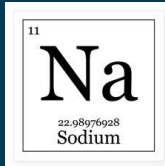


KC2 webinar: Data GUIDS

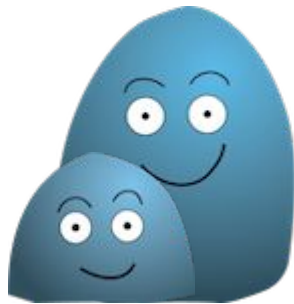
Info



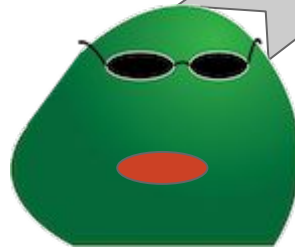
Meet the DCPPC GUID cast



DataCite DOI



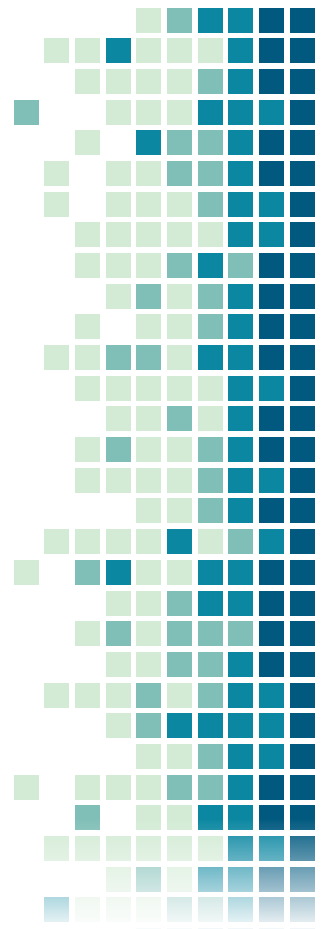
Minid
ARK



Data GUID



Compact
Identifier



Data GUIDs



dg.4503/0008cb73-d9eb-4c54-bc6a-ac140b05ac79



Namespace



Local Identifier

Data GUIDs are designed for scalable object referencing for millions of objects in a data commons.



Data GUIDs

Scalable

In-commons resolution means lookups can stay entirely local

Free

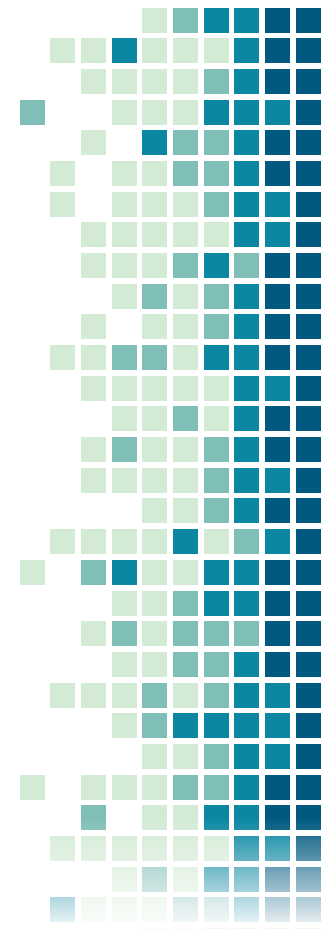
Lookups and minting are free.

Non-authoritative namespaces

Decentralized lookups and minting

Standardized API for lookups

Driving the GA4GH DOS standard for resolution



Who uses them?

dg.4503

DATA COMMONS PILOT

DATA STAGE

dg.7C5B


**Environmental
Data Commons**

dg.F82A1A


Gabriella Miller
Kids First
PEDIATRIC RESEARCH PROGRAM
Data Resource Center



When should you use them?

