

# The ELIXIR Beacon Project

## Context, status & next steps

Michael Baudis -ELIXIR HDC HON - 2018-02-20



University of  
Zurich<sup>UZH</sup>



# GA4GH API promotes sharing

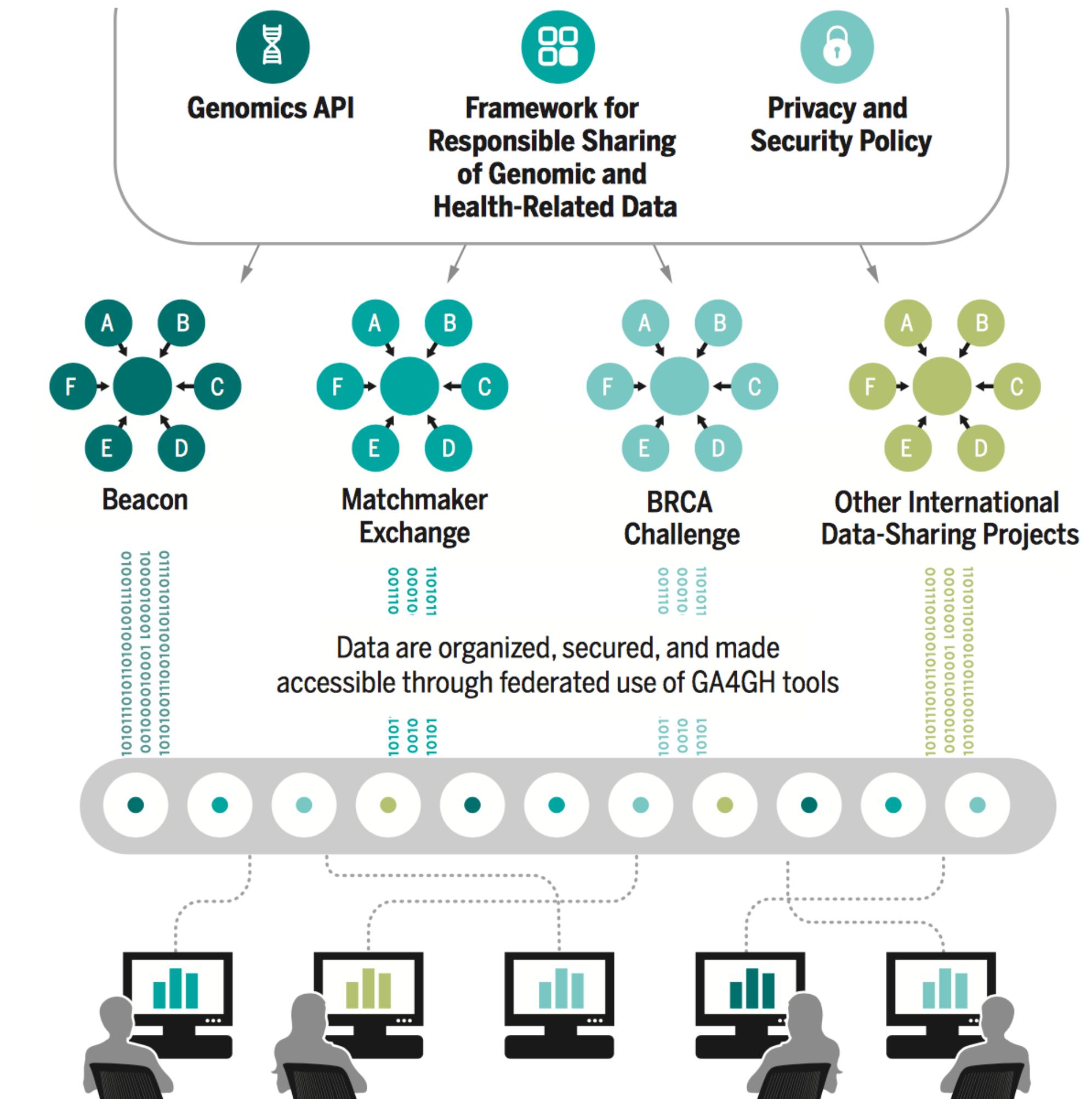
**A federated data ecosystem.** To share genomic data globally, this approach furthers medical research without requiring compatible data sets or compromising patient identity.



