TT	Resource Type	Description *(what it is)*	Comments / Relevance *(why/how it might be useful)*	Link(s)	Contributors *(to the records/info in this doc)*
BeaconAPI v2.0	API	API for discovery data (genomic, cohorts, individuals, biosamples...) and facilitate the next step in the data sharing process.	It suggest a default schema for cohorts as a "meeti	https://beacon-project.io/	Jordi Rambla (EGA-CRG)
Data Connect API v1.0	API	Compact API for storing, accessing, listing, and optionally querying Tables of semantically annotated data. A Table row need not be flat; it can be an arbitrary JSON object. Each table is described by a JSON Schema, which constitutes a semantic description of the contents of a Table by referencing externally defined concepts.	- Extensibility where needed: We could agree on a mandatory core set of concepts that are essential to all cohorts, and also take advantage of the extensibility offered by the ability to reference external concepts, even one-off concepts that are specific to a single study. - Storage, transmission, and snapshots: a Data Connect query result is itself a Table, which is a self-describing unit that can be stored, shared, and published as is.	https://ga4gh-discovery.github.io/data-connect/ https://github.com/ga4gh-discovery/data-connect	Jonathan Fuerth (DNAstack)
Phenopackets v2.0 - Cohort Element	Data Model				
Cohort Statistical Profiles	Cohort Summary	A public data source about clinical data that is completely open, transparent and has nothing identifiable. It is de facto a "statistical summary" of real data for a defined phenotype. Includes counts, std devs, distributions, percentages from the usual quantifiable suspects: labs, meds, diagnoses, demographics. Tries to preserve first order correlation among variables.	-Freely shareable, PHI-free	https://clinicalprofiles.org/	
CINECA/IHCC Ontology and Cohort Discovery Platform	Cohort Registry	Built a simple model built on the consensus overlap across CINECA cohorts. Looked at overlap between dictionaries. Formalized as an ontology. The usage of that model right now has been for cohort discovery. Built a UI on top of the data to find cohorts of interest	Would be interested to know what others are doing for value harmonization.		
HDR UK Phenotype Library	Data Model	Building a phenotype library, collect phenotype definitions for disease. Making it computable by generating Phenotyping algorithm workflows in Phenoflow. Once we had defined a phenotype with all of these codes, we have built a cohort discovery tool. Not at the level of IHCC but more at the granular query level.		https://portal.caliberresearch.org/ https://kclhi.org/phenoflow/phenotype/all/ https://www.healthdatagateway.org/	Susheel Varma (HDR UK)
National COVID Cohort Collaborative (N3C)	Data Model	N3C, is building a centralized national data resource — the NCATS N3C Data Enclave — that the research community can use to study COVID-19 and identify potential treatments as the pandemic continues to evolve. Specifically, the N3C will enable the rapid collection and analysis of clinical, laboratory and diagnostic data from hospitals and health care plans.	Patients are selected with a computable phenotype, there is a need for this phenotype to evolve. Have a local script run to extract those records to then provision those data	https://github.com/National-COVID-Cohort-Collaborative/Phenotype_Data_Acquisition/wiki/Latest-Phenotype	
International HundredK+ Cohorts Consortium (IHCC)	Cohort Registry			https://ihccglobal.org/, https://atlas.ihccglobal.org	
Answer ALS	Application / Knowledgebase	A public data portal with filtering options allowing extraction of data at the level of participants thus creating cohorts.		https://dataportal.answerals.org/search	
Open MRS Cohort Builder	Application / Knowledgebase	"perform ad-hoc queries for patients with defined characteristics, and combines multiple queries into more complex ones"		https://guide.openmrs.org/en/Using%20Data/cohort-builder.html	
Deep 6 AI Cohort Builder	Application / Knowledgebase	"Deep 6 AI's Cohort Builder is a software tool researchers can use to find more, better-matching patients for clinical trials. It allows research staff to search structured and unstructured data in the Electronic Medical Record (EMR), looking at millions of documents in a matter of seconds."		https://deep6.ai/update-cohort-builder-3-1/	
NCI ISB-CGC Platform Cohort feature	Application / Knowledgebase	"The Cohort Builder/Data Explorer is an ISB-CGC web interface which allows you to build cohorts based on clinical demographics and molecular filters. Compare patient cohorts with various exploration tools including IGV viewer, image viewers, and analytical visualization."		https://isb-cancer-genomics-cloud.readthedocs.io/en/latest/sections/DataExplorer.html	