Any data

Track any sort of data, whether publication results or temporary files

When lookup speed matters

1000s of lookups per second when doing computational work



Resolution and Brokering Software

Indexd

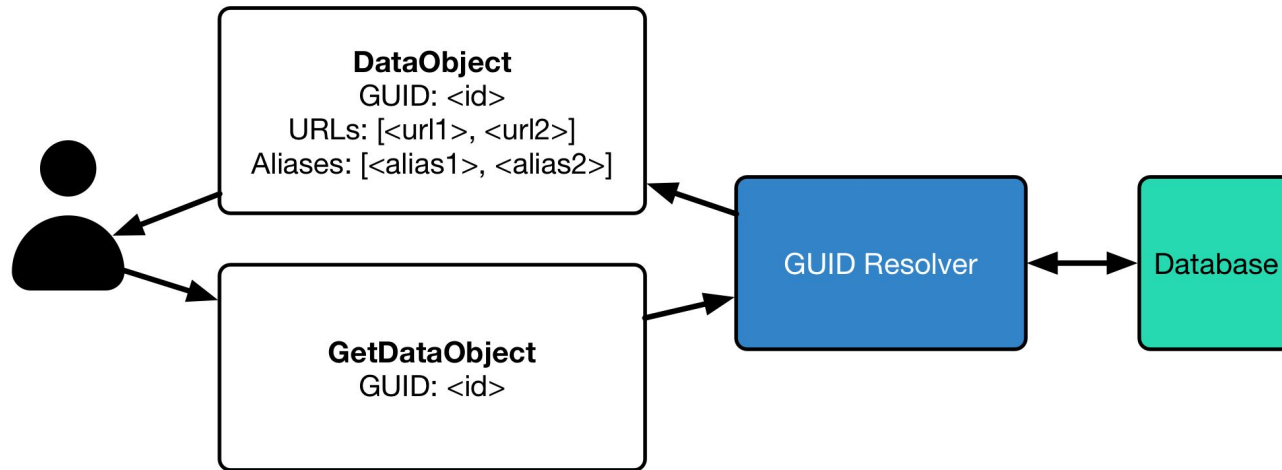
Open source index service which supports brokering developed for Gen3 Data Commons

<https://github.com/uc-cdis/indexd>

PostgreSQL backend, supports multiple checksum types, object ACLs, and much more.



Resolution and Brokering Software



<https://github.com/ga4gh/data-object-service-schemas>

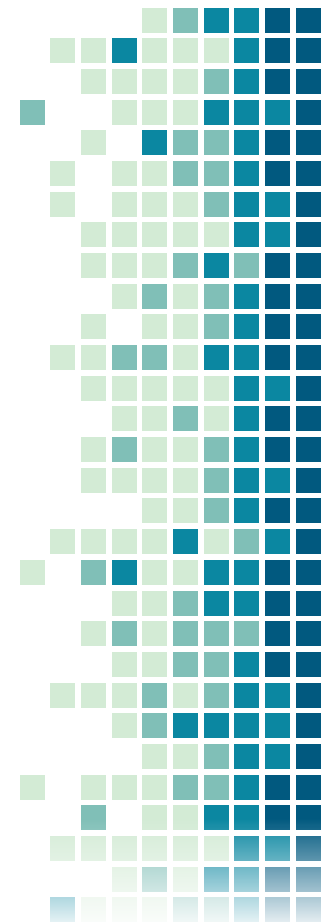
Resolution and Brokering Software

```
{
  "acl": ["phs000424.v7.p2", "gtex"],
  "baseid": "98ec8de2-fdb1-4717-8b36-416463533d14",
  "created_date": "2018-07-17T20:16:28.142644",
  "did": "dg.4503/0008cb73-d9eb-4c54-bc6a-ac140b05ac79",
  "file_name": "GTEx-131YS-0126-SM-5KM1M.Aligned.sortedByCoord.out.patched.cram.crai",
  "form": "object",
  "hashes": {
    "md5": "c0036ee1b032e43cf39b7bae1e34da75"
  },
  "metadata": {},
  "rev": "27cf6283",
  "size": 149171,
  "updated_date": "2018-07-17T20:16:28.142653",
  "urls":
["gs://gtex-v7-exchange-011618/rnaseq_cram_files/GTEx-131YS-0126-SM-5KM1M.Aligned.sortedByCoord.out.patched.cram.crai",
"s3://nih-nhgri-datacommons/gtex_v7/rnaseq/GTEx-131YS-0126-SM-5KM1M.Aligned.sortedByCoord.out.patched.cram.crai"],
  "urls_metadata": {
    "gs://gtex-v7-exchange-011618/rnaseq_cram_files/GTEx-131YS-0126-SM-5KM1M.Aligned.sortedByCoord.out.patched.cra
m.crai": {},
    "s3://nih-nhgri-datacommons/gtex_v7/rnaseq/GTEx-131YS-0126-SM-5KM1M.Aligned.sortedByCoord.out.patched.cram.crai
": {}
  },
  "version": null
}
```