**GENOMICS**

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

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



# Beacon Project

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



Beacon Network

Search Beacons

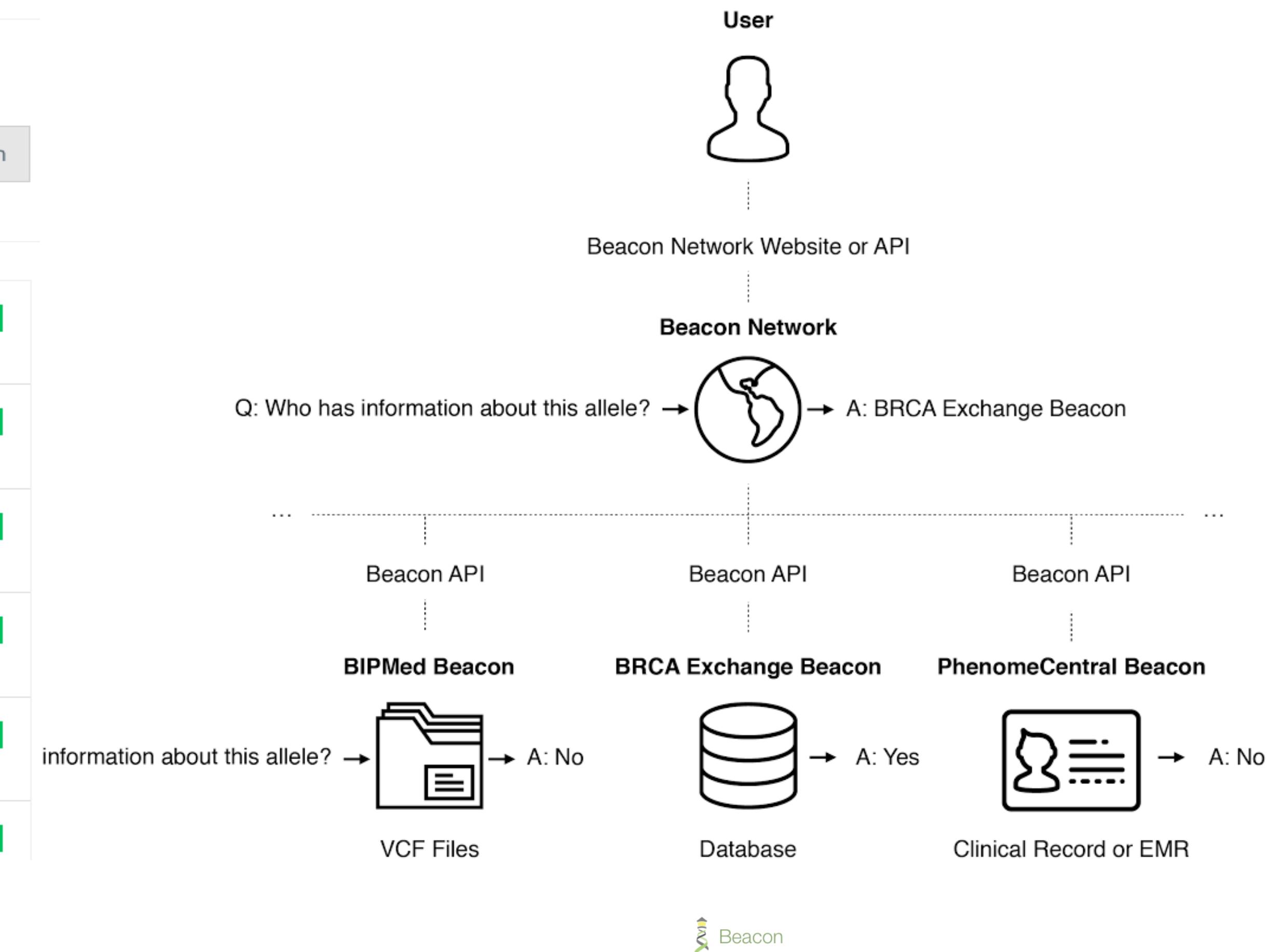
Search all beacons for allele

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 ...

