

Contextual Data ClearingHouse Datastore and API specification

| | |
|--|----|
| Change History | 2 |
| Introduction | 4 |
| Authentication | 5 |
| Base URL | 6 |
| Write endpoints | 6 |
| Read endpoints | 7 |
| Use cases and required attribute and value fields | 7 |
| Status codes | 8 |
| Example Commands | 8 |
| Create Curation Data | 8 |
| Update Curation Status (Only by archive users) | 14 |
| Suppress Curation Status (Only by the curation data owner) | 15 |
| Retrieve all curation data | 16 |
| Retrieve curation by Curation ID | 19 |
| Retrieve curation by Record ID | 21 |
| Duplication Handling | 24 |
| Error Response messages | 24 |
| Database fields | 25 |
| Curation record | 25 |
| Assertion Evidence | 27 |
| Archive Processing | 28 |
| User | 29 |
| Appendix - A | 30 |
| Table -1 RecordId Validation Patterns | 30 |

Change History

| Document Version | Date | Author | Notes |
|------------------|------------|---|--|
| 1.0.0 | 2019-02-28 | Vishnukumar Balavenkataraman Kadhirvelu | Initial version. |
| 1.0.1 | 2019-10-30 | Vishnukumar Balavenkataraman Kadhirvelu | <ul style="list-style-type: none"> • Curation bulk insertion • Retrieve curation by attribute pre or post name • Multiple Assertion Evidence support with Identifier • Upgraded to latest AAP services |
| 1.0.2 | 2019-11-15 | Vishnukumar Balavenkataraman Kadhirvelu | <ul style="list-style-type: none"> • Validation implemented to avoid duplicate curation |
| 1.0.3 | 2020-10-05 | Vishnukumar Balavenkataraman Kadhirvelu | <ul style="list-style-type: none"> • Removed datasubs email support |
| 1.0.4 | 2020-10-28 | Vishnukumar Balavenkataraman Kadhirvelu | <ul style="list-style-type: none"> • Added data type for covid19 • Skip OLS validation for covid19 data type |
| 1.0.5 | 2021-02-03 | Vishnukumar Balavenkataraman Kadhirvelu | <ul style="list-style-type: none"> • Added filters provider name, record type, start and end time |
| 1.0.6 | 2021-11-30 | Vishnukumar Balavenkataraman Kadhirvelu | <ul style="list-style-type: none"> • Made the curations post response ordered and consistent |
| 1.0.7 | 2022-01-27 | Vishnukumar Balavenkataraman Kadhirvelu | <ul style="list-style-type: none"> • Changed application log format |
| 1.0.8 | 2022-02-14 | Vikas Gupta | <ul style="list-style-type: none"> • Added record Id validation for adding and reading the Curations. • Introduced 'study' as |

| | | | |
|--------|------------|---|---|
| | | | a new recordType. |
| 1.0.9 | 2022-05-26 | Vishnukumar Balavenkataraman Kadhirvelu | <ul style="list-style-type: none"> Removed totalRecords field from the readAPI |
| 1.0.10 | 2022-06-28 | Vishnukumar Balavenkataraman Kadhirvelu | <ul style="list-style-type: none"> Update Life science login in the document |

Introduction

This document describes the datastore and Application Programming Interface (API) specification for the Contextual Data ClearingHouse (CDCH), the purpose of which is to enable extension, correction and improvement of publicly available annotations on sample and sequence records available in Elixir archive data resources. The CDCH data store and APIs aim to provide a seamless method of exchange for curated contextual data available in external resources and community curation efforts, with ELIXIR data resources.

Figure 1 outlines the flow of contextual data into the CDCH data store. The data store is a MongoDB database accessed through read and write APIs managed by a load balanced VM architecture. The data store captures contextual data, associated metadata and provenance information, with incoming blocks of curation data separated into individual rows of curated contextual data attribute pairs.

