

Building a global search engine for genetic data

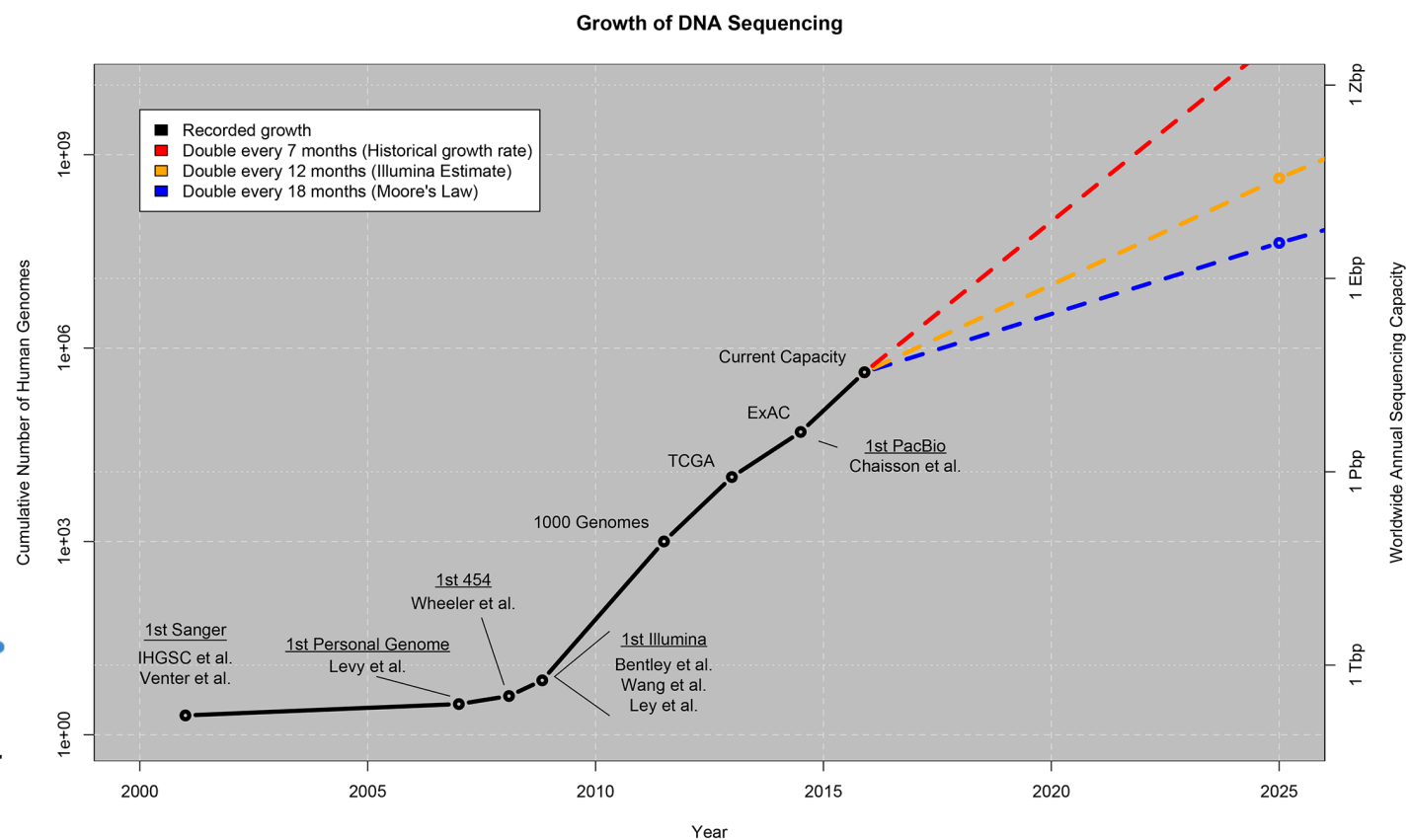
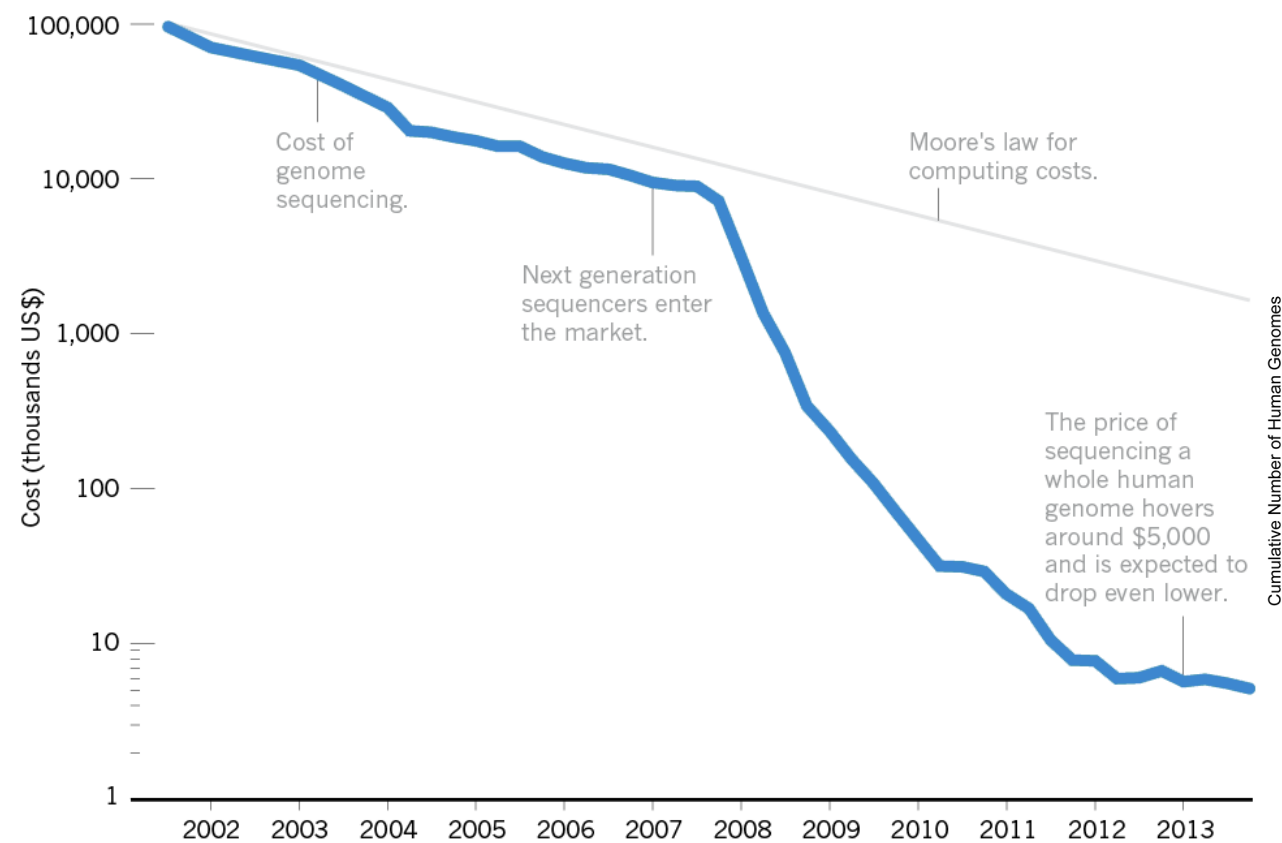
Miro Cupak

