resources
- ▶ WP2 - hCNV Resources and Beacon
- ▶ WP3 - Galaxy Community Intersection and Data Exchange
- ▶ WP4 - Workflows and Tools for hCNV Data Exchange Procedures
- ▶ WP5 - Training and dissemination



# hCNV Implementation Studies 2021-2023 No. 2



Beacon and beyond – Implementation-driven standards and protocols for CNV discovery and data exchange

The screenshot shows a web interface for the Progenetix database. At the top, it says "Cancer genome data @ progenetix.org". Below that, a text block states: "The Progenetix database provides an overview of mutation data in cancer, with a focus on copy number abnormalities (CNV / CNA), for all types of human malignancies. The data is based on *individual sample data* from currently 139448 samples." A chart titled "Sezary syndrome (icdom-97013)" displays CNV frequencies across 22 chromosomes for 166 samples. The chart has a y-axis ranging from -75 to 75. Below the chart, there are links to "Download SVG", "Go to icdom-97013", and "Download CNV Frequencies". A note below the chart says: "Example for aggregated CNV data in 166 samples in Sezary syndrome. Here the frequency of regional **copy number gains** and **losses** are displayed for all 22 autosomes." At the bottom, there is a dark header with "BANCCO" and navigation links for "Accueil", "Statistique", "Contact", and "Inscription". A login form with fields for "Email" and "Mot de passe" is shown. Below the header, a call-to-action button says "Importer vos données et partager les avec les partenaires du réseau" and "BANCCO - Banque Nationale de CNV Constitutionnelles".

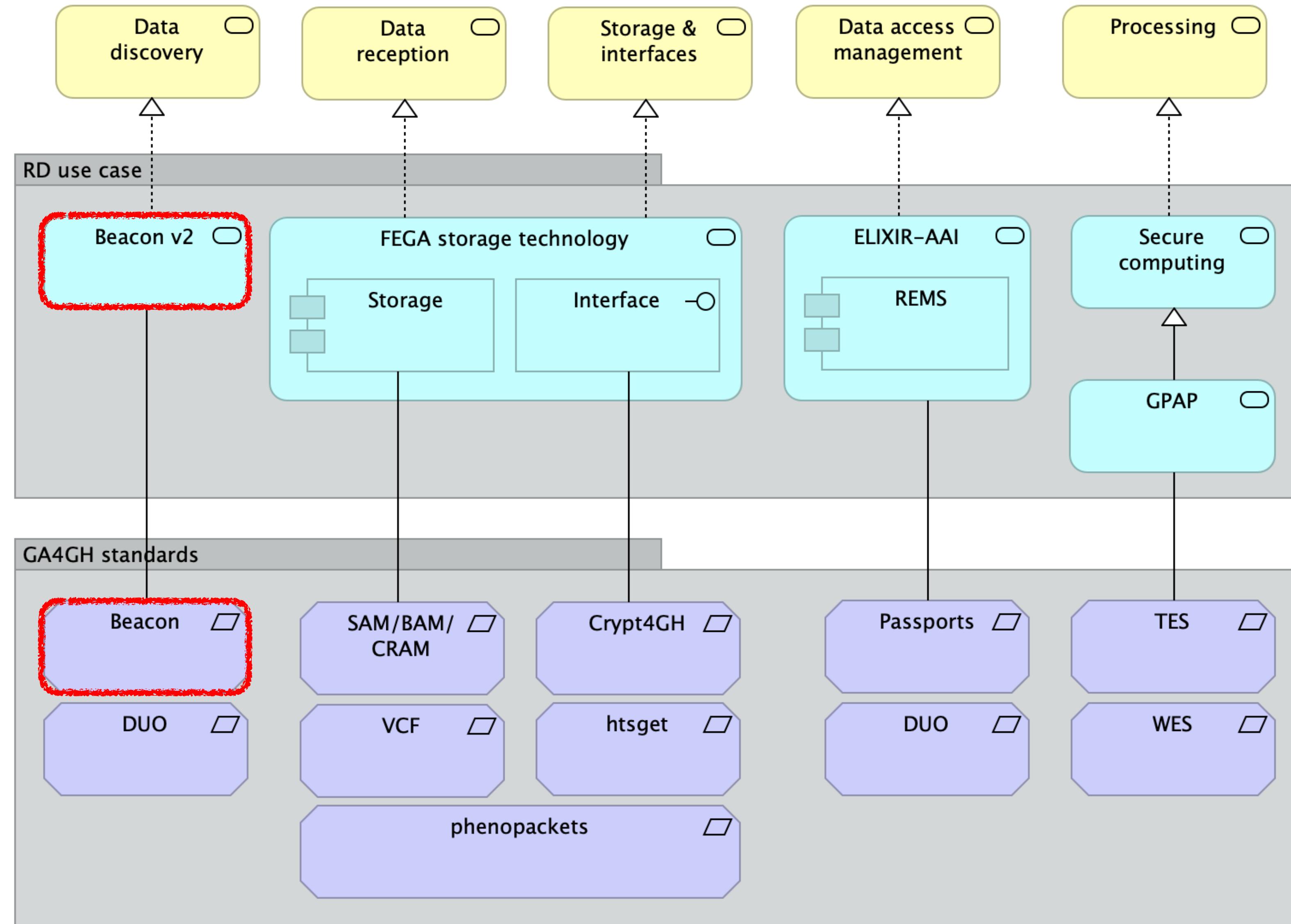
"Galaxify", "Beaconize" & "Phenopack"  
Progenetix & RD-CNVdb prototypes