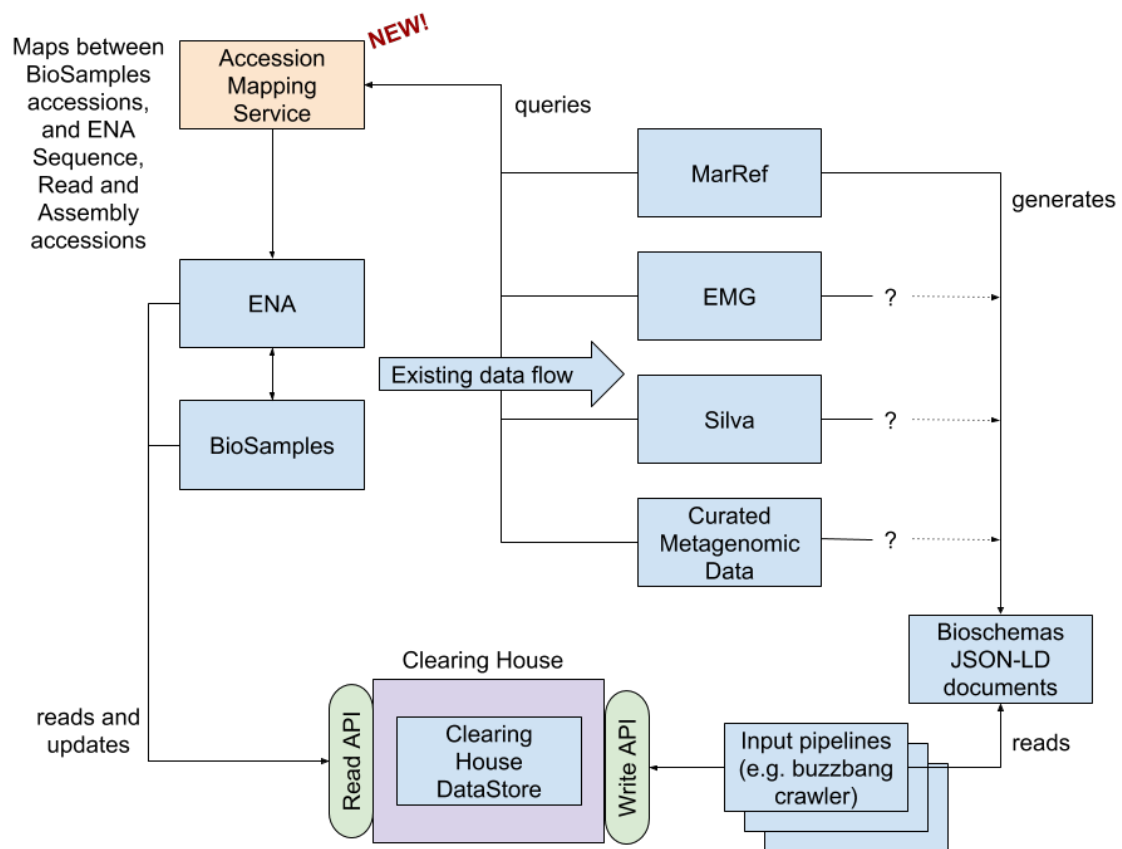


Figure 1. Contextual data flow surrounding Clearing House Data Store.

Authentication

The CDCH is an authenticated service in order to ensure contextual data provenance and authenticity.

Users/services wanting to supply contextual data to the CDCH need to first register for a Life Science ID here: <https://signup.aai.lifescience-ri.eu/fed/registrar/?vo=lifescience>

Alternatively, users can directly create the AAP account here:
<https://aai.ebi.ac.uk/registerUser>

Once you have an account you will need to obtain an authentication token in order to interact with the CDCH APIs. To authenticate you first need to obtain a bearer authentication token, this is obtained by using the link <https://api.aai.ebi.ac.uk/sso> in the browser and supplying the credentials from when you registered for an account.

Alternatively, you can obtain a bearer authentication token by supplying your AAP username and password programmatically, for example:

```
curl -u <your_aap_username>:<your_aap_password> https://api.aai.ebi.ac.uk/auth
```

The response will provide a bearer token, for example:

```
eyJhbGciOiJIc3EiLCJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJ1b2N0b3R5b250b211LTQzOTgtOGRjOS02ZmU4MWVkbWZlbnRvbmFtZSI6Imthcm9seSBFcmRvcyIsImRvbWFpbniMiOiJlYXVzZXJzLWRvbWFpbjdfQmNleqmpdxGM y7ln3mlMFqhdNVDrf0uvilbe2pqngper56sjUYUOMnzcCK6ht2Q65P25VdJAA_nSOQJ03ZEqp2ZB0urGpCY_sK0JV834 ZSe1atLKww7wa2Ntq5EJakUfxrma9IHLTylGxJzYc3ej1s8pTE5Jj6tprvcV5s3eNEmYFUJL9RwhKmDA9- 5E7TfOVYvEQSMw7lqXMiVSSQNU1o_dfalp- PYlqer9mTR2MPIMAAa2wdr7jKVDXzI7zKlprIRLVyJAiEPIeWzNYN8FC2RdigwqbcBt9UTTDrdwDH5Op3fAJPrYWNp RD4-wV8mn5DWzd7cNm6dMqwm1C12WEg
```

NOTE: The token will be valid for 1 hour, after which you will need to request another token.