		Found
	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

# ELIXIR Beacon Project



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

The slide shows the GA4GH Driver Projects section. It features a red circular icon with a white DNA helix and the text "Driver Projects". Below it, a blue bar contains the text "GA4GH Driver Projects are real-world genomic data initiatives that help our development efforts and pilot our tools. Stakeholders around the globe advocate, mandate, implement, and use our frameworks and standards in local contexts." To the right, there is a white box with the ELIXIR logo and the text "ELIXIR Beacon" and "www.elixir-europe.org". It also lists "Europe" as the region and "Champions: Serena Scollen, Ilkka Lappalainen, Michael Baudis".

## Beacon forward

- **structural variations** (DUP, DEL) in addition to SNV
- **metadata** queries
- layered **authentication** system using **ELIXIR** AAI
- **quantitative** responses
- Beacon queries as entry for **data delivery** (outside Beacon protocol)
- Ubiquitous **deployment** (e.g. throughout **ELIXIR** network)

# ELIXIR Beacon Project 2018



- WP1 - Development of Beacon API specification with the GA4GH
- WP2 - Supporting new ELIXIR Beacon development and deployment across Europe
- WP3 - Implementation of an ELIXIR Beacon Network & Registry
- WP4 - Security of ELIXIR Beacon and Beacon Network
- WP5 - Developing Indicators to establish Registry and Beacon as an Emerging ELIXIR Service
- WP6 - Strategic partnerships with national data cohorts and biobanks
- WP7 - Project Management and Communication

Node	Name of PI	Role	PMs	Other Contrib	Work Packages						
					1	2	3	4	5	6	7
FI	Ilkka Lappalainen	co-lead	5		X	X	X	X	X	X	X
ES	Jordi Rambla	lead	5		X	X	X				X
CH	Michael Baudis, (Heinz Stockinger)	co-lead	5		X						X X
EBI	Dylan Spalding	co-lead	5		X	X	X	X			
SE	Niclas Jareborg	member	1			X					
BE	Yves Moreau	member	1				X				
FR	Macha Nikolski	member	1					X			
NL	Morris Swertz	member	1						X		
IT	David Horner	member	-								
FI	ELIXIR Compute Platform - Tommi Nyrönen								2		
TBD <sup>6</sup>	ELIXIR Training Platform							1		X	
HUB	Serena Scollen, Susheel Varma	member coordinator	0.5		X				X	X	X