VP Engineering, DNASTack

30/04/2018

What and why?

- Beacon Network (<https://beacon-network.org/>)
 - Beacon Project from the Global Alliance for Genomics and Health (GA4GH)
 - largest search and discovery engine of human genomic variation



Beacon

- experiment to test the willingness of international sites to share genetic data in the simplest of all technical contexts
- design principles
 - A beacon has to be technically simple.
 - A beacon has to minimize risks associated with genomic data sharing.
 - It has to be possible to make a beacon publicly available.
- simple web service allowing users to query institution's databases to determine whether they contain a genetic variant of interest
- started in March 2014, quickly gained traction
- receives questions of the form *Do you have information about this mutation?*
- responds with *yes* or *no*, optionally with additional information about the mutation



Global Alliance for Genomics and Health Beacon of Beacons

Beacon: all

Chromosome: 15

Position: 41087869

Alt: A

Beacon	Result
Wellcome Trust Sanger Institute	No
EMBL-EBI	No
UniProt	No
NCBI ClinVar	No
NCBI	No
Leiden Open Variation	No
AMPLab	Yes

A global search engine for genetic mutations.

GRCh37 ▾

e.g. 1 : 100,000 A>C

Search

Quickstart: [Search for a BRCA2 variant](#)

Find genetic mutations shared by these organizations



<https://beacon-network.org>

Search all beac... for allele

GRCh37 ▾
13 : 32936732 G > C

Cancel



Response **All** **None**

- ☒ Found 8
- ☒ Not Found 28
- ☐ Not Applicable 13



Organization **All** **None**

- ☒ AMPLab, UC Berk...
- ☒ Australian Genomics ...
- ☒ Belgian Medical Gen...
- ☒ Bioinformatics Are...
- ☒ BioReference Labo...
- ☒ Brazilian Initiative on ...
- ☒ BRCA Exchange
- ☒ Broad Institute
- ☒ Centre for Genomic ...
- ☒ Centro Nacional de...
- ☒ Children's Mercy Hos...
- ☒ DNASTack
- ☒ ELIXIR
- ☒ EMBL European Bioi...
- ☒ Global Alliance for ...



Altruist

Hosted by [Sequencing.com](#)

Not Found



AMPLab - 1000 Genomes Project

Hosted by [AMPLab](#), [UC Berkeley](#)

Not Found



BioReference

Hosted by [BioReference Laboratories](#)

Not Found



BRCA Exchange

Hosted by [BRCA Exchange](#)

[Hide Metadata](#)

Found

Clinical significance citations	PMID:23108138
HGVS cDNA	c.7878G>C
Condition ID type	OMIM
Date last evaluated	8/10/15
HGVS protein	p.(Trp2626Cys)
Assertion method citation	http://enigmaconsortium.org/documents/ENIGMA_Rules_2015-03-26.pdf