You need to supply this token as an Authorisation header in all of your write API requests. The format is “Authorization: Bearer <TOKEN>”. There are examples of its usage in the examples section below.

The write API can be utilised to supply contextual data by any user with a valid Authentication, Authorisation and Profile service (AAP) account. Further information on AAP accounts can be found here <https://aai.ebi.ac.uk/home>. The EMBL-EBI AAP service is fully compatible with Life Science Authentication and Authorization Infrastructure (AAI; <https://lifescience-ri.eu/ls-login.html>) and accepts Life Science credentials for authentication.

The authenticated AAP account ID will be recorded in the CDCH datastore against your supplied data to verify authenticity and provenance of the provided information.

The read API does not require authentication.

The archive processing status write endpoint will be restricted to specific AAP accounts associated with each Elixir Archive or external data resource that is consuming the contextual data. To register for the ability to record processing status for your archive/resource in the CDCH please contact our helpdesk using <https://www.ebi.ac.uk/ena/browser/support> with the

subject “Contextual Data ClearingHouse archive authenticated access” which will add you to the AAP domain that authorises this functionality.

A test version of the CDCH is also available to enable users to develop against the available APIs and setup their workflows. The test version utilises a different authentication service. For the test version, users will need to register for a separate Explore AAP account at <https://explore.aai.ebi.ac.uk/registerUser>

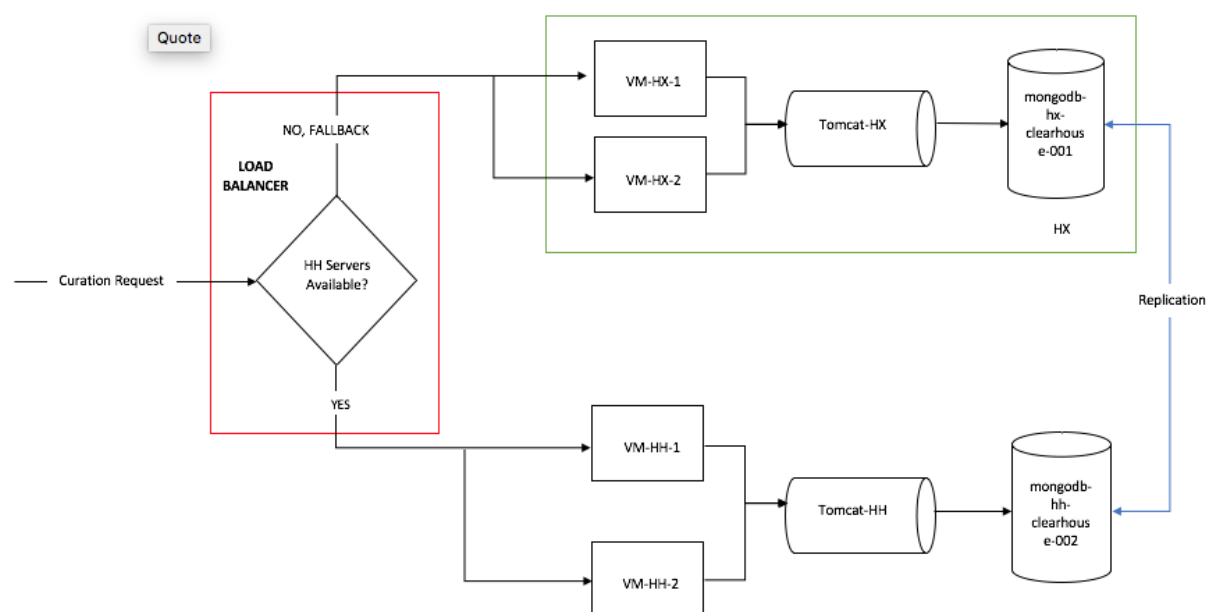
For the test version of the CDCH the bearer token is obtained from the explore API:
 curl -u <your_aap_username>:<your_aap_password> https://explore.api.aai.ebi.ac.uk/auth

Base URL

Production API URL: <https://www.ebi.ac.uk/ena/clearinghouse/api>

Test API URL : <https://wwwdev.ebi.ac.uk/ena/clearinghouse/api>
 (Note: For developer use/feature testing only. This instance is backed by a development database and data stored here may be removed without notice.)

