

FHIR Bulk Data API

Dan Gottlieb (@gotdan), Central Square Solutions LLC



HL7 FHIR DevDays 2020, Virtual Edition US, June 15–18, 2020 | @HL7 @FirelyTeam | #fhirdevdays | www.devdays.com/us

slides: <https://bit.ly/dd-2020-bulk>

Agenda

- Why a FHIR Bulk Data API?
- Technical Architecture (v1.0)
- Open Source Tools
- Adoption!
- 2020 Argonaut Project (v1.2)
- Next steps and how to get involved

Extending FHIR to Population Level Datasets



Patient

FHIR Rest API



Panel

FHIR Bulk Data API



Population

Use Cases

- Internal clinical data warehouse for study cohort identification
- Machine learning startup obtaining training data from cloud EHR
- Claims in EHR to provide comprehensive view
- Integration population health system with EHR system
- Transferring records from one EHR to another
- Payer database to assess care quality
- Reportable disease submission or other registry

Let's enhance FHIR® to support population level data access

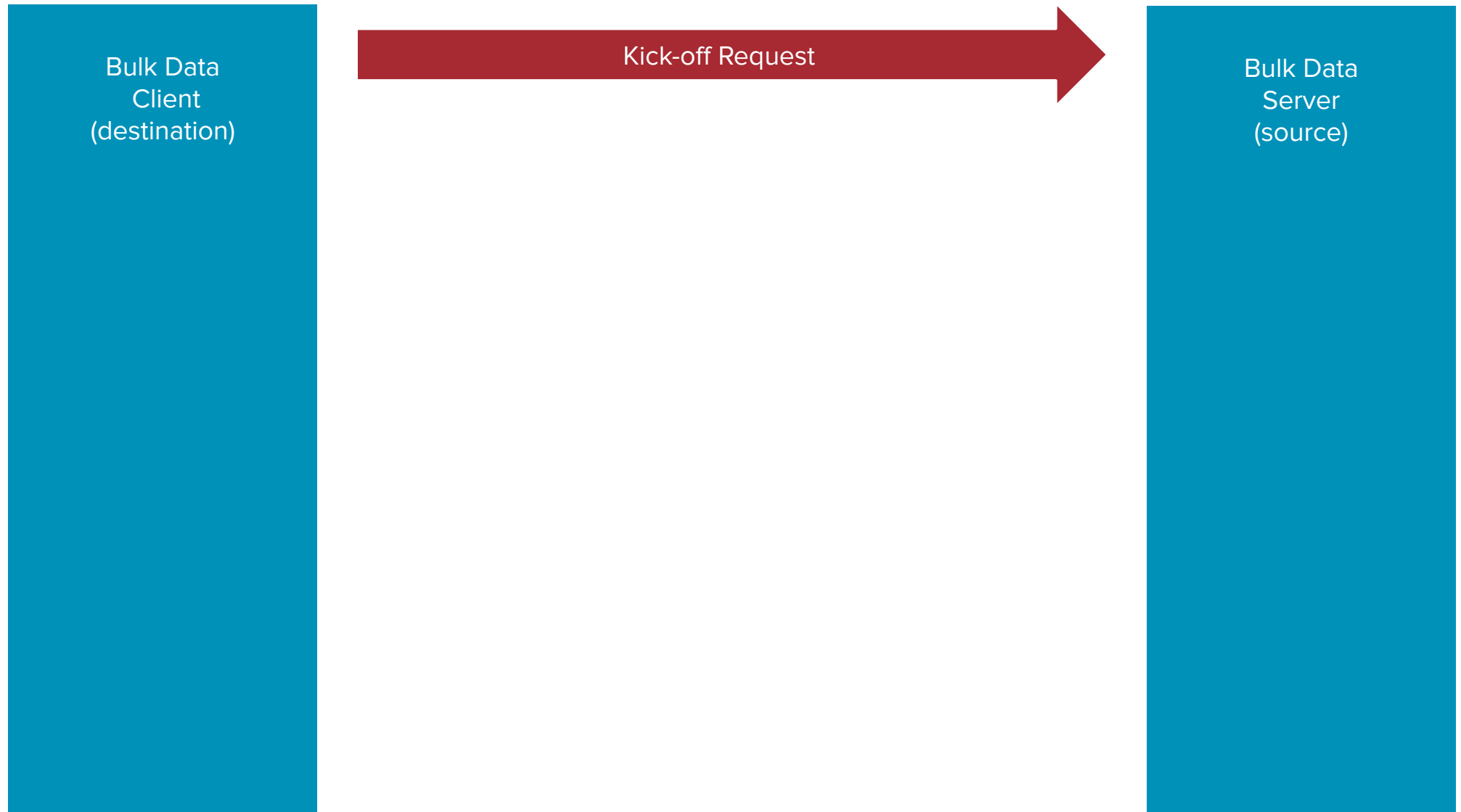
- **FHIR Resources** as a standard data model to simplify data parsing and mapping
- **FHIR Operation API** to initiate the data extracts
- **SMART Backend Services Authorization** as security model

Focused Scope - out of scope in v1

- Legal framework for sharing data between partners needs to be set up out-of-band (BAAs, SLAs, DUAs)
- Real-time data (but, data loaded through bulk data can be supplemented at with synchronous FHIR REST API calls)
- Patient matching (but, it's possible to include identifiers like subscriber number in FHIR resources)
- Data transformation (but, can serve as a foundation for this)

Technical Architecture Bulk Data Access Implementation Guide STU1 (v1.0)