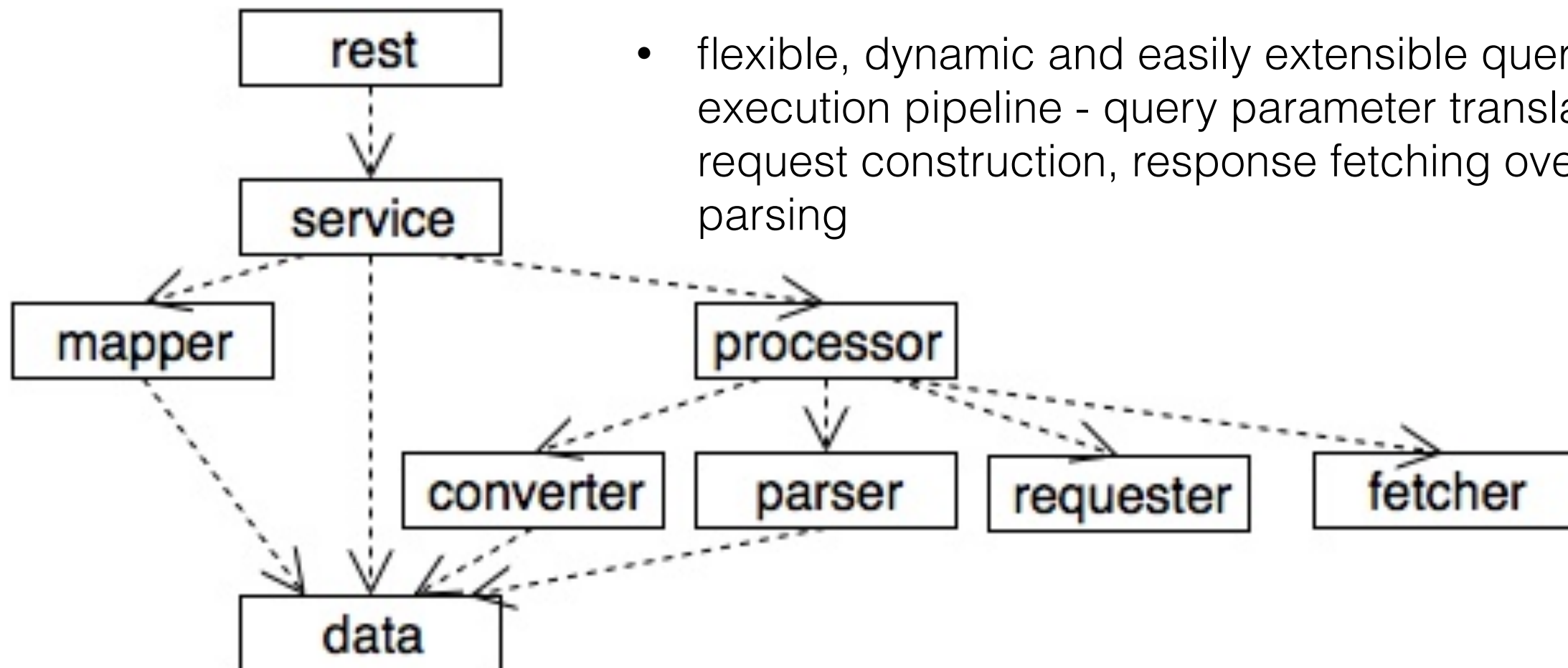
<https://beacon-network.org>

Standard

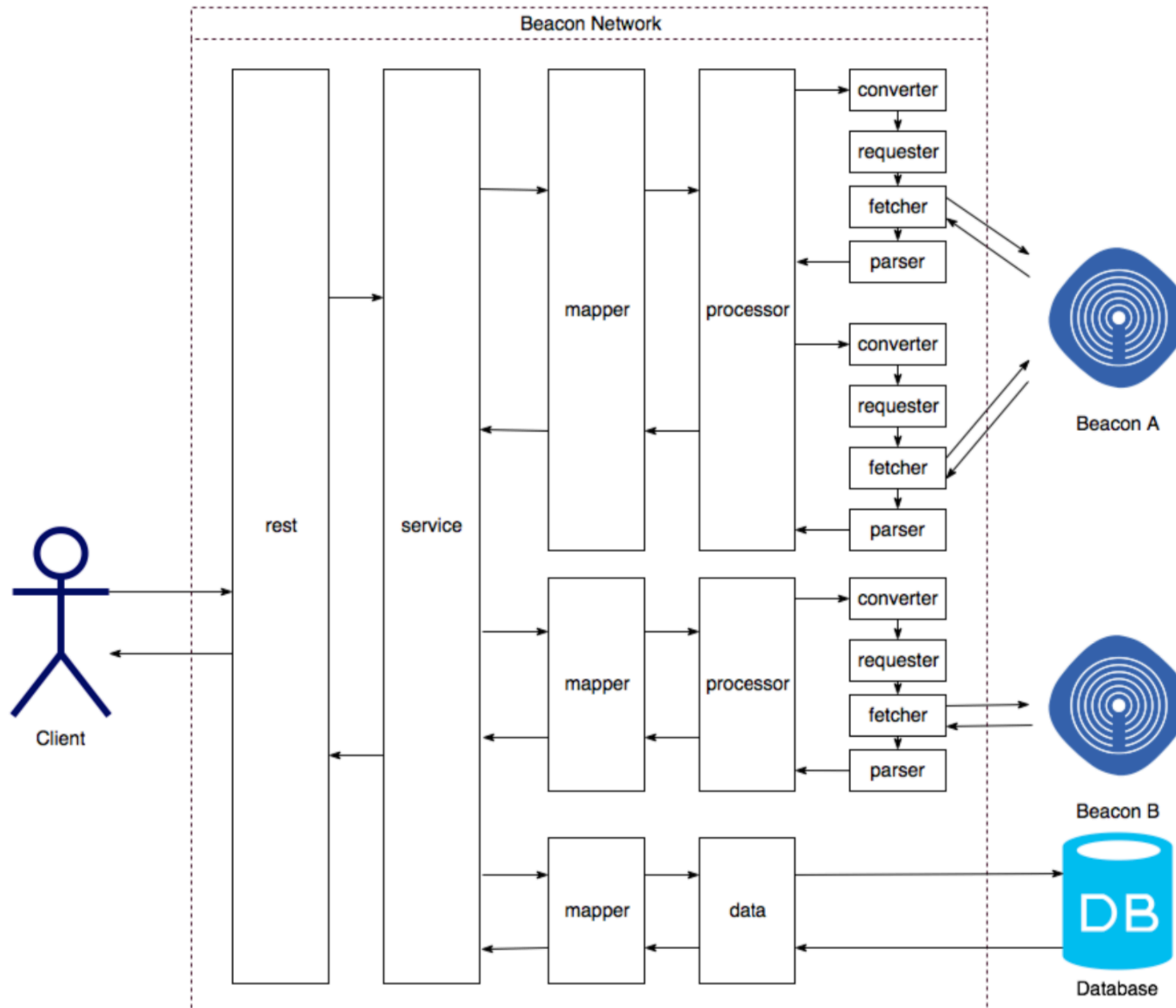
- 0.1 (2014): simple, Apache Avro
- 0.2 (2015): complex, datasets, self description, Apache Avro
- 0.3 (2016): simplified, improved, modular and extensible, tooling, moving towards Protocol Buffers
- 0.4 (2018): flexible, complex variants, data use conditions, developer-friendly, OpenAPI
- now working on 0.4.1, 0.5 coming soon, 1.0 in 2018