The load balancer will process incoming API request routing according to this server VM design (approved by EMBL-EBI web production) the schema for this architecture is available here:



Write endpoints

Base URL: <https://www.ebi.ac.uk/ena/clearinghouse/api>

List of API endpoints for the following URI: base_url / endpoint ? filter

| Endpoint | Verb | Action | Filters | Success Status-code |
|--------------------------|-------|---|---------|---------------------|
| /curations | POST | Creates new curation record(s) | | 201 Created |
| /curations/{id} | PATCH | Changes the status of the curation record | | 200 OK |
| /curations/{id}/suppress | PATCH | Suppresses the curation record | | 200 OK |

Read endpoints

Base URL: <https://www.ebi.ac.uk/ena/clearinghouse/api>

List of API endpoints for the following URI: base_url / endpoint ? filter

| Endpoint | Verb | Action | Filters | Success Status-code |
|-----------------|------|---|---|---------------------|
| /curations | GET | Retrieve the curation records | offset limit dataTypeId entifier endTime startTime providerName recordType | 200 OK |
| /curations/{id} | GET | Retrieve the curation by Curation ID or Record ID and attribute names (pre or post) | attributeName | 200 OK |

Use cases and required attribute and value fields

To add a new attribute and value pair:

Required: attribute_post & value_post

To modify the name of an existing attribute:

Required: attribute_pre, attribute_post, value_pre & value_post (same as value_pre)

To replace an attribute/value pair with a new attribute/value pair:

Required: attribute_pre, attribute_post, value_pre & value_post

To remove an attribute:

Required: attribute_pre & attribute_delete (true)

To suppress a previously supplied curation

Required: UUID. Use write endpoint /curations/{uuid}/suppress

Status codes

General convention for status codes.

GET - Read:

- 200 (OK)
- 401 (Unauthorised for this action, please authenticate)
- 404 (Record Not Found, if ID not found or invalid).
- 400 (Bad Request, if ID is failed in pattern matching)

POST - Create:

- 201 (CREATED, if all the curations of the batch are successful)
- 207 (MULTI-STATUS, if the curation of the batch are partially successful)
- 401 (Unauthorised for this action, please authenticate)
- 403 (Forbidden, when non-archive users attempt to edit archive information)
- 404 (Record Not Found, if ID not found or invalid)
- 400 (Bad request, will include error response from error response messages below, including malformed JSON or missing mandatory information).

Example Commands

Key

| |
|--------------------------|
| Example CURL request |
| Example success response |
| Example failure response |

Create Curation Data

| Field | Example | Type | Requirement | Notes |
|------------|---|--|-------------|---|
| recordType | 'sample'/'sequence'/'study' | String. controlled list (sample sequence study) | Mandatory | Target type to determine which archives will be interested. API will throw error if not value from controlled list. |
| recordId | SAMEXXXX XXX/ERRXX XXX/PRJXX XXXXX | free text string (validated – Please Refer to Table-1 in Appendix-A) | Mandatory | archive record identifier. Initially will be free text string with no validation. Future versions of CDCH may utilise dependent archive accession |

| | | | | |
|-----------------|---------------------------|-----------------------------------|--|---|
| | | | | validation code/APIs. |
| dataType | covid19 | String. controlled list (covid19) | Optional | Target type to determine the data type of the record. API will throw error if not value from controlled list. |
| dataIdentifier | COVIDI26 | free text string | Optional (Mandatory only if datatype is mentioned) | Data identifier for the given dataType. |
| attributePre | sex/location/breed | free text string | Optional (at least one of attribute_pre or attribute_post is required) | Preexisting attribute to correct. See use cases section for appropriate usage. |
| valuePre | male/UK/Rambouillet | free text string | Optional | Pre-existing value to curate. See use cases section for appropriate usage. |
| attributePost | sex/location/breed | free text string | Optional (at least one of attribute_pre or attribute_post is required) | Replace current attribute with this one. See use cases section for appropriate usage. |
| valuePost | male/UK/Rambouillet | free text string | Optional | Replace current value with this one. See use cases section for appropriate usage. |
| attributeDelete | true | boolean | Optional | Flag for deleting the attribute listed in attribute_pre. Must not specify attribute_post or value_post |
| assertionMethod | 'automatic assertion'/'ma | Valid text or synonym that is | Mandatory | Whether curation was determined |

| | | | | |
|-------------------------|---|--|---|---|
| | nual assertion' | ontology child of ECO_0000217 assertion method | | manually or automatically taken from Evidence & Conclusion Ontology (ECO) |
| assertionEvidences | AssertionEvidence | List. See below. | Mandatory | Supporting evidence for the assertion taken from Evidence & Conclusion Ontology (ECO) |
| assertionSource | | Valid URI | Optional (could make mandatory for some values of assertion_evidence) | Valid URI link either to the supporting publication or external reference database that supports the stated assertion |
| assertionAdditionalInfo | | free text | Optional | Additional free text info to support assertion |
| providerName | MarRef/EMG/Silva | free text | Mandatory | name of provider |
| providerUrl | https://mmp.sfb.uit.no/databases/marref/ | Valid URI | Optional | URL of the provider |