Health Catalyst	Terminology / Identifier System	Describe the nuanced approaches to building cohorts in either a patient-centric ("asthma"), episode-centric ("pregnancy") or encounter-centric ("appendectomy") way. "Most organizations involved in healthcare analytics simply use groupings of administrative codes (e.g., ICD9, CPT codes) to define patient cohorts. While this approach represents a valid starting point, relying solely on administrative code groupings often misses patients who should have been included in the cohort to whom the population health strategies should have been applied. Much more robust and clinically credible cohort definitions can be achieved by the addition of other factors such as supplemental administrative codes (e.g., ICD9 code for wheezing for the asthma cohort), sentinel medications (e.g., patients taking metformin for the Type 2 diabetes cohort), and clinical observations such as results of imaging studies or lab tests (e.g., cardiac ejection fraction and brain natriuretic peptide [BNP] for the heart failure cohort)."	More healthcare centric than research centric but valuable insights.	https://www.healthcatalyst.com/Defining-Patient-Populations	
HDR UK Cohort Discovery Service	API	A cohort discovery services that federates cohort queries across multiple nodes of healthcare datasets across UK	Supports OMOP / FIHR / i2b2 with a relatively simplistic cohort query API with inclusion and exclusion criteria that uses the standard terminologies and Phenotype definitions	https://www.healthdatagateway.org/about/cohort-discovery	Susheel Varma (HDR UK)
OHDSI APHRODITE	Data Model	OHDSI's Auto- mated PHenotype Routine for Observational Definition, Iden- tification, Training and Evaluation		https://github.com/OHDSI/Aphrodite	Susheel Varma (HDR UK)
PheKB	Application / Knowledgebase	Knowledgebase for discovering phenotypes for EHRs		https://phekb.org/	Susheel Varma (HDR UK)
SAIL Concept Library	Data Model	Next version of the HDR UK Concept Library		https://conceptlibrary.demo.saildatabank.com/phenotypes/	Susheel Varma (HDR UK)
Finngen Risteys	Application / Knowledgebase	Finngen's Risteys platform relates phenotypes temporally, listing those phenotypes that a patient is likely to exhibit either before or after exhibiting an- other (e.g. the onset of depression after exhibiting bipolar dis- order)		https://risteys.finngen.fi/	Susheel Varma (HDR UK)
Phenoflow	Data Model	Database extraction code generated from the HDR UK Phenotype Library		https://kclhi.org/phenoflow/phenotype/all/	Susheel Varma (HDR UK)
OpenSafely Codelists / Study Definitions	Data Model	Pythonic Cohort Definition and extraction system		https://www.opencodelists.org/	Susheel Varma (HDR UK)
Pathling	API	Advanced FIHR Data Analytics Server - Aggregation, Search		https://pathling.csiro.au/	John Grimes (CISRO)
HL7 Clincal Quality Language (CQL)	Data Model	CQL has matured into a capable and flexible specification for representing clinically-focused logic, and is being used in a broad variety of applications from cohort definition and quality measurement, to clinical decision and cognitive support, computable guidelines, and public health reporting. (from https://cql.hl7.org/)	Intended mainly to support clincal quality reporting, it is a tool that can be used to define cohorts	https://cql.hl7.org/	Robert Carroll (VUMC / AoU DRC)
OHDSI Atlas	Application / Knowledgebase	Atlas is a web tool built to support querying and analytics on OMOP backends.	Atlas includes a cohort definition tool and json specification that defines cohorts that can be shared and used among OMOP/Atlas instances.	https://atlas.ohdsi.org/	Robert Carroll (VUMC / AoU DRC)
Criteria2Query	Application / Knowledgebase	Criteria2Query (C2Q) is an automatic cohort identification system. It enhances human-computer collaboration to convert complex eligibility criteria text into more accurate and feasible cohort SQL queries.	C2Q uses NLP to transform free text into structured queries. Could be helpful to quickly assess/transform cohort definitions into structured details.	https://github.com/OHDSI/Criteria2Query	Robert Carroll (VUMC / AoU DRC)
Clincal Trial Knowledge Base	Application / Knowledgebase	Knowledge base of discrete clinical trial eligibility criteria equipped with a web-based user interface for querying and aggregate analysis of common eligibility criteria.	Built using C2Q, it could be a helpful reference for common clinical trial enrollment criteria from clinicaltrials.gov	http://ctkb.io/?#/	Robert Carroll (VUMC / AoU DRC)
HL7 FHIR Group Resource	Data Model	Core FHIR resource that defines a group of specific people, animals, devices etc. by enumerating them, or by describing qualities that group members have. The group resource refers to other resources, possibly implicitly. Groups are intended to be acted upon or observed as a whole; e.g. performing therapy on a group, calculating risk for a group, etc. This resource will commonly be used for public health (e.g. describing an at-risk population), clinical trials (e.g. defining a test subject pool) and similar purposes.	The functionality is limited with regards to defining cohort logic, but it does cover the basics and can also integrate with explicit references to data.	https://www.hl7.org/fhir/group.html	Robert Carroll (VUMC / AoU DRC)