Kick-off Request



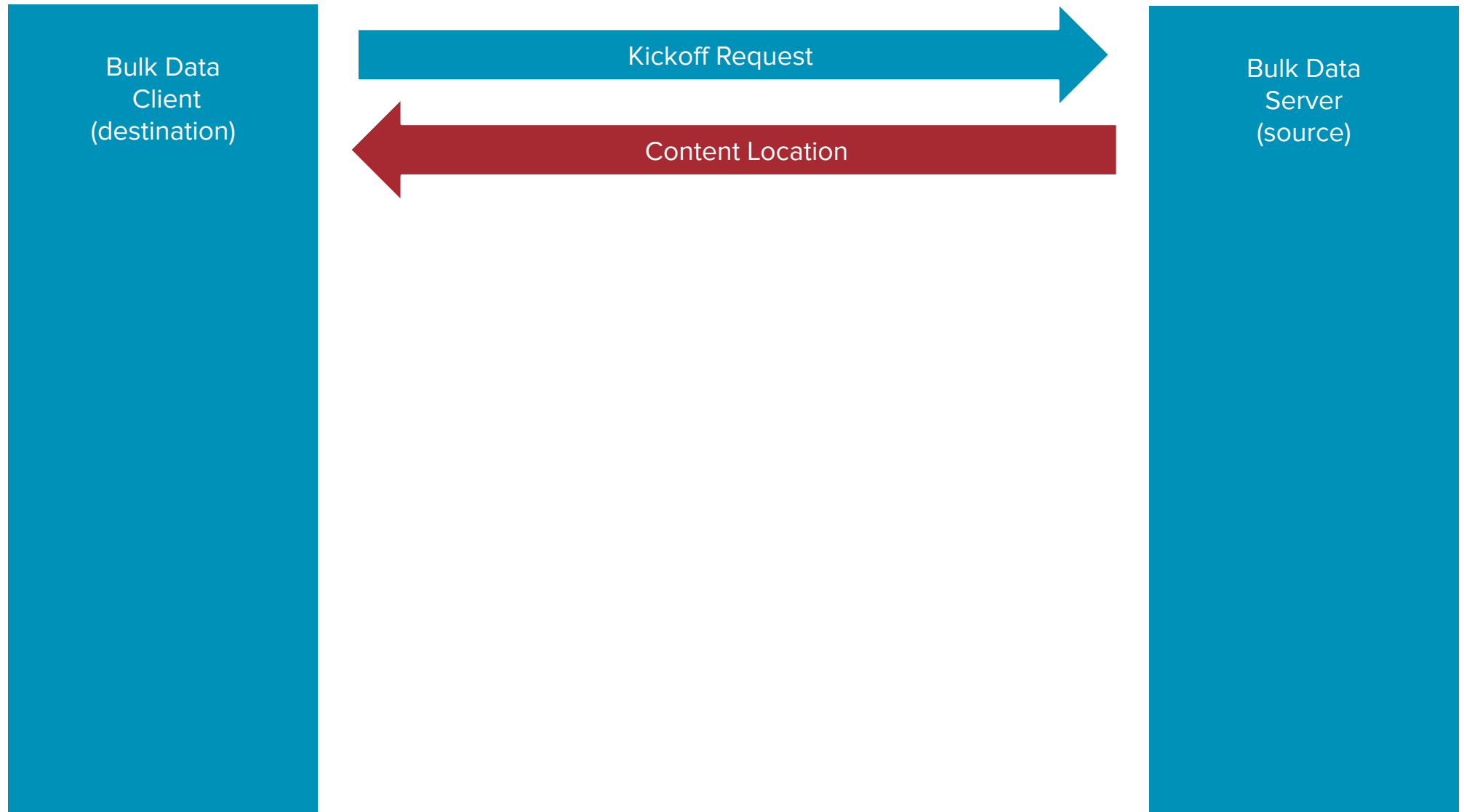
Kick-off Request

- Asynchronous requests with status polling
Prefer: respond-async
- FHIR Operation for all data on all patients
[FHIR Server Base]/Patient/\$export
- FHIR Operation for all data on a group of patients (eg. research cohort, plan members)
[FHIR Server Base]/Group/[group id]/\$export
- FHIR Operation for all data on the server
[FHIR Server Base]/\$export

Kick-off Operation Parameters

<code>_outputFormat</code>	The format for the generated bulk data files Currently, only ndjson is supported
<code>_since</code>	Filter results by FHIR resource modified date FHIR instant timestamp
<code>_type</code>	Filter results by comma delimited list of FHIR resource types
<code>_typeFilter</code>	Experimental syntax to limit data returned using FHIR search parameters