Creation

```
curl -X POST "https://www.ebi.ac.uk/ena/clearinghouse/api/curations" -H "accept: */*" -H "Authorization: Bearer eyJhbGciOiJSUzI1NiJ9.eyJpc3MiOiJodHRwczovL2V4cGxvcmUuYWVpLmViaS5hYy51ay9zcCIzImp0aSI6IIBBMFNMVWxoaHNzZVhzVHRabWF0OEElLCJpYXQiOiE1NDg3NzQ3ODksInN1YiI6InVzci05OGY5NDlkMi1hNTY1LTQzOWQtYWVwMy01MjI3NDNIN2VIZjkiLCJlbWFpbCI6InZpc2hidGVjaEBnbWFpbC5jb20iLCJuaWNrbmFtZSI6InZpc2hidGVjaEBnbWFpbC5jb20iLCJ1Y291IjoiaWZlZG51a3VtYXN1IiLCJkb21haW5zIjpbXSswiZXhwIjoxNTQ4Nzc4Mzg5fQ.ONF2xH42n3_5EiILfPUMAJmuu-0gehAyWQF1Z8lgatA025yM6xAORGb2QK9ZZ61aleESiwwk0tt684fVrNm_Ju1y-F2-9YjS3DxPfmMYKNFk4Y4RPnVUUHjYAFV9v2eR_4wWxVGx6Y1o9vzCkzSgQgkYlvTQ3vn0XBINC5tNEB-1VlzaeFWuy1SicP5dBz2hIOPf0uhd20MbFtwsHuG-evlkQzsl_yZsR396qe7LoUFKVF8crk2jDi6gIjoDJw8-6QwNREN54mypoDBMjOPEKxRu8E0nDh4mVmqPleiayluzD0iMB5uiIk0BHIPPfk1ZXkLrVmKTQOhynO9W83kwa" -H "Content-Type: application/json" -d "{ \"curations\": [ { \"assertionAdditionalInfo\": \"Curator spotted error during project review\", \"assertionEvidences\": [{ \"label\": \"inference based on individual clinical experience used in manual assertion\" }], \"assertionSource\": \"https://mmp.sfb.uit.no/databases/marref/#/records/MMP02469391\", \"attributePost\": \"geographic location\", \"attributePre\": \"isolation source\", \"providerName\": \"MarRef\", \"providerUrl\": \"https://mmp.sfb.uit.no/databases/marref/\" } ] }
```

```
\\"https://mmp.sfb.uit.no/databases/marref/#/records/MMP02469391\\", \\"recordId\\": \\"SAMN02469391\\", \\"recordType\\": \\"sample\\", \\"attributeDelete\\": \\"true\\"}]"
```

Status: 201 CREATED

```
{
  "timestamp": "2019-01-29T16:18:05.423+0000",
  "curationIds": [
    "d7b6eed1-945e-48be-b7ec-66375813ef31"
  ],
  "errors": []
}
```

Status: 400 BAD REQUEST

```
{
  "timestamp": "2019-01-29T16:21:31.586+0000",
  "curationIds": [
    "ERR-10b60c4a-9469-443f-9179-86bc075078f7"
  ],
  "errors": [{
    "errorId": "ERR-10b60c4a-9469-443f-9179-86bc075078f7",
    "message": "recordType must not be null for the curation record: 1"
  }]
}
```

Status: 400 BAD REQUEST

```
{
  "timestamp": "2021-11-23T13:05:35.098+0000",
  "curationIds": ["ERR-cafe3d70-2498-4103-bde4-f40f94daaaa9"],
  "errors": [{
    "errorId": "ERR-cafe3d70-2498-4103-bde4-f40f94daaaa9",
    "recordId": "ERR00001",
    "message": "System Error"
  }]
}
```