- ▶ WP1 - hCNV community reference resources
- ▶ WP2 - hCNV Resources and Beacon
- ▶ WP3 - Galaxy Community Intersection and Data Exchange
- ▶ WP4 - Workflows and Tools for hCNV Data Exchange Procedures
- ▶ WP5 - Training and dissemination



# Functionalities, Standards, Components





# Onboarding

## Demonstrating Compliance

- onboarding server run by CRG (EGA team)
- registering the URI of a server's map document will initiate traversal and testing of services
- blueprint for Beacon service registries
- to be used as demonstrator in GA4GH approval process for the Spring 2022 session



The figure displays four screenshots of Beacon service registries, each showing a list of supported service endpoints and their status relative to the GA4GH specification.

**CNAG (Centre Nacional Analisis Genomica):**

- Configuration: ✗ Configuration
- analysis: ✗ analysis
- dataset: ✗ dataset
- interactor: ✗ interactor
- variantInSample: ✗ variantInSample
- BeaconMap: ✗ BeaconMap
- biosample: ✗ biosample
- genomicVariant: ✗ genomicVariant
- run: ✗ run
- variantInterpretation: ✗ variantInterpretation
- EntryTypes: ✓ EntryTypes
- cohort: ✗ cohort
- individual: ✗ individual
- variantAnnotation: ✗ variantAnnotation

**European Genome-Phenome Archive (EGA):**

- Configuration: ✓ Configuration
- analysis: ✓ analysis
- dataset: ✓ dataset
- interactor: ✓ interactor
- variantInSample: ✓ variantInSample
- BeaconMap: ✓ BeaconMap
- biosample: ✓ biosample
- genomicVariant: ✓ genomicVariant
- run: ✓ run
- variantInterpretation: ✓ variantInterpretation
- EntryTypes: ✓ EntryTypes
- cohort: ✓ cohort
- individual: ✓ individual
- variantAnnotation: ✓ variantAnnotation