Kick-off Response

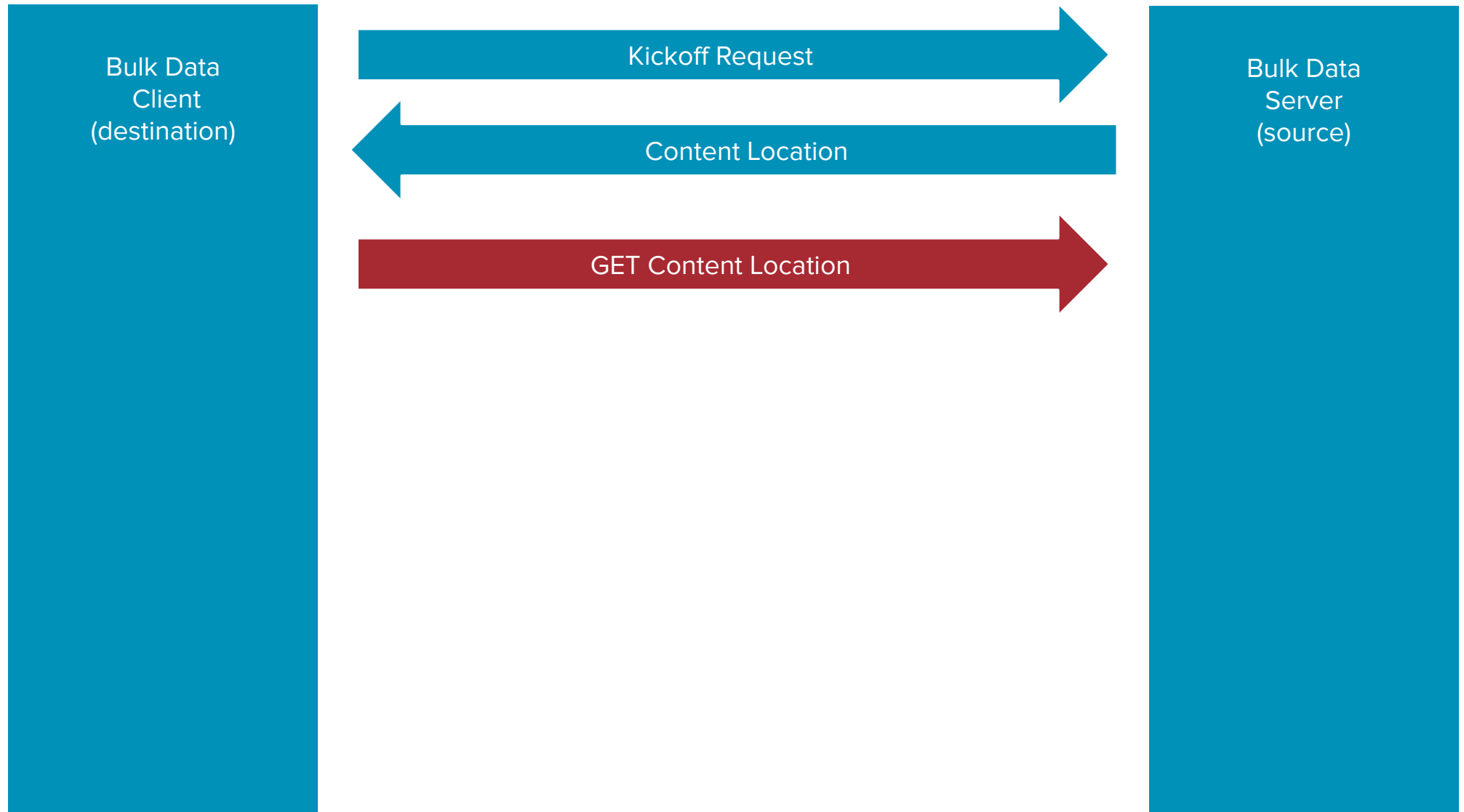


Kick-off Response Header

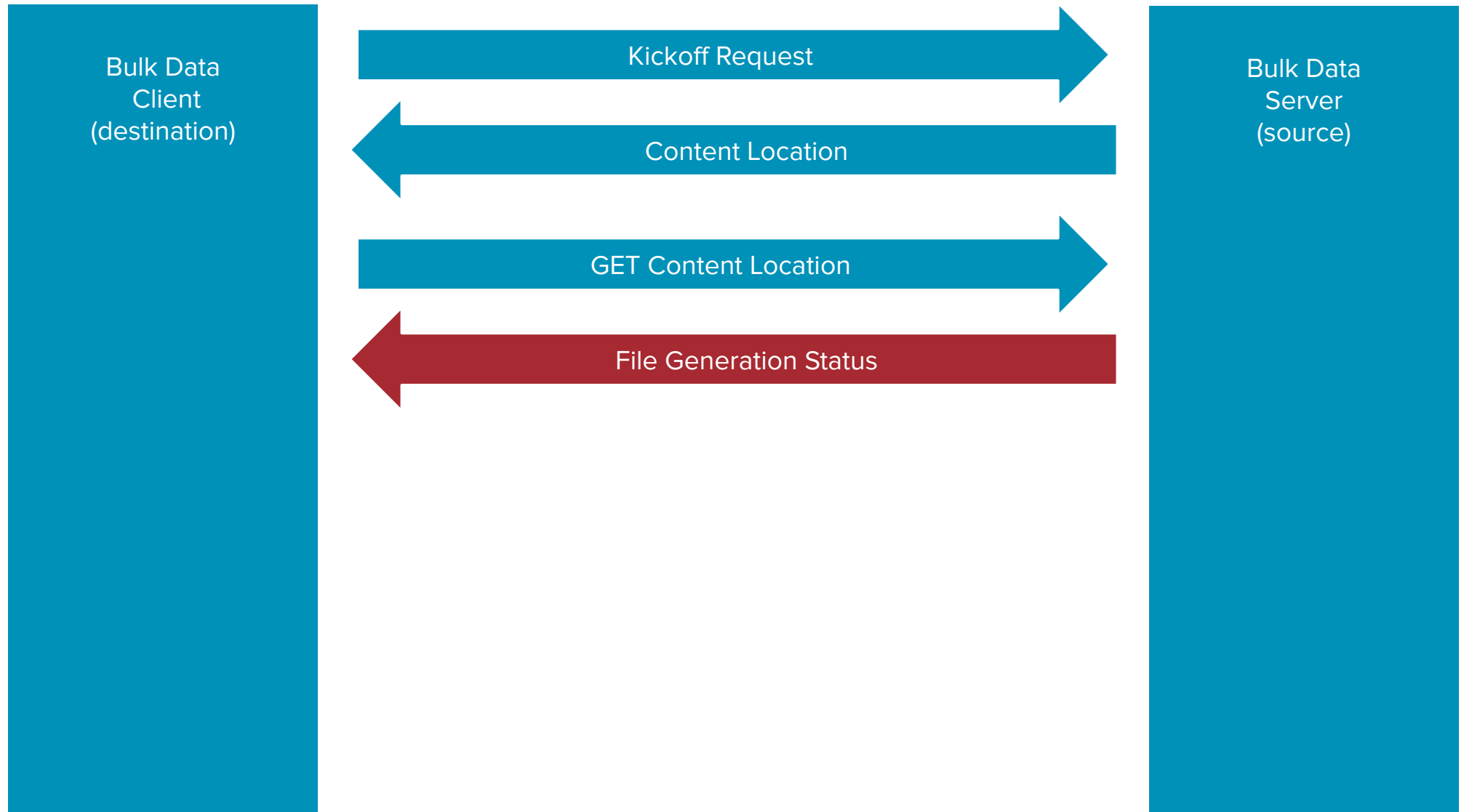
Status: 202 Accepted

Content-Location: [URL for status or deleting request]