Status: 400 BAD REQUEST

```
{
  "timestamp": "2022-02-14T11:04:06.153+0000",
  "curationIds": [
    "ERR-3bd47252-042c-4a34-abe8-cb428ca0c1a4"],
  "errors": [{
    "errorId": "ERR-3bd47252-042c-4a34-abe8-cb428ca0c1a4",
    "message": "curations[0] For RecordType 'sequence', supplied Record Id 'ERR0001'
format is not valid for the curation record: 1"
  }]
}
```

```
{
  "timestamp": "2019-09-26T10:49:51.882+0000",
  "curationIds": [
    "ERR-tufc3d70-1234-5678-hut3-d98f94dbbbb9",
    "ERR-cafe3d70-2498-4103-bde4-f40f94daaaa9"
  ],
  "errors": [ {
    "errorId": "ERR-tufc3d70-1234-5678-hut3-d98f94dbbbb9",
    "recordId": " SAMN02469391",
```

```

    "message": " System Error "
  }, {
    "errorId": "ERR-cafe3d70-2498-4103-bde4-f40f94daaaa9",
    "recordId": "SAMN02469391",
    "message": "System Error"
  }
}

```

Status: 207 MULTI-STATUS

```

{
  "timestamp": "2019-09-26T10:49:51.882+0000",
  "curationIds": [
    "1d75f552-e9e2-4f27-9e0d-62c0a88d6925",
    "ERR - cafe3d70 - 2498 - 4103 - bde4 - f40f94daaaa9"
  ],
  "errors": [{
    "errorId": "ERR-cafe3d70-2498-4103-bde4-f40f94daaaa9",
    "recordId": " SAMN02469391",
    "message": "System Error"
  }
]
}

```

NOTE: The order of the response has been maintained based on the requested curation order irrespective of success or failure. Failures of the curation are responded with the Error code (E.g., ERR-cafe3d70-2498-4103-bde4-f40f94daaaa9) in the curationIds list and errors list of the json.



Update Curation Status (Only by archive users)

| Field | Example | Type | Requirement | Notes |
|--------------------------|---|-----------------|-------------|--|
| Curation ID | 623a1df1- eeeb-483a- 9938- abf24f2f5187 | String | Mandatory | ID of the submitted curation as path variable. It can only be added by authenticated archive |
| archiveName | biosamples | Free text | Mandatory | can only be added by authenticated archive |
| archiveProcessing Status | processed/on_ hold/rejected | Controlled list | Mandatory | can only be added by authenticated archive |
| archiveProcessing Notes | Accepted but formatted to meet archive convention | free text | Optional | can only be added by authenticated archive |

```
curl -X PATCH "https://www.ebi.ac.uk/en/clearinghouse/api/curations/faf1032a-8166-44e1-b197-6d2e58a6dc14" -H "accept: */*" -H "Authorization: Bearer eyJhbGciOiJSUzI1NiJ9.eyJpc3MiOiJodHRwczovL2V4cGxvcmUuYWFPpLmViaS5hYy51ay9zcGlscmp0aS16Ink2cGMlY0hValhxaH15VFpYeHZqdEEiLCJpYXQiOiEjNDg3NzkwMjk5LnN1Yil6InVzci0tMDZlbnZlc0Ni03NTgwLTRiZDYtOWQ0MC0tMDNjNWQ5ZTQzM2UuLiClJlbWFPbWpCImlmthZGhpencZlhbHVhZAZWJpLmFmFjLnVrIiwibmJja25hbWUiOiJ2aXNobnVrdWlhcilslm5hbWUuOiJWZXNobnVrdWlhcilBCYyXhdmdVua2F0YXJhbWFlETHZGhpencZlhbHUuLiClJk21haW5zIjpblnNlbGYuY2xlYXJpbmdob3VzZSIslm5hbWUuOiJWZXNobnVrdWlhcilBmdob3VzZV9hemNoaXZlcjldLjCjLeHAiOiEjNDg3ODI2Mjl9.lyTh_XTsJf2UONMilcoNYZdnc7RpN1RRpv1Lfj-32r5XGV8NBmD4643QH-VckxMXATUOrEq8rpjYdp6Hze2W9g_wEKTOHnw0qiUqH54vZoRrhuz8O1cpzcqbh-HfyUaqWmeqXGZBPm4i2kaH5vongDCXBVTdaV9OUKEDI5iffwtU06E-qPz9wnoHaekabWHFPtH_dN9yllgsIU62P12yV3IP_ZYPdl_DX8CNY0-yXWIHL_e6BV59ejt1KRHBi3HjGHCHjdJhyFDYgHzdfucguPAQAuCMrmqnSP3FmhMeeMfywPZplCqrIxG9lg6EiDSE7aQZAcrf2dQi3JD9ha" -H "Content-Type: application/json" -d '{"archiveName": "biosamples", "archiveProcessingNotes": "processed the curation", "archiveProcessingStatus": "processed"}'
```

```
Status: 200 OK
{
  "id": "faf1032a-8166-44e1-b197-6d2e58a6dc14",
  "message": "Curation Status added",
  "success": true,
  "timestamp": "2019-01-29T16:24:21.772+0000"
}
```

```
Status: 404 NOT FOUND
{
  "id": "faf1032a-8166-44e1-b197-6d2e58a6dc141",
  "message": "Record not found",
}
```

```
"success": false,  
"timestamp": "2019-01-29T16:25:30.743+0000"  
}
```