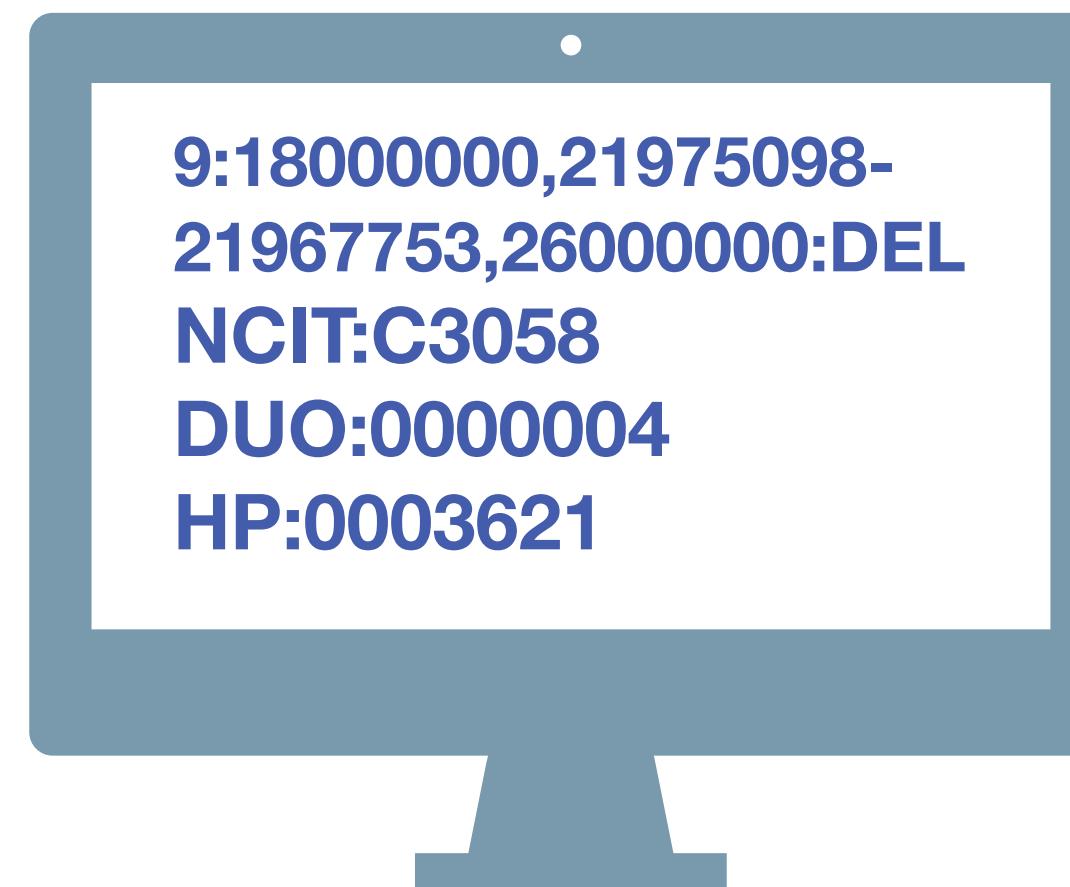
**European Genome-Phenome Archive (EGA) (Test Instance):**

- Configuration: ✓ Configuration
- analysis: ✓ analysis
- dataset: ✓ dataset
- interactor: ✓ interactor
- variantInSample: ✓ variantInSample
- BeaconMap: ✓ BeaconMap
- biosample: ✓ biosample
- genomicVariant: ✓ genomicVariant
- run: ✓ run
- variantInterpretation: ✓ variantInterpretation
- EntryTypes: ✓ EntryTypes
- cohort: ✓ cohort
- individual: ✓ individual
- variantAnnotation: ✓ variantAnnotation

**Canadian Distributed Infrastructure for Genomics:**

- Configuration: ✗ Configuration
- analysis: ✗ analysis
- dataset: ✗ dataset
- interactor: ✗ interactor
- variantInSample: ✗ variantInSample
- BeaconMap: ✗ BeaconMap
- biosample: ✗ biosample
- genomicVariant: ✗ genomicVariant
- run: ✗ run
- variantInterpretation: ✗ variantInterpretation
- EntryTypes: ✗ EntryTypes
- cohort: ✗ cohort
- individual: ✗ individual
- variantAnnotation: ✗ variantAnnotation