Status Request #1



Status Response #1



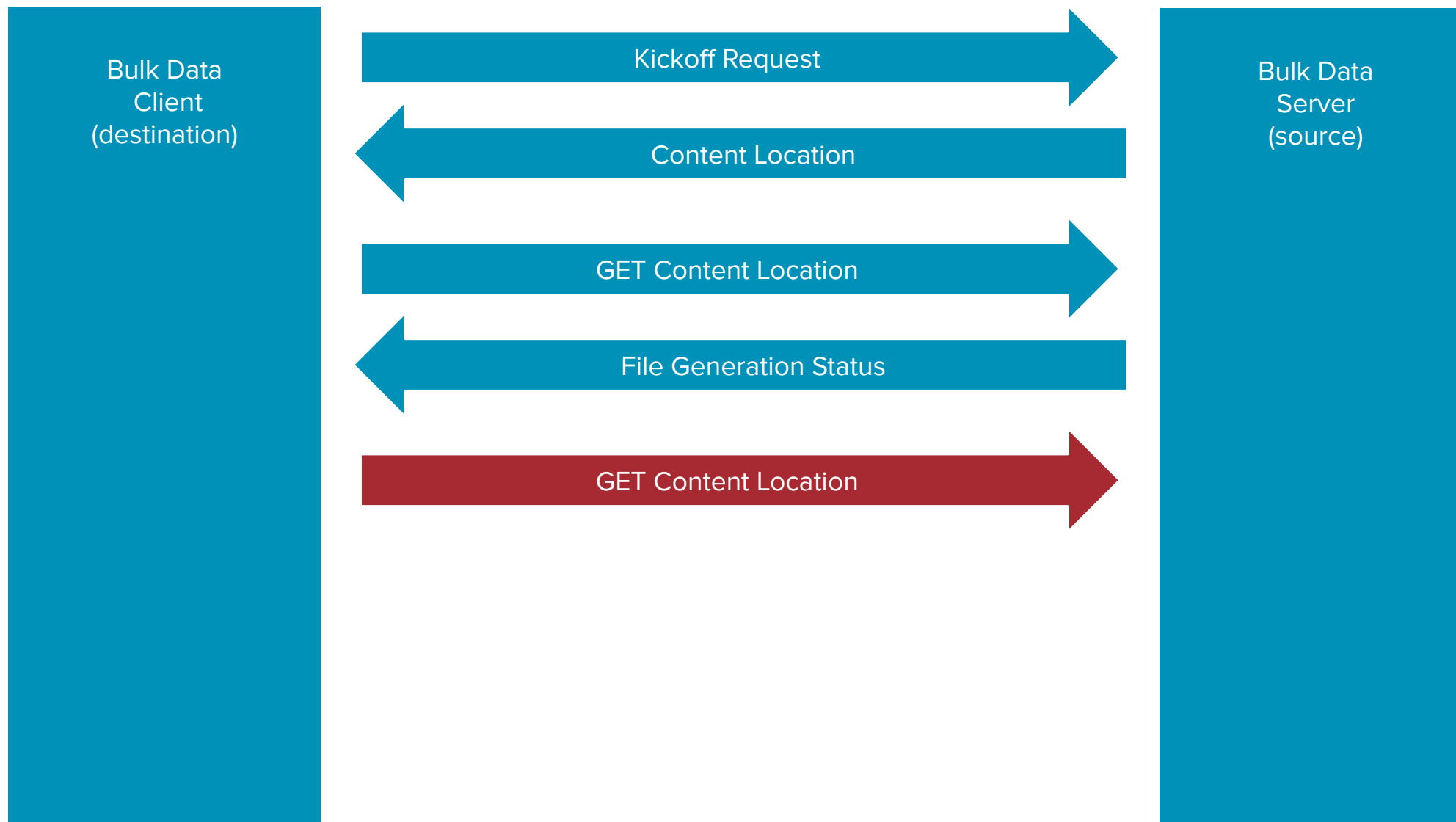
In-Progress Status Response Header

Status: 202 Accepted

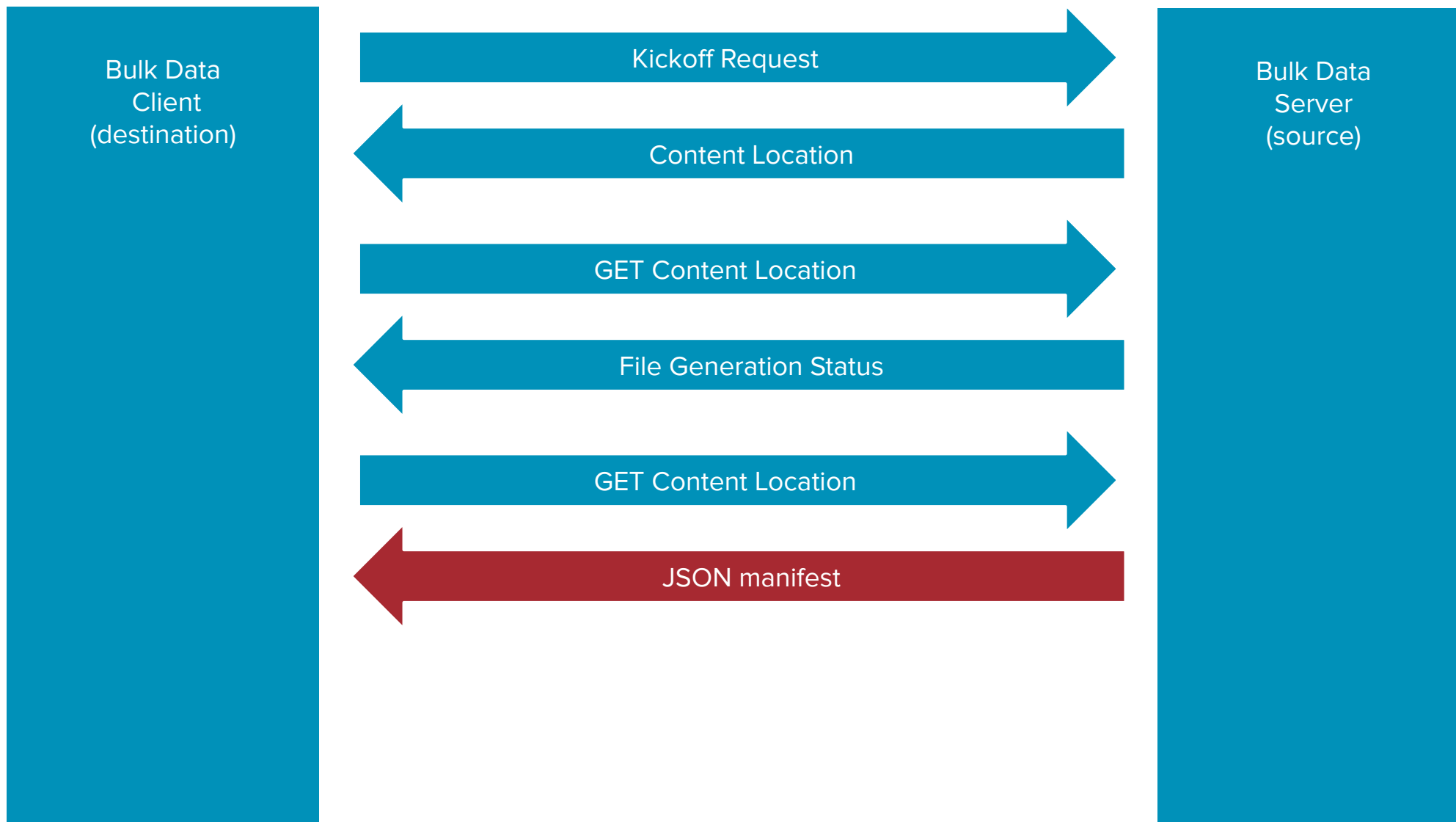
X-Progress: "50% complete"