Suppress Curation Status (Only by the curation data owner)

| Field | Example | Type | Requirement | Notes |
|-------------|--|--------|-------------|---|
| Curation ID | 623a1df1- eeeb-483a- 9938- abf24f2f5187 | String | Mandatory | ID of the submitted curation to be suppressed as path variable |

```
curl -X PATCH "https://www.ebi.ac.uk/ena/clearinghouse/api/curations/623a1df1-eeeb-483a-9938-abf24f2f5187/suppress" -H "accept: */*" -H "Authorization: Bearer eyJhbGciOiJSUzI1NiJ9.eyJpc3MiOiJodHRwczovL2V4cGxvcmUuYWFpLmViaS5hYy5lay9zcCIsImp0aSI6InVENkFGM3YtaUICUWF0b3VoYlhh4UVEiLCJpYXQoIjEINDg3Nzg1OTESlnN1Yil6InVzci05OGY5NDlkMi1hNTY1LTQzOWQtYVWVmMy01Mj13NDNIN2VlZjkiLCJlbWVubGl6InZpe2hidGJvaEBnbWFBpC5jb20iLCJuYVWI1IjoimlmlzaGV1a3VtYXhlLCJkb21haW51bjpbXSwhIjoxNTZlNGzyMTkxfQ.SQYdSALTcxXR2vNEECSDioF6v5eyMUBCbl_g1MaSbmLarIaIDbLEtZ6rJB9IPZL9O6xW8bg51Z2-EDyGeLDLF0ohHtEMCOIWsa4Y6jzzPAaFastP5Bg-FahX8xb5if3SYtGTs0uzJohUcW7Tt9PSreDdV5VaQ8nKlrjtALvmhH-SQBj9s6ekbz8le3LGMsYAklwpFrR316Ppb_L_-2m-gs0imY553Xv3bm8YX-dQCXIIdjYx0eui2c2C9Y-7gqQIpNQas7Bq6WAShrDPooGwO3yPBgbtb8KFdcXgIJ5se7FYt_WF0cKQk72vDeflM7ob5R3keN2IX7yyXP1RPA"
```

```
Status: 200 OK
{
  "id": "623a1df1-eeeb-483a-9938-abf24f2f5187",
  "message": "Suppressed the data",
  "success": true,
  "timestamp": "2019-01-29T16:28:08.763+0000"
}
```

```
Status: 404 NOT FOUND
{
  "id": "623a1df1-eeeb-483a-9938-abf24f2f51871",
  "message": "Record not found",
  "success": false,
  "timestamp": "2019-01-29T16:28:45.335+0000"
}
```

Retrieve all curation data

| Field | Example | Type | Requirement | Notes |
|--------------------|------------------------------|---|-------------|--|
| offset | 0 | number | Optional | Number of records to be skipped |
| limit | 10 | number | Optional | Number of records to be limited |
| dataTypeIdentifier | Covid19/COVID26 | string | Optional | Number of records to be fetched based on data type or data identifier |
| providerName | MarRef | string | Optional | Number of records to be fetched based on provider name |
| recordType | sequence | string. controlled list (sample sequence) | Optional | Number of records to be fetched based on the record type |
| startTime | 2019-01-29T14:20:44.109+0000 | ISO Date | Optional | Number of records to be fetched based on the start time |
| endTime | 2019-01-29T14:20:44.109+0000 | ISO Date | Optional | Number of records to be fetched based on the end time. If the startTime has not given, then the endTime becomes redundant. |

```
curl -X GET "https://www.ebi.ac.uk/ena/clearinghouse/api/curations?limit=2&offset=0" -H "accept: */*"
```

```
Status: 200 OK
{
  "curations": [{
```



```

      "id": "faf1032a-8166-44e1-b197-6d2e58a6dc14",
      "recordType": "sample",
      "recordId": "SAMN02469391",
      "attributePre": "geographic location",
      "valuePre": "Norway: Trondheim fjord",
      "attributePost": "geographic location",
      "valuePost": "Norway: Lysefjord fjord",
      "attributeDelete": false,
      "assertionMethod": "manual assertion",
      "assertionEvidences": [{"identifier": "ECO:0006019",
"shortForm": "0006019", "label": "inference based on individual clinical experience used in
manual assertion"}],
      "assertionSource":
"https://mmp.sfb.uit.no/databases/marref/#/records/MMP02469391",
      "assertionAdditionalInfo": "Curator spotted error during project review",
      "providerName": "MarRef",
      "providerUrl":
"https://mmp.sfb.uit.no/databases/marref/#/records/MMP02469391",
      "submittedTimestamp": "2019-01-29T14:04:37.498+0000",
      "archiveProcessing": [{
        "archiveName": "biosamples",
        "archiveProcessingStatus": "processed",
        "archiveProcessingNotes": "processed the curation",
        "timestamp": "2019-01-29T16:24:21.633+0000"
      }],
      "updatedTimestamp": "2019-01-29T16:24:21.633+0000",
      "suppressed": false
    }, {
      "id": "36619ac9-88bb-435f-bece-40d09205e149",
      "recordType": "sample",
      "recordId": "SAMN02469391",
      "attributePre": "isolation source",
      "attributePost": "geographic location",
      "attributeDelete": true,
      "assertionMethod": "manual assertion",
      "assertionEvidences": [{"identifier": "ECO:0006019", "shortForm": "
ECO_0006019", "label": "inference based on individual clinical experience used in manual
assertion"}],
      "assertionSource":
"https://mmp.sfb.uit.no/databases/marref/#/records/MMP02469391",
      "assertionAdditionalInfo": "Curator spotted error during project review",
      "providerName": "MarRef",
      "providerUrl":
"https://mmp.sfb.uit.no/databases/marref/#/records/MMP02469391",
      "submittedTimestamp": "2019-01-29T14:20:44.109+0000",
      "updatedTimestamp": "2019-01-29T14:20:44.109+0000",
      "suppressed": false
    }
  ],
  "totalAttributes": "2"