Legend: ✓ Matches the Spec, ✗ Not Match the Spec, ? Not Implemented

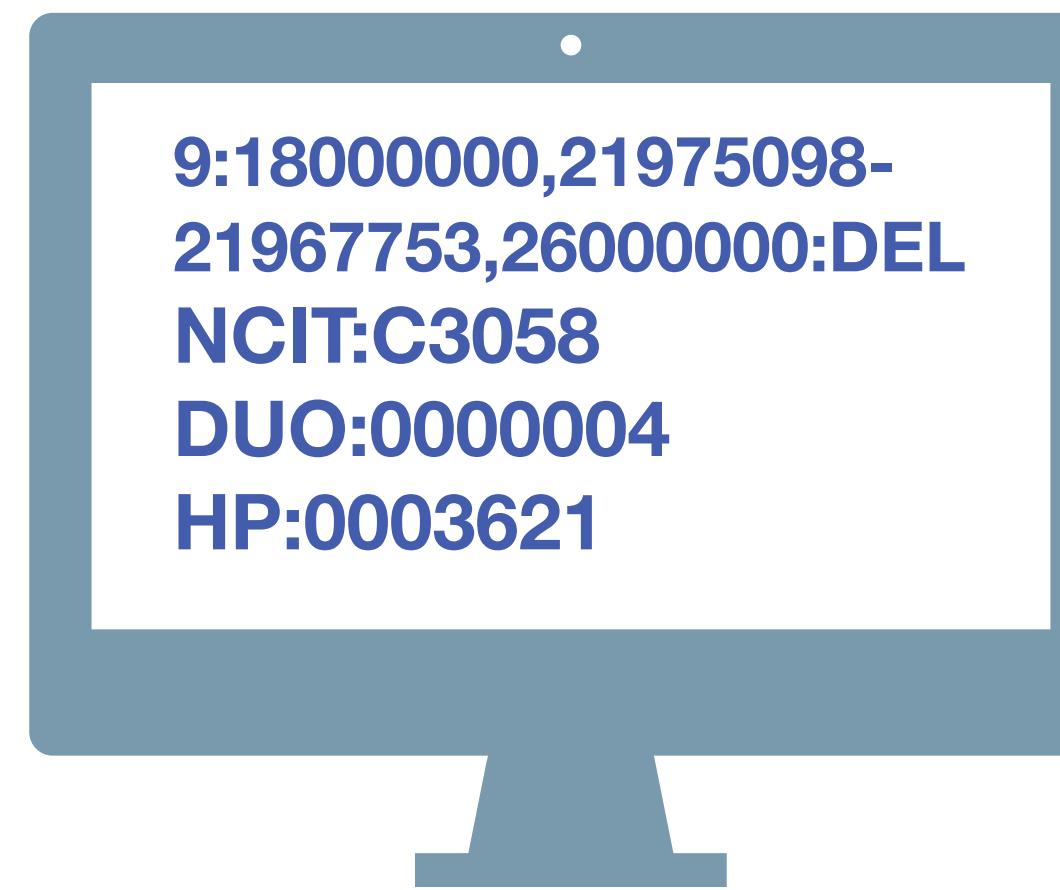


Have you seen deletions in this region on chromosome 9 in Glioblastomas from a juvenile patient, in a dataset with unrestricted access?

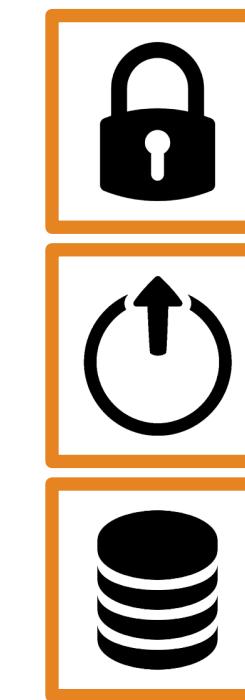


## Beacon v2 API

The Beacon API v2 proposal opens the way for the design of a simple but powerful "genomics API".



Have you seen deletions in this region on chromosome 9 in Glioblastomas from a juvenile patient, in a dataset with unrestricted access?



## Beacon v2 API

The Beacon API v2 proposal opens the way for the design of a simple but powerful "genomics API".



## GA4GH Genome Beacons

A Driver Project of the Global Alliance for Genomics and Health GA4GH and supported through ELIXIR

[News](#)

[Specification & Roadmap](#)

[Beacon Networks](#)

[Events](#)

[Examples, Guides & FAQ](#)

[Contributors & Teams](#)

[Contacts](#)

[Meeting Minutes](#)

[Related Sites](#)

[ELIXIR BeaconNetwork](#)

[Beacon @ ELIXIR](#)

[GA4GH](#)

[beacon-network.org](#)

[Beacon+](#)

[GA4GH::SchemaBlocks](#)

[GA4GH::Discovery](#)

[Github Projects](#)

[Beacon API and Tools](#)

[SchemaBlocks](#)

[Tags](#)

[CNV](#) [EB](#) [FAQ](#) [SV](#) [VCF](#) [beacon](#) [clinical](#)

[code](#) [compliance](#) [contacts](#) [definitions](#)

[developers](#) [development](#) [events](#) [filters](#)

[minutes](#) [network](#) [press](#) [proposal](#)

[queries](#) [releases](#) [roadmap](#)

[specification](#) [teams](#) [v2](#) [versions](#)

[website](#)

## Beacon Protocol for Genomic Data Sharing

Beacons provide discovery services for genomic data using the Beacon API developed by the [Global Alliance for Genomics and Health \(GA4GH\)](#). The [Beacon protocol](#) itself standard for genomics data discovery. It provides a framework for public web service against genomic data collections, for instance from population based or disease specific repositories.