Retry-After: 120

Status Request #2



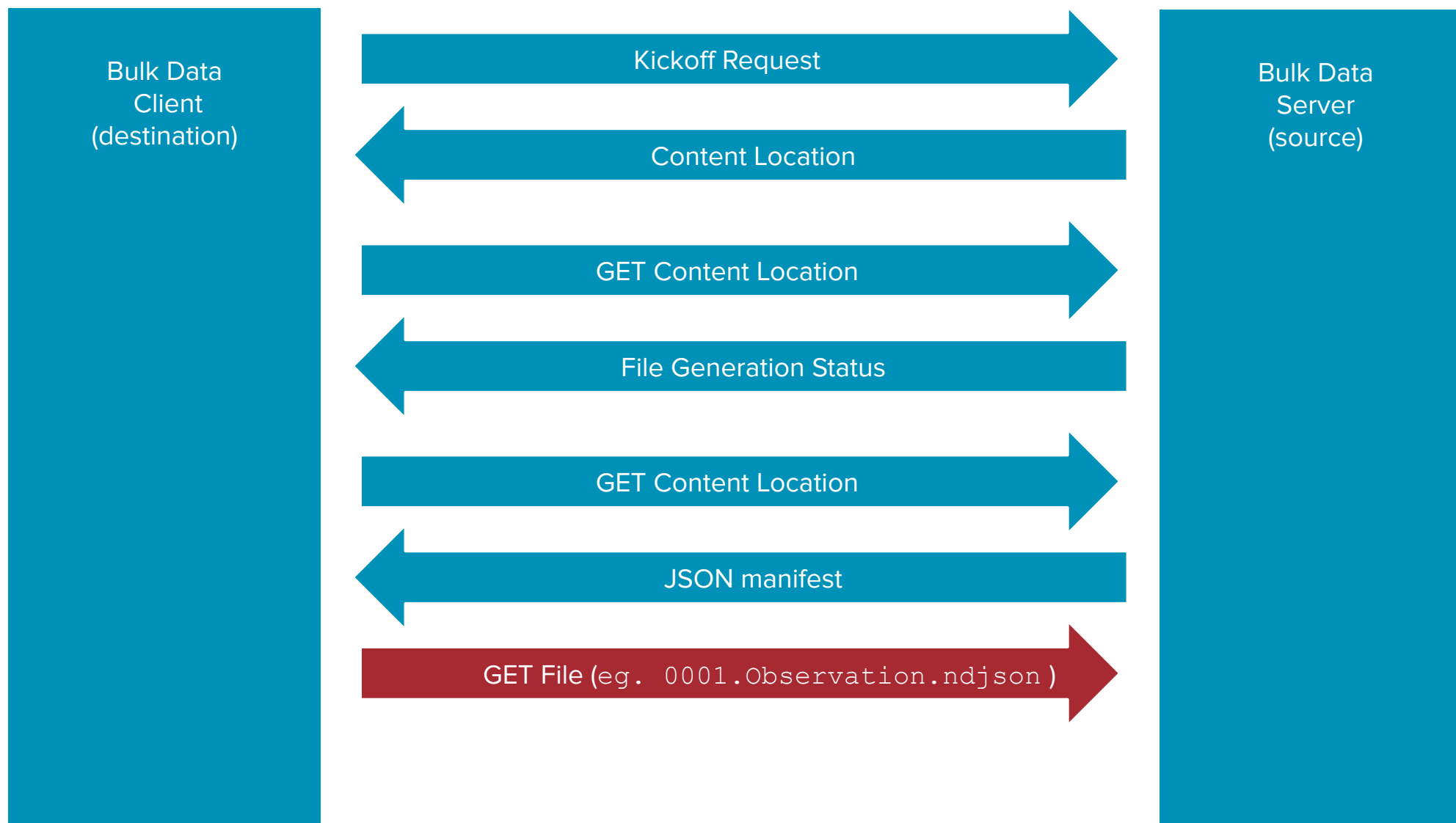
Status Response #2



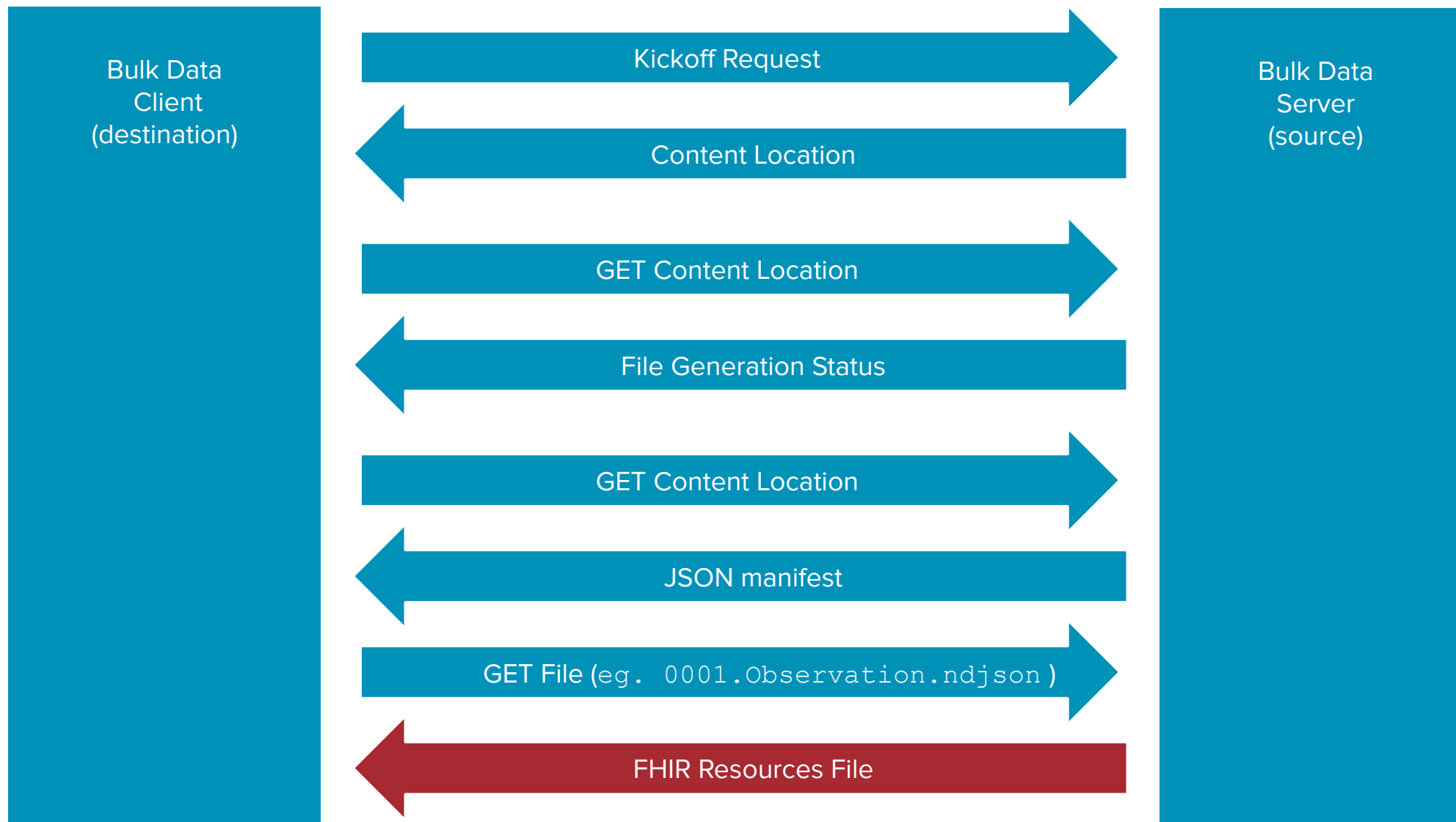
Status Complete Response Body

```
{
1  "transactionTime" : "[instant]",
2  "request" : "[base]/Patient/$export?_type=Patient,Observation",
3  "requiresAccessToken" : true,
4  "output" : [{
      "type" : "Patient",
      "url" : "http://serverpath2/patient_file_1.ndjson"
    },{
      "type" : "Patient",
      "url" : "http://serverpath2/patient_file_2.ndjson"
    },{
      "type" : "Observation",
      "url" : "http://serverpath2/observation_file_1.ndjson"
    }
  ],
5  "error" : [{
      "type" : "OperationOutcome",
      "url" : "http://serverpath2/error_file_1.ndjson"
    }
  ]
}
```