TT	Resource Type	Description *(what it is)*	Comments / Relevance *(why/how it might be useful)*	Link(s)	Contributors *(to the records/info in this doc)*

TT	Resource Type	Description *(what it is)*	Comments / Relevance *(why/how it might be useful)*	Link(s)	Contributors *(to the records/info in this doc)*

TT	Resource Type	Description *(what it is)*	Comments / Relevance *(why/how it might be useful)*	Link(s)	Contributors *(to the records/info in this doc)*

TT	Resource Type	Description *(what it is)*	Comments / Relevance *(why/how it might be useful)*	Link(s)	Contributors *(to the records/info in this doc)*

TT	Resource Type	Description *(what it is)*	Comments / Relevance *(why/how it might be useful)*	Link(s)	Contributors *(to the records/info in this doc)*

TT	Resource Type	Description *(what it is)*	Comments / Relevance *(why/how it might be useful)*	Link(s)	Contributors *(to the records/info in this doc)*

TT	Resource Type	Description *(what it is)*	Comments / Relevance *(why/how it might be useful)*	Link(s)	Contributors *(to the records/info in this doc)*

TT	Resource Type	Description *(what it is)*	Comments / Relevance *(why/how it might be useful)*	Link(s)	Contributors *(to the records/info in this doc)*

TT	Resource Type	Description *(what it is)*	Comments / Relevance *(why/how it might be useful)*	Link(s)	Contributors *(to the records/info in this doc)*

ENUM NAME	ENUM VALUES	VALUE DEFINITION/EXAMPLES	NOTES/COMMENTS
Resource Type	Data Model	Community standard providing a model for representation of data (use cases may include data collection, integration, reasoning, or exchange, etc (e.g. GA4GH-G2P, FHIR).	Most valueable/relevant models are those for structuring assertions and their evidence/provenance. Scope may also include models for representation of variants or conditions/phenotypes (the entities variant annotations/assertoins are typically about).
Resource Type	API	e.g. Beacon, Search	
Resource Type	Terminology / Identifier System	Community standard that provides a set of terms, codes, or identifiers for concepts in the domain (e.g. ECO, SO). Sources of identifiers/codes for concepts, not patterns/schema for representing data.	
Resource Type	Application / Knowledgebase	A knoweldge resource or application that may use its own data model and/or community standards	
Resource Type	Cohort Study	A pre-defined collection of cohorts conducted by an organisation with a specific theme	
Resource Type	Cohort Registry	a pre-defined collection of cohort studies curated by an organisation	
Resource Type	Cohort Summary		
Resource Type	Other (enter new type)	If the resource doesnt fit the aboove types, add new type here.	