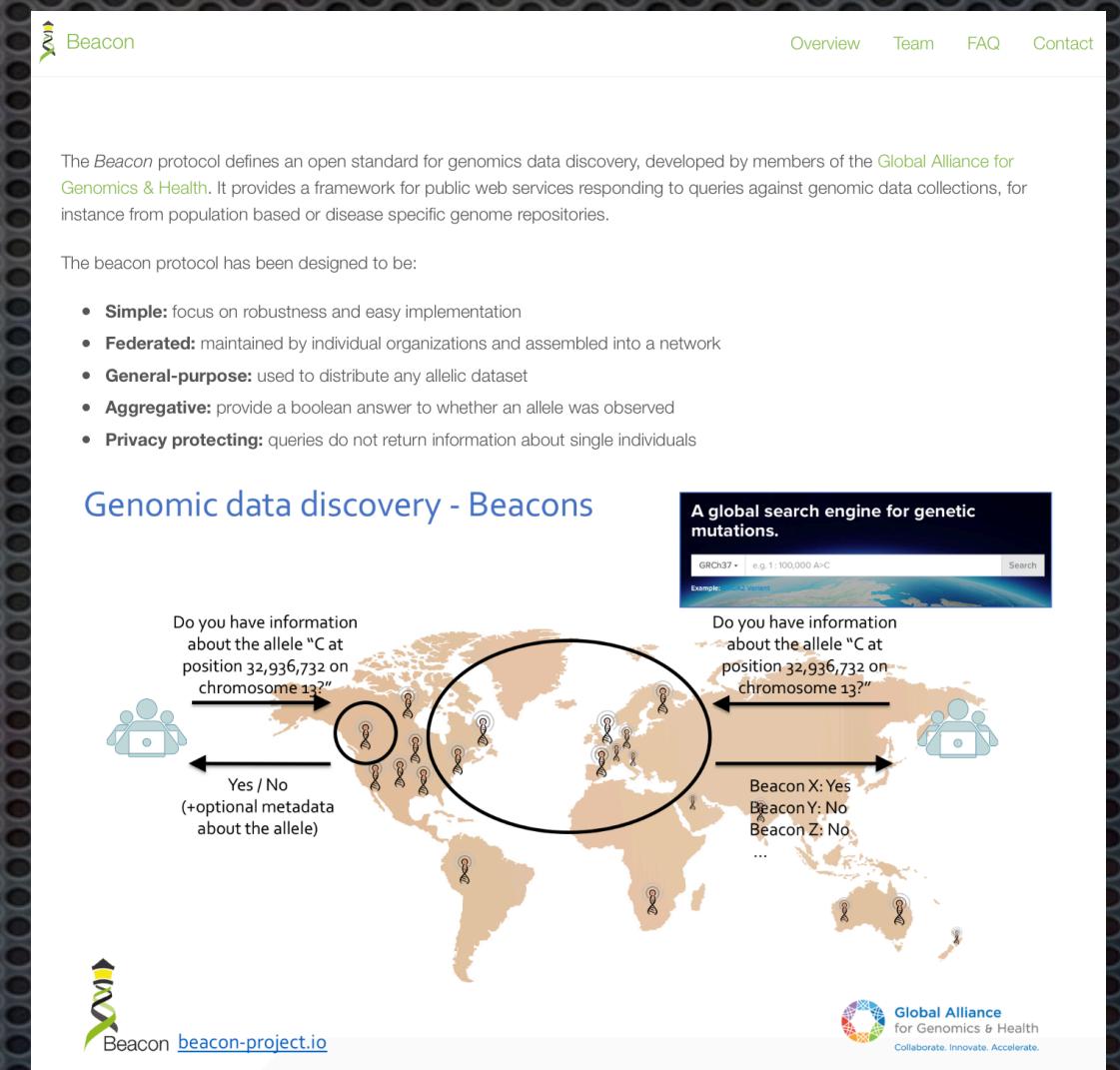
Node work package participation, as in 2018 project plan





# ELIXIR Beacon - Feb 2018

- Release of Beacon specification v0.4
- Github project re-organisation under way
- Launch of Beacon project website at [beacon-project.io](http://beacon-project.io)
- Beacon workshop planned
  - 2018-03-19/2018-03-20, Barcelona
  - strategy:
    - ▶ Beacon implementation wishlist
    - ▶ next steps
    - ▶ roadmap alignment
  - hands-on coding
- New technical coordinator: Gary Saunders



- WP1 - Development of Beacon API specification with the GA4GH
- WP2 - Supporting new ELIXIR Beacon development and deployment across Europe
- WP3 - Implementation of an ELIXIR Beacon Network & Registry
- WP4 - Security of ELIXIR Beacon and Beacon Network
- WP5 - Developing Indicators to establish Registry and Beacon as an Emerging ELIXIR Service
- WP6 - Strategic partnerships with national data cohorts and biobanks
- WP7 - Project Management and Communication