File Request



File Response



FHIR Resources

Data models representing discrete clinical and administrative units (patient, practitioner, allergy, medication order, etc.)

- Currently around 100 have been defined
- Can reference other resources by their URL
- Don't include the kitchen sink, but support extensions
 - “We only include data elements if we are confident that most normal implementations using that resource will make use of the element”
 - Grahame Grieve (FHIR Product Director)
- MU3 Common Clinical Dataset (and soon USCDI) defines subset

NDJSON

[

```

{"id": "06eb35fc-09e6-48 ... "given": ["Lucille"],"family": "Bluth"}} ,
{"id": "cf53f382-6eb6-4f ... "given": ["George", "Oscar"],"family": "Bluth","suffix": ["Senior"]}] } ,
{"id": "406a9c3e-50f9-4c ... "given": ["Michael"],"family": "Bluth"]}

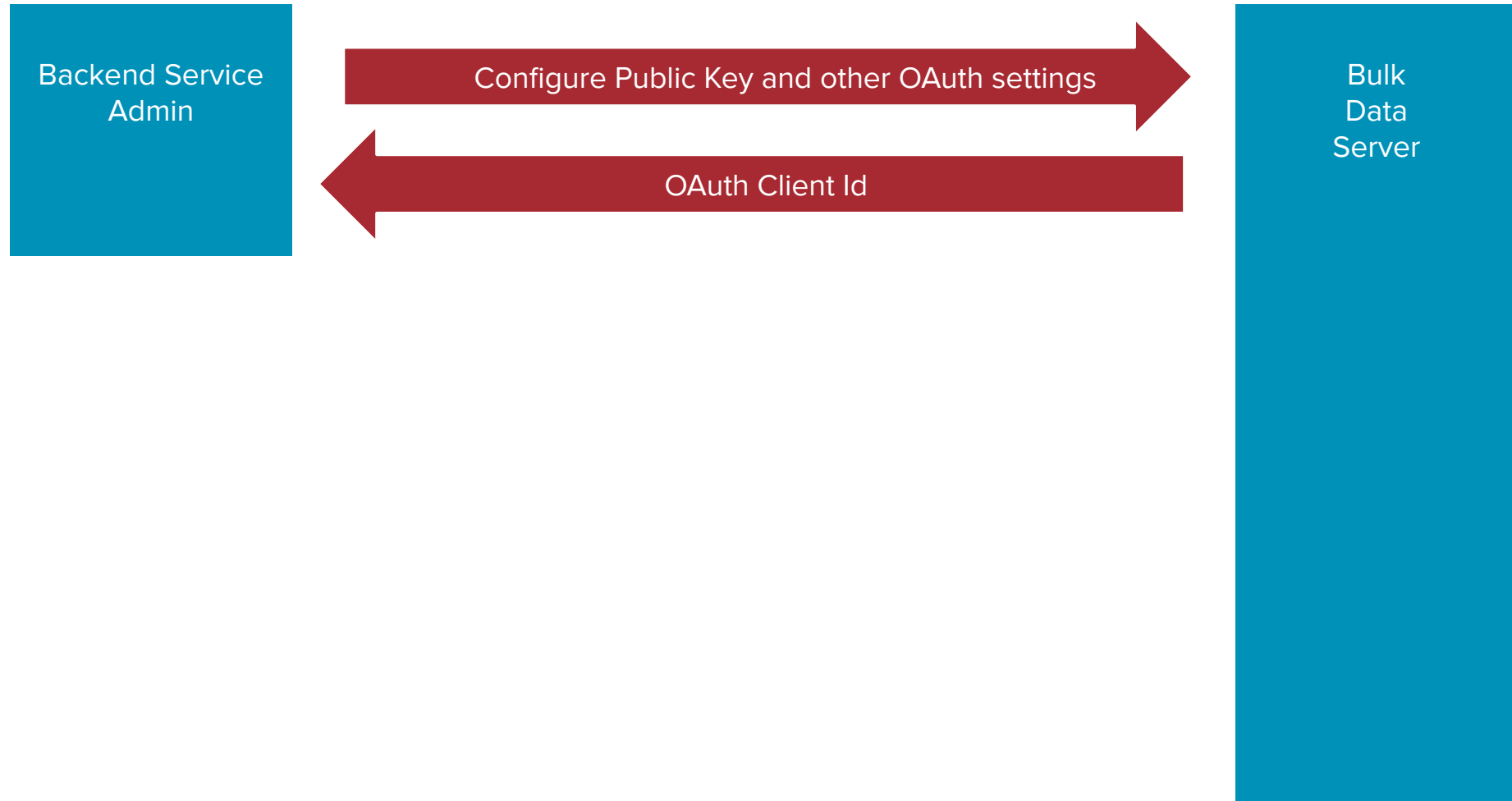
```

]