Beacon Network

- federation of queries across beacons
- de-facto registry
- programmatically accessible, unified beacon API
- aggregation
- participant resolution
- flexible, dynamic and easily extensible query execution pipeline - query parameter translation, request construction, response fetching over network, parsing

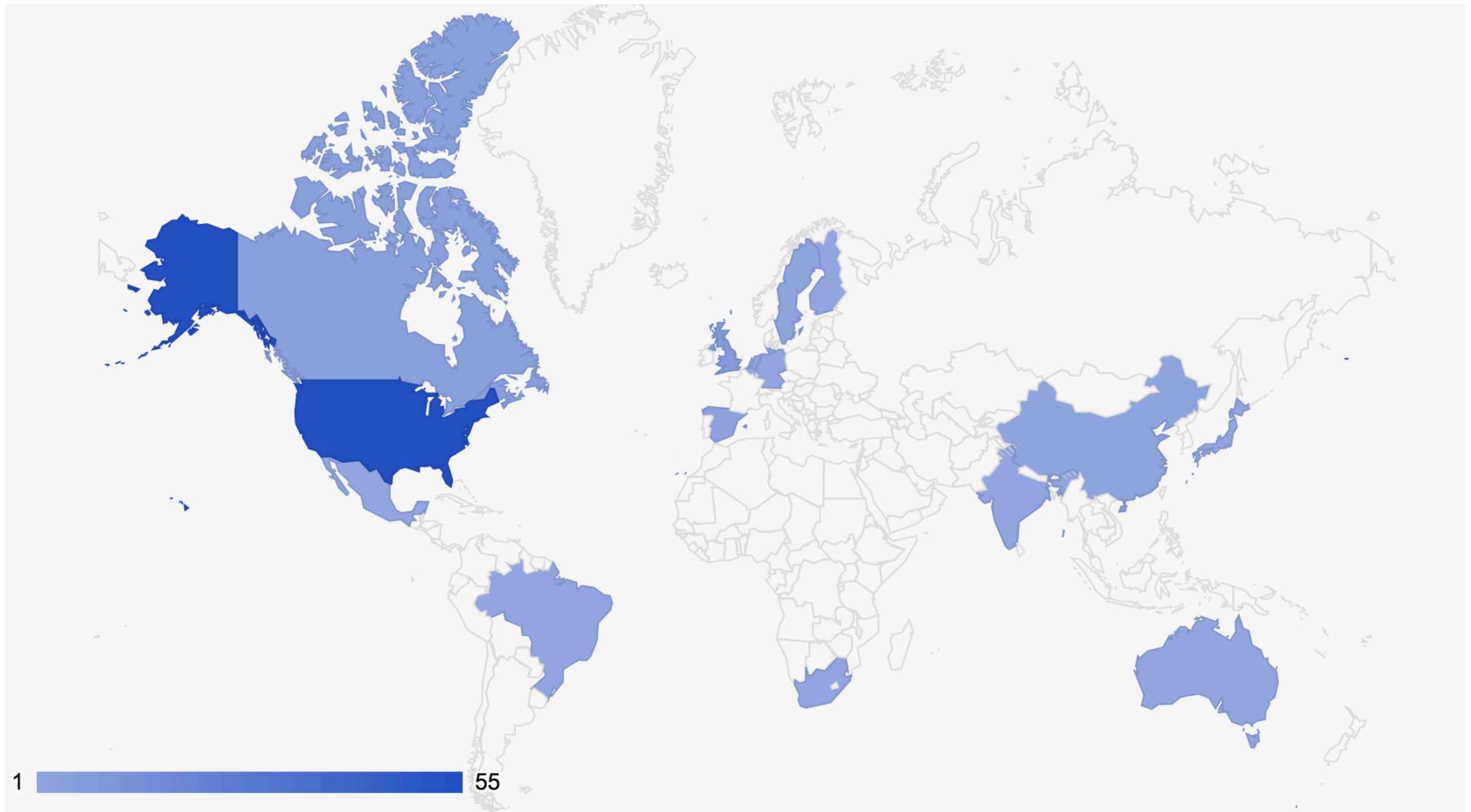


Search execution



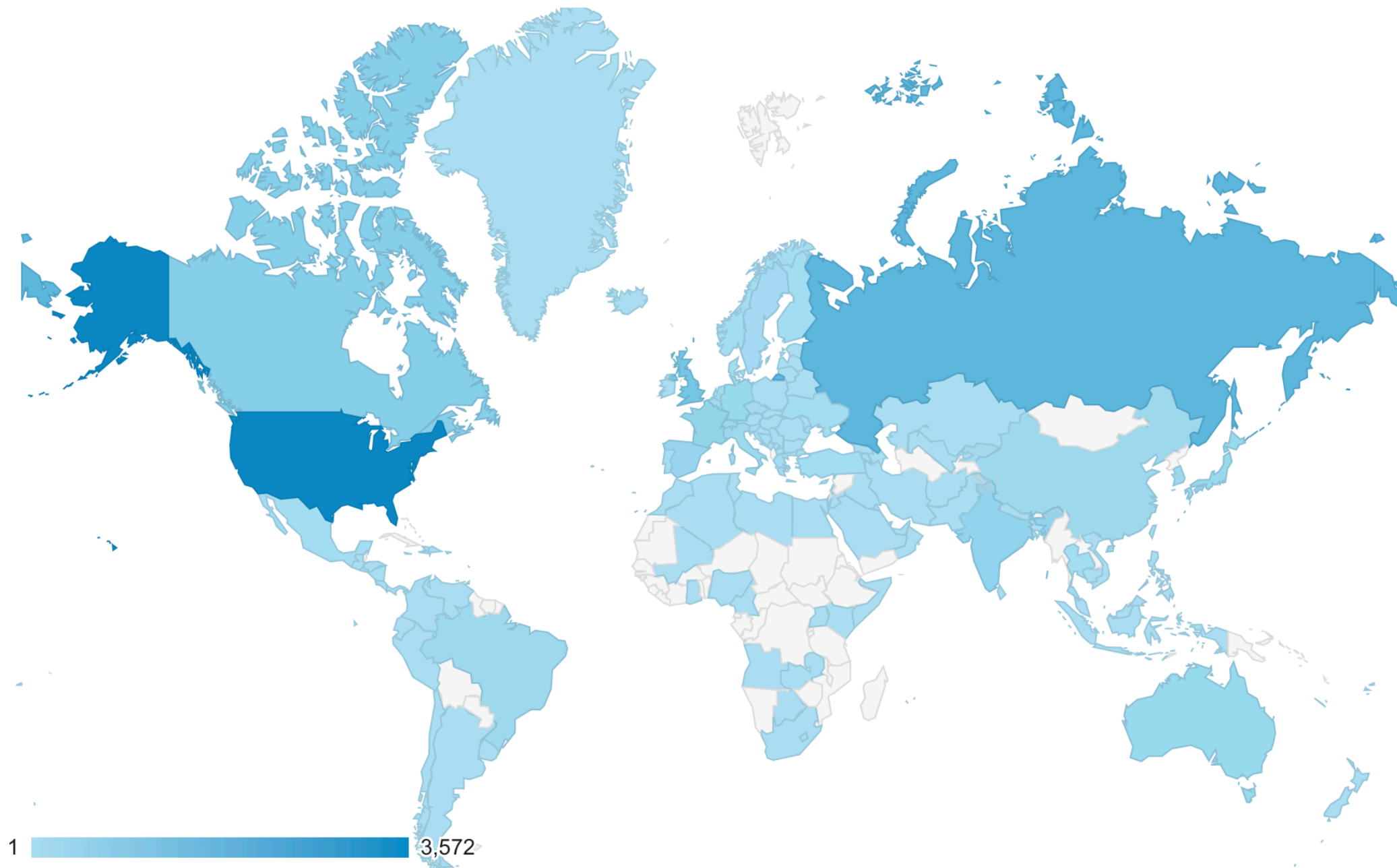
Size

- ~100 installations, 40 institutions, 18 countries, 6 continents