# ELIXIR Beacon - Challenges

- Engaging ELIXIR nodes w/ **human genome resources** regarding Beacon implementations
- Propagating protocol extensions (variants, metadata, authentication) throughout existing & upcoming Beacon instances
- ELIXIR Beacon network & registry
- Alignment of Beacon development and dissemination w/ HDC priority topics
  - **rare disease resources**
  - **structural genome variants**
- WP1 - Development of Beacon API specification with the GA4GH
- WP2 - Supporting new ELIXIR Beacon development and deployment across Europe
- WP3 - Implementation of an ELIXIR Beacon Network & Registry
- WP4 - Security of ELIXIR Beacon and Beacon Network
- WP5 - Developing Indicators to establish Registry and Beacon as an Emerging ELIXIR Service
- WP6 - Strategic partnerships with national data cohorts and biobanks
- WP7 - Project Management and Communication



# Beacon<sup>+</sup>

This forward looking Beacon interface implements additional, planned features.

**Query**

<b>Dataset</b>	tcga
<b>Reference name*</b>	9
<b>Genome Assembly*</b>	GRCh38 / hg38
<b>Start min Position*</b>	19,500,000
<b>Start max Position</b>	21,975,098
<b>End min Position</b>	21,967,753
<b>End max Position</b>	24,500,000
<b>Alt. Base(s)*</b>	DEL
<b>Bio-ontology</b>	icdot:c50.9: (4065)

## Beacon Implementations

- implementing existing resources with Beacon protocol
- e.g. TCGA cancer variants (structural and SNV)



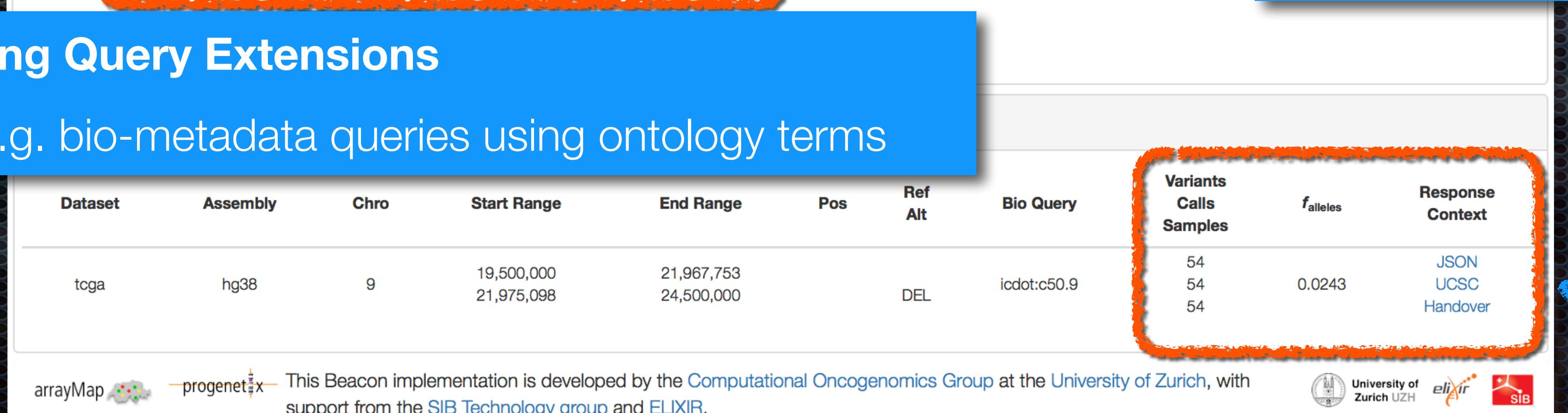
The screenshot shows a header with a green ribbon logo and three tabs: 'Example' (selected), 'DGV Example', and 'CNV Example'. Below the tabs is a button labeled 'Info'.

## Beacon Response

- quantitative (counts for variants, callsets and samples)
- *Handover* to authentication system for data retrieval
- **no exposure** of data beyond standard Beacon response and additional pointer to matched data

## Prototyping Query Extensions

- testing e.g. bio-metadata queries using ontology terms



The screenshot shows a table with columns: Dataset, Assembly, Chro, Start Range, End Range, Pos, Ref Alt, Bio Query, Variants, Calls, Samples, f<sub>alleles</sub>, and Response Context. The 'Bio Query' column contains 'icdot:c50.9'. The 'Variants', 'Calls', and 'Samples' row under 'Response Context' is highlighted with a red box. A blue arrow points from the 'Variants' cell to the 'f<sub>alleles</sub>' cell.