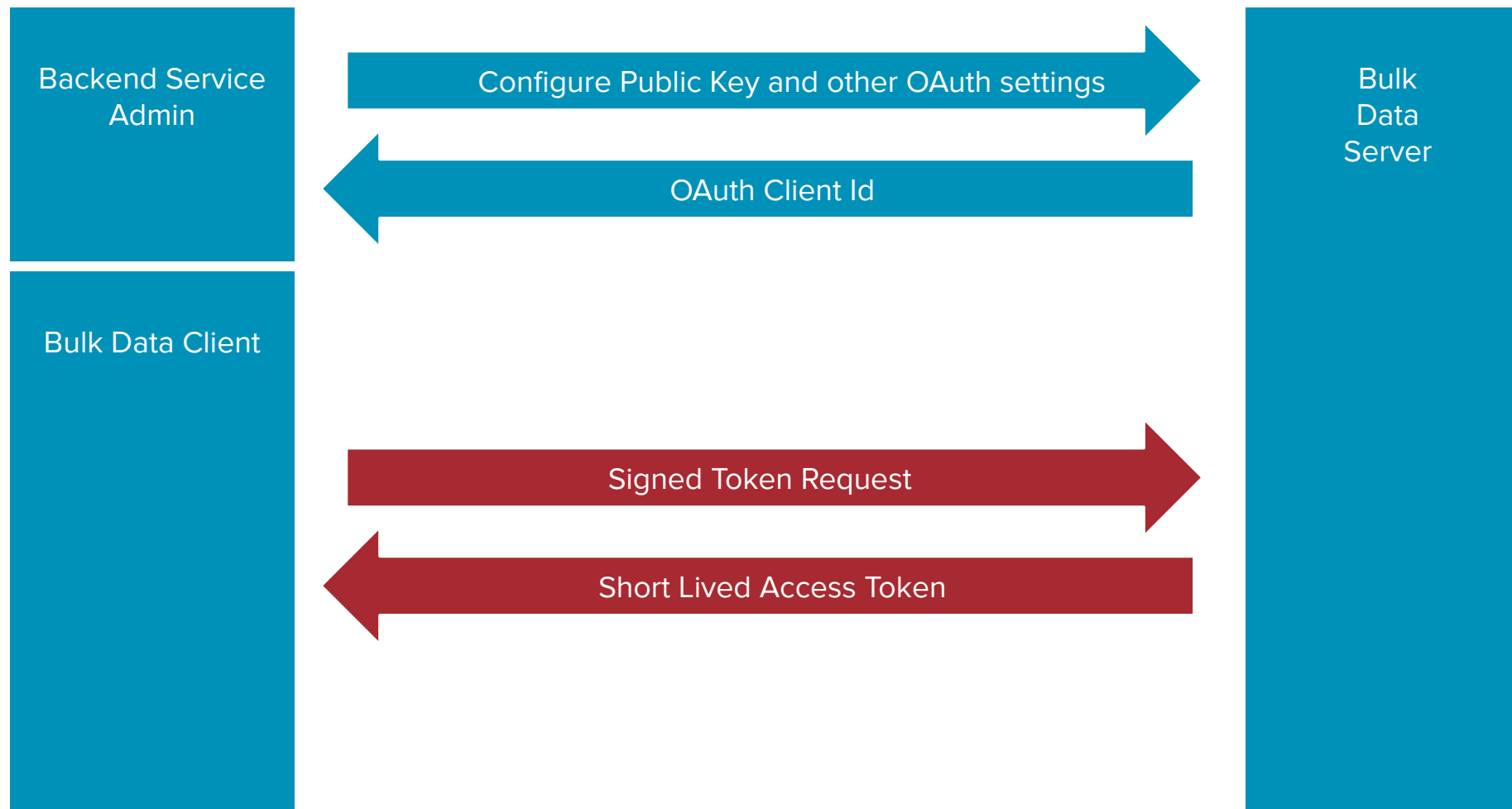
SMART Backend Services Authorization

- Out-of-band app registration (can use Dynamic Client Registration or portal)
- Apps can register public key (JWKS format) or URL for public key
- Token requests signed with private key
- System level scope (parallels SMART “user” and “patient” scopes)
`system/ [resourceType] . read`
- Short-lived access tokens

Registration Flow (once)



Authorization Flow (min. once per request)