Users

- 13k users, 136 countries



Searches



DNASTack

Follow



Sign in

Get started



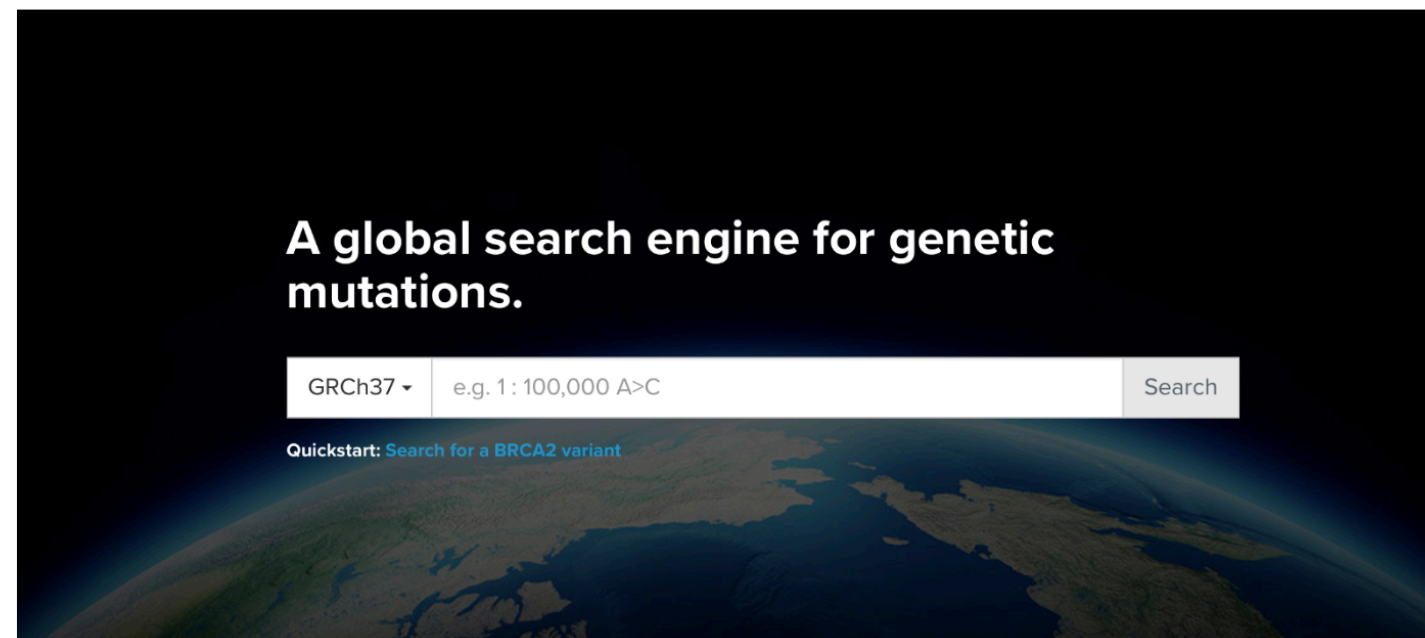
DNASTack

Follow

We believe that making cutting edge #genomics infrastructure broadly accessible will catalyze the realization of #precisionmedicine and save lives globally.

Sep 20, 2017 · 2 min read

Beacon Network served its Millionth search result



Today, the [Beacon Network](https://beacon-network.org) (beacon-network.org) reached 1M searches, passing a significant milestone in our mission to democratize access to the world's genomics data. The DNASTack-developed Beacon Network enables

Other fun stats

- popular parameter values
- variants
- deleteriousness
- rarity
- genes
- disorders and clinical abnormalities

Questions?

<https://github.com/mcupak/beacon-of-beacons>

<https://github.com/ga4gh-beacon/>

<https://dnastack.com/#/team/careers>

<https://mirocupak.com>