Dataset	Assembly	Chro	Start Range	End Range	Pos	Ref Alt	Bio Query	Variants	Calls	Samples	f <sub>alleles</sub>	Response Context
tcga	hg38	9	19,500,000 21,975,098	21,967,753 24,500,000		DEL	icdot:c50.9	54	54	54	0.0243	JSON UCSC Handover

arrayMap  progenetix  This Beacon implementation is developed by the Computational Oncogenomics Group at the University of Zurich, with support from the SIB Technology group and ELIXIR.

University of Zurich UZH  elixir  SIB 

```
"dataset_allele_responses" : [  
  {  
    "dataset_id" : "arraymap",  
    "call_count" : 171,  
    "sample_count" : 171,  
    "variant_count" : 157,  
    "error" : null,  
    "exists" : true,  
    "external_url" : "http://beacon.arraymap.org",  
    "frequency" : 0.1557,  
    "info" : {  
      "callset_access_handle" : "d5850347-d411-11e7-8c89-ec436516cb41",  
      "description" : "The query was against database \"arraymap_ga4gh\",  
      variant collection \"variants_cnv_grch36\". 171 matched calls for 157 distinct  
      variants. Out of 62033 biosamples in the database, 1098 matched the biosample  
      query; of those, 171 had the variant.",  
    },  
    "note" : "",  
  }]
```

## Beacon Handover Demonstrator

- only exposure of access handle to data stored in secure system
- here one-step authentication and selection of *handover* action; other scenarios possible / likely
- *handover response* **outside of Beacon protocol / system**



# Beacon query => Handover Handle =>

# Beacon+ example implementation using public somatic variation data

Beacon Handover Demonstrator

- only exposure of access handle to data stored in secure system
  - here one-step authentication and selection of *handover* action; other scenarios possible / likely
  - *handover response not managed by Beacon* protocol / system => "Discovery" protocol?



Global Alliance  
for Genomics & Health



# Beaconize Your Resource

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

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

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

[ga4gh.org](http://ga4gh.org)



University of  
Zurich UZH



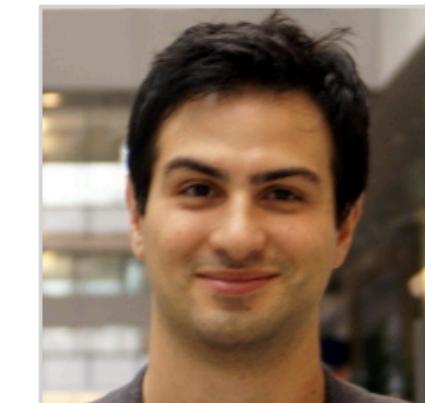
Serena Scollen



Susheel Varma



Michael Baudis



Marc Fiume



Ilkka Lappalainen



Jordi Rambla



David Lloyd



Dylan Spalding



Knox Carey



Anthony Brookes



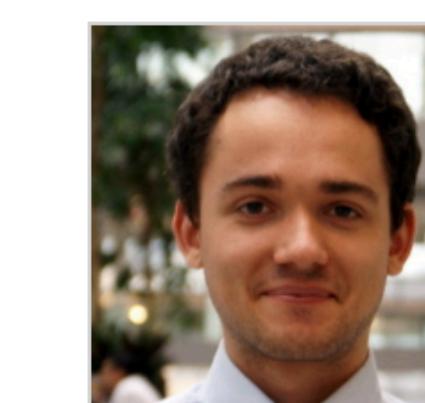
Stephanie Dyke



Aurélien Barré



Rishi Nag



Miro Cupak



Juha Törnroos



Sabela de la Torre



Marc Duby



Saif Ur-Rehman



Bo Gao