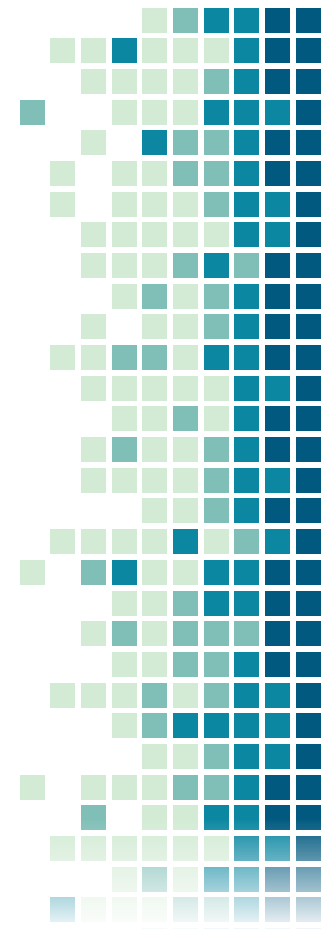
Core Metadata

Name	Description	Required
@id	Primary identifier expressed as URL/URI. See discussion of identifier . JSON-LD uses @id , RDFa 1.1 uses resource , microdata uses itemid .	Y
@type	Should in most cases be Dataset .	Y
identifier	Data GUID expressed as URL. One of the identifiers can be a checksum using propertyValue with the specific checksum algorithm and checksum.	Y
contentSize	The number of bytes representing the actual object	Y
contentUrl	Actual bytes of the media object, for example the image file or video file.	N

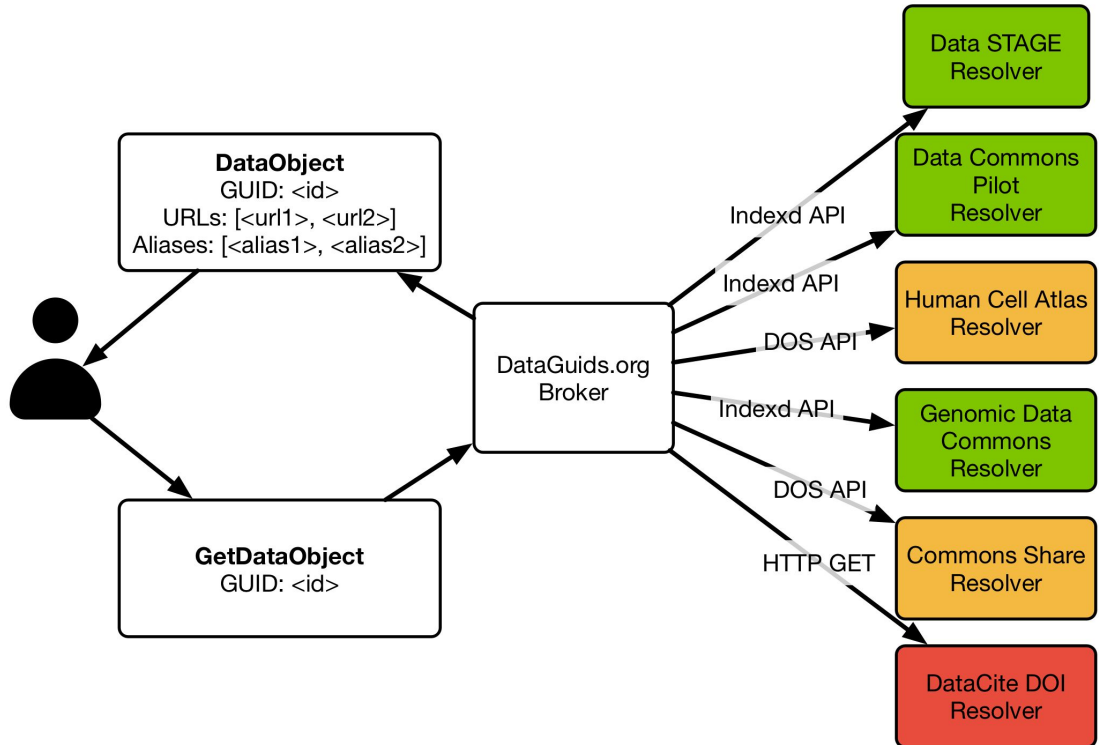


Core Metadata

Name	Description	Required
dateCreated	The date on which the CreativeWork was created.	Optional
name	The name or title of the resource.	Optional
author	The author(s) of the dataset. Schema.org uses creator as synonym.	Optional



Resolution and Brokering Software



Resolve Data GUIDs Elsewhere

<http://identifiers.org/ga4ghdos/dg.4503/a5d79375-1ba8-418f-9dda-eb981375e516>

<http://n2t.net/ga4ghdos:/dg.4503/a5d79375-1ba8-418f-9dda-eb981375e516>



DataGUIDs.org

Data GUID

[HOME](#) [ABOUT](#) [FEATURES](#) [RESOLVER](#)

**dg.4503/
00e6cfa9-
a183-
42f6-
bb44-
b70347106bbe**

Domain Neutral Persistent Identifiers for the Modern Web

Data GUIDs enable you to mint globally unique identifiers for data in a scalable distributed architecture.

[Learn more](#)