Open Source Tools

SMART Reference Server Implementation

<https://bulk-data.smarthealthit.org>

SMART Bulk Data Server alpha
Save

Launch Options

Public Key Generate Keys

base64-encoded RSA-256 public key

Paste your RSA-256 public key or use the button above to generate a new one

Service URL

Service URL

Resources per File

10,000
▾

Advanced

Access Token Lifetime

15 minutes
▾

Database Size

100 Patients
▾

Simulate Error for Testing

None
▾

Simulated file generation duration

10 Seconds
▾

Launch Configuration Download as JSON

FHIR Server URL Try Sample App

https://bulk-data.smarthealthit.org/eyJlcnliOlllLCJwYVWdljoxMDAwMCwiZHVyIjI

Client ID

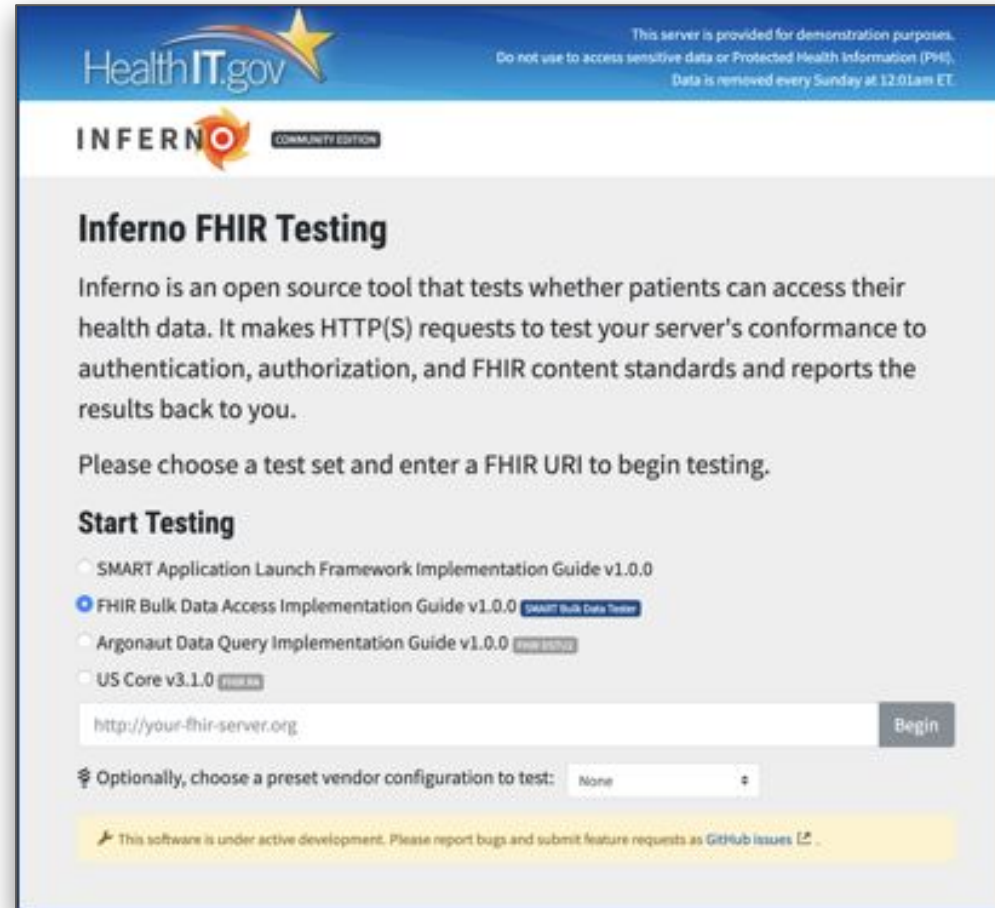
Authentication URL

https://bulk-data.smarthealthit.org/auth/token

[Bulk Data Docs](#) |
 [Backend Services Docs](#) |
 [View NodeJS Sample App](#)

ONC Inferno Testing Tool

<https://inferno.healthit.gov/community>



The screenshot shows the web interface for the Inferno FHIR Testing tool. At the top, there is a blue header with the HealthIT.gov logo and a disclaimer: "This server is provided for demonstration purposes. Do not use to access sensitive data or Protected Health Information (PHI). Data is removed every Sunday at 12:01am ET." Below the header is the "INFERNO COMMUNITY EDITION" logo. The main content area is titled "Inferno FHIR Testing" and contains the following text: "Inferno is an open source tool that tests whether patients can access their health data. It makes HTTP(S) requests to test your server's conformance to authentication, authorization, and FHIR content standards and reports the results back to you." Below this, it says "Please choose a test set and enter a FHIR URI to begin testing." Under the heading "Start Testing", there are four radio button options: "SMART Application Launch Framework Implementation Guide v1.0.0", "FHIR Bulk Data Access Implementation Guide v1.0.0" (which is selected and has a "SMART Bulk Data Tester" tag), "Argonaut Data Query Implementation Guide v1.0.0" (with a "FHIR BULK" tag), and "US Core v3.1.0" (with a "US CORE" tag). Below the options is a text input field containing "http://your-fhir-server.org" and a "Begin" button. At the bottom, there is a dropdown menu for "Optionally, choose a preset vendor configuration to test:" with "None" selected. A yellow banner at the very bottom states: "This software is under active development. Please report bugs and submit feature requests as GitHub issues." with a link icon.

Adoption

Growing number of implementations!

FHIR Servers

- [Microsoft](#)
- [HAPI](#)
- [IBM](#)
- [CareEvolution](#)
- Google (prototype)

Payor Data Servers

- [CMS ACO Beneficiary Claims Data \(pilot\)](#)
- [CMS Data at the Point of Care \(pilot\)](#)
- BCH Payor Analytics (open source prototype)

EHR FHIR Server Prototypes

- Epic
- Cerner
- T-System

US Regulatory Requirements

170.215 [EHR Certification] Application Programming Interface Standards.

The Secretary adopts the following application programming interface (API) standards and associated implementation specifications [...] FHIR Bulk Data Access (Flat FHIR) (v1.0.0: STU 1), including mandatory support for the “group-export” “OperationDefinition”













Implementation Date: 5/1/2022

Argonaut & Next Steps

HL7 Argonaut - Bulk Data IG v1.2

- Open process
 - Meetings schedule at <http://2020.argo.run> with updates on Bulk Data Zulip discussion channel
 - Work-in-progress IG at <https://github.com/hl7/bulk-data>
 - Bulk data track at FHIR Connectathon events for testing
- Active participants include: CareEvolution, Cerner, CMS (Adhoc), Epic, Google, IBM, Meditech, Microsoft, SMART, T-System

Historical Group Data - the challenge

	Jan 1, 2020	Feb 1, 2020
Group 1 Data	 Patient 1  EoB 1-1 lastUpdated 2019-12-31	 Patient 1  EoB 1-1 lastUpdated 2019-12-31  EoB 1-2 lastUpdated 2020-01-15  Patient 2  EoB 2-1 lastUpdated 2019-12-31  EoB 2-2 lastUpdated 2020-01-15
Query	GET Group/1/\$export	GET Group/1/\$export?_since=2020-01-01
Results	 EoB 1-1	 EoB 1-2  Missing: EoB 2-1  EoB 2-2

Historical Group Data - approach

- Revised guidance on the “**_since**” parameter

“In the case of a Group level export, servers MAY return additional resources modified prior to the supplied time if the resources belong to the patient compartment of a patient added to the Group after the supplied time (this behavior should be clearly documented by the server).”

- Addition of optional “**patients**” parameter
 - Limit historical data requests to only new patients

HL7 Argonaut - Bulk Data IG v1.2

- Historical data for new members of a group ✓
- Binary content in attachments (including handling of large files) ✓
- Propagate resource deletions ✓
- Documentation clarifications
 - Parameter optionality for servers ✓
 - Use of POST requests with Parameters resource ✓

HL7 Argonaut - Bulk Data IG v1.2

- Handling Provenance resources
- Export job management
 - Clearer errors
 - Cancellation and export file deletion

HL7 Argonaut - Next Steps

- Consensus language for remaining IG updates
- Prototype and pilot implementations of new functionality
- Update infrastructure to support new IG build process
- Align computable definitions with narrative changes
- Ballot as v1.2

Get Involved!

- Use the APIs in real-world use cases and collect ideas for v2
- Argonaut Project (<http://2020.argo.run>)
- Open source modules (eg. [de-identification](#), filtering, NLP)
- Define Import Operation
Early draft at <https://github.com/smart-on-fhir/bulk-import/blob/master/import.md>
- Standardize analytic approaches (identify cohorts, quality measures)
Draft proposal at <https://github.com/FHIR/sql-on-fhir/blob/master/sql-on-fhir.md>

slides: <https://bit.ly/dd-2020-bulk>

Resources

- Bulk Data Implementation Guide
<https://hl7.org/fhir/uv/bulkdata/>
- SMART Server Reference Implementation
<https://bulk-data.smarthealthit.org>
- Bulk Data Discussion Group (Bulk Data Stream on FHIR Zulip Chat)
<https://chat.fhir.org/#narrow/stream/bulk.20data>

ORGANIZED BY



PARTNER



HOST SPONSOR