Timestamp	Name	Email	Driver Project or Initiative Affiliation	Use Case/User Story Description	Use Case Pain Points	Functional Requirements	Additional Comments	Would you like to be involved in GA4GH Cohort Representation efforts?
5/26/2021 12:19:54	Susheel Varma		Health Data Research UK	Computable Cohort Representation for cohort discovery, extraction and comparatative statistics	Lack of standardised cohort representation, lack of interoperability			
5/31/2021 12:08:50	Francis Jeanson		NA	I currently consult with a number of research organizations and cohorts are typically organized by studies. Cross-study cohort discovery is essential for optimized discovery and identifying key standard raw variables and computed variables to include in the cohort discovery across studies would be of great value.	For example, some cohort studies will encode months since birth, or weeks, or years, or date of birth preventing cross cohort grouping. Other common demographic variables may also be non-standard. Could a meta mapping provide a solution to these cross-study and cross-modality groupings?			
6/15/2021 16:01:07	Ian Fore	ian.fore@nih.gov	Cancer Research Data Commons	From a CRDC DP point of view there is not a single answer to these questions. From a work stream point of view Data Connect aims to provide the ability to federate and aggregate data from multiple sources to assemble datasets which are often, but not only, cohorts. Data Connect aims to integrate with model descriptions, semantic services and mappings in order to achieve this.	Just as a cohort is a particular collection of subjects which meet specific criteria , the data required to be represented is similarly specific to the scientific purpose of the investigator.	Allow for model representation of different sources, including annotation linking to semantic definitions (data elements, vocabularies, ontologies). Ability to create, share and use mappings. Likely need to deal with cohorts where data on individuals will need to stay within specific repositories. i.e. the cohort cannot be assembled in one place.	These comments are brief, need more expansion than is possible in a form.	I am already involved
7/6/2021 16:25:42	Kees van Bochove	kees@thehyve.nl	CINECA / The Hyve	Cohort data for OHDSI studies is represented using OMOP, the cohort definitions in ATLAS JSON	Would be great (although not easy...) to have an interoperable computable cohort definition standard that goes across OMOP / FHIR / HL7 CQL etc.	Please have a look at the cohort definition structure in OMOP (see https://ohdsi.github.io/TheBookOfOhdsi/Cohorts.html)	Would love to join the effort if more people are thinking in this direction!	Yes
7/13/2021 2:58:45	Ronald Cornet	r.cornet@amsterdamumc	EJP-RD	Rare diseases registries, organized by ERNs, aim to provide complete coverage of the (EU) patient population. Overall data quality is a challenge that may be out of scope, but case completeness is hard to determine.	One of the main challenges I see is the baseline comparison. I.e., what is the reference population from which included patients come? Without clear cohort description, such basic figures are hard to determine, which means that many analyses are hard to perform.	Ideally provide a "computational" description of the cohort, e.g., one that could be run as a query on the baseline population, should all data be available. This will help to make selection criteria explicit and potentially bias as well.	In my opinion, a basic schema relying on a few ontologies could do much of the job. This could lead to, e.g., a SPARQL-query as a way of cohort description.	Yes
7/30/2021 17:39:43	Nicola Mulder	nicola.mulder@uct.ac.za	H3Africa	Metadata from H3Africa datasets and biospecimens are collected from the H3Africa archive (which submits data to EGA) and the 3 H3Africa biorepositories into a database. The data are searchable in the H3Africa catalogue from which users can select datasets and biospecimens and submit an access request to the H3Africa DBAC. We have now mapped information on the studies (cohorts), phenotypes and other data types to the GECKO ontology and will soon use this to better display and describe the individual project/cohort information.	One of the issues is that the projects are collecting more data than they are submitting to the EGA. The information we have on what is collected is from case report forms. It is difficult to dive deeper into the studies to be able to represent the data more accurately due to sharing issues. The metadata simply states what data is collected, the values themselves for each individual are not available. The projects may also be collecting and representing the phenotypes differently. Representing longitudinal data is also a challenge.	Provide a suggested model and methods for mapping to that model. This won't overcome the sharing issue to enable individual-level information. We also have issues with matching IDs between participants and samples in datasets versus biorepositories as some biorepositories have their own sample IDs based on their LIMS, and these are assigned after data are submitted.		Yes

DOI/URI	Title	Author	Journal/Source	Comments		
https://doi.org/10.	Transforming and evaluating electronic health record disease phenotyping algorithms using the OMOP common data model: a case study in heart failure	Papez et al	JAMIA Open	Compares phenotype definitions in source vs in OMOP mapped data		
https://discovery.u	Evaluation of Semantic Web Technologies for Storing Computable Definitions of Electronic Health Records Phenotyping Algorithms.	Papez et al	AMIA Proceedings	Use RDF/OWL for defining computable representations of phenotype algorithms		
https://doi.org/10.	Toward cross-platform electronic health record-driven phenotyping using Clinical Quality Language	Brandt et al	Learning Health Systems	CQL to OMOP strategy		
https://doi.org/10.	Extracting research-quality phenotypes from electronic health records to support precision medicine	Wei et al	BMC Genome Medicine	Based on eMERGE		
https://www.youtu	Phenotype Development & Evaluation WG Update (Presentation by Gowtham Rao • Oct. 5 OHDSI Community Call)	Gowtham Rao	YouTube	Useful to get an overview of phenotype definitions in OHDSI/OMOP		

Standard	Name of Potential Speaker	Planner				
FHIR Research Study	Bob or Hugh Glover	IF	https://www.hl7.org/fhir/researchstudy.html			
OMOP	Kees / perhaps someone from health da	JR, LF	https://ohdsi.github.io/TheBookOfOhdsi/Cohorts.html			
IHCC Cohort Atlas	Melanie or Nicky	NM	https://ihccglobal.org/membercohorts/			
Bioschemas			https://bioschemas.org/profiles/Dataset/0.3-RELEASE-2019_06_14			
NYU Data Catalog			https://datacatalog.med.nyu.edu			
DATS			https://www.nature.com/articles/sdata201759			
OMICS DI			https://www.omicsdi.org			
DataMed			https://datamed.org			
Childhood Cancer Initiative (CCDI) Data Catalog						
NCPI Data Catalog			https://anvilproject.org/ncpi/data			
Maelstrom		LF, JR	https://www.maelstrom-research.org/			
CKAN		IF	https://ckan.org/			
Google Dataset Search						
https://datasetsearch.research.google.com						
Searches like the following are interesting						
“framingham cohort”						
“copdgene”						
These aren’t the kind of cohort queries we want to enable – but is a useful route into how cohorts are represented.						