## Baudisgroup @ UZH

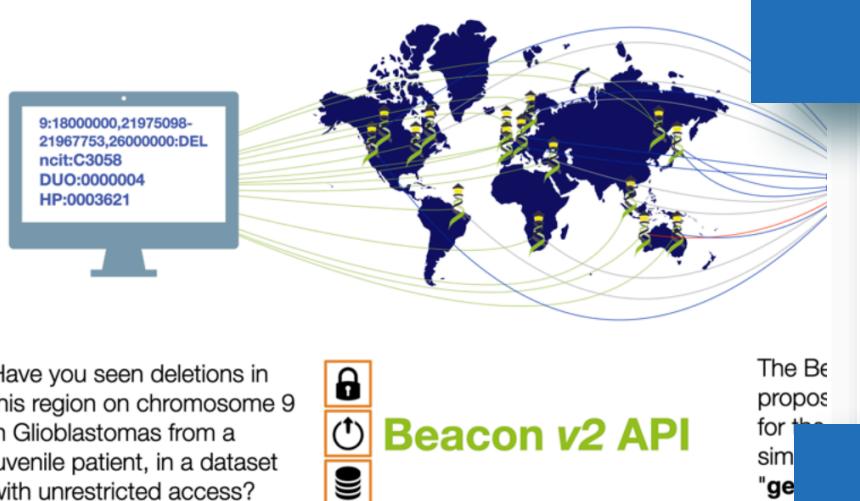
Michael Baudis  
(Paula Carrio Cordo)  
(Bo Gao)  
Qingyao Huang  
Sofia Pfund  
Rahel Paloots  
Hangjia Zhao

Pierre-Henri Toussaint

The original Beacon protocol had been designed to be:

- **Simple:** focus on robustness and easy implementation
  - **Federated:** maintained by individual organizations and assembled into networks
  - **General-purpose:** used to report on any variant collection
  - **Aggregative:** provide a boolean (or quantitative) answer about the observed variants
  - **Privacy protecting:** queries do not return information about single individuals
- Sites offering *beacons* can scale through aggregation *Beacon Networks*, which aggregate queries among a potentially large number of international *beacons* and assemble them into a single response. Since 2015 the development of the Beacon protocol has been led by [ELIXIR](#) in close collaboration with international participants. Recent versions of the *Beacon* protocol have expanded its scope:
- providing a framework for other types of genome variation data (i.e. rare variants)
  - allowing for data delivery using *handover* protocol, e.g. to link with clinical environments and allow for data delivery and visualisation services

### Beacon v2 - Towards Flexible Use and Clinical Applications



Have you seen deletions in this region on chromosome 9 in Glioblastomas from a juvenile patient, in a dataset with unrestricted access?

[beacon-project.io](http://beacon-project.io)

**{S}[B] and GA4GH**  
Melanie Courtot  
Helen Parkinson  
many more ...

[progenetix.org](http://progenetix.org)



## Beacon API Leads

Jordi Rambla  
Anthony Brooks  
Juha Törnroos

## Discovery WS

Michael Baudis (Beacon)  
Marc Fiume (Networks)

## ELIXIR

Gary Saunders  
David Lloyd  
Serena Scollen  
Dylan Spalding

## Beacon Team CRG

Laureen Fromont  
Babita Singh  
Sabela de la Torre Pernas

...

## Beacon v2 Scouts

Tim Beck  
Joaquin Dopazo  
Veronique Geoffroy  
Jean Muller  
David Salgado  
Alex Wagner

...

[github.com/ga4gh-beacon/](https://github.com/ga4gh-beacon/)

The screenshot shows the GitHub repository page for 'ga4gh-beacon'. At the top, there are buttons for 'Unwatch' (7), 'Star' (1), 'Fork' (2), and a bell icon. Below these are tabs for 'Actions', 'Wiki', 'Security', 'Insights', and a dropdown for 'About'. The 'About' section contains the following text: 'GA4GH Beacon v2 specification.', with tags 'ga4gh', 'beacon', and 'openapi'. It also includes sections for 'Readme', 'Apache-2.0 License', 'Releases' (no releases published), 'Create a new release', 'Packages' (no packages published), 'Publish your first package', and 'Contributors' (3 contributors: sdelatorrep, mbaudis, and blankdots). The main content area shows a table of contents for the specification, including 'Introduction', 'Protocol Overview', 'Data Model', 'Implementation', 'API', 'Security', 'Performance', 'Future Work', and 'Conclusion'. Each section has a brief description and a 'View file' link. The 'API' section is currently expanded, showing code snippets for 'Structural Variants' and 'Allele Frequency'. The 'Implementation' section links to 'Beacon API' and 'Beacon v2 API'.