```

}

[]

Retrieve curation by Curation ID

| Field | Example | Type | Requirement | Notes |
|-------------|--|--------|-------------|------------------------------|
| Curation ID | 623a1df1- eeeb-483a- 9938- abf24f2f5187 | String | Mandatory | ID of the submitted curation |

```
curl -X GET "https://www.ebi.ac.uk/ena/clearinghouse/api/curations/faf1032a-8166-44e1-b197-6d2e58a6dc14" -H "accept: */*"
```

Status: 200 OK

```
{
  "curations": [{
    "id": "3051ca86-8584-4acb-8f49-c8cc75eeec4b",
    "recordType": "sample",
    "recordId": "SAMN02469391",
    "attributePre": "geographic location",
    "valuePre": "Norway: Trondheim fjord",
    "attributePost": "geographic location",
    "valuePost": "Norway: Lysefjord fjord",
    "attributeDelete": false,
    "assertionMethod": "manual assertion",
    "assertionEvidences": [{"identifier": "ECO:0006019",
"shortForm": "ECO_0006019", "label": "inference based on individual clinical experience
used in manual assertion"}],
    "assertionSource":
"https://mmp.sfb.uit.no/databases/marref/#/records/MMP02469391",
    "assertionAdditionalInfo": "Curator spotted error during project review",
    "providerName": "MarRef",
    "providerUrl":
"https://mmp.sfb.uit.no/databases/marref/#/records/MMP02469391",
    "archiveProcessing": [{
      "archiveName": "biosamples",
      "archiveProcessingStatus": "on_hold",
      "archiveProcessingNotes": "notes for biosamples",
      "timestamp": "2019-02-15T13:23:41.414+0000"
    },
    {
      "archiveName": "string",
      "archiveProcessingStatus": "processed",
      "archiveProcessingNotes": "string",
      "timestamp": "2019-02-18T16:27:41.643+0000"
    },
    {
      "archiveName": "biosamples",
```

```
        "archiveProcessingStatus": "processed",
        "archiveProcessingNotes": "sample accepted",
        "timestamp": "2019-02-22T14:29:59.175+0000"
      }
    ],
    "submittedTimestamp": "2019-02-15T13:09:26.847+0000",
    "updatedTimestamp": "2019-02-22T14:36:12.065+0000",
    "suppressed": true
  }],
  "totalAttributes": 1
}
```

Status: 404 NOT FOUND

```
{
  "success": false,
  "timestamp": "2019-01-29T16:31:46.467+0000",
  "errors": [
    {
      "message": "Record Not Found"
    }
  ]
}
```

Retrieve curation by Record ID

| Field | Example | Type | Requirement | Notes |
|---------------|---------------------|--------|-------------|---|
| Record ID | SAMN02469391 | String | Mandatory | Archive record identifier as path variable |
| attributeName | geographic location | String | Optional | Attribute Pre or Attribute Post of the submitted curation |

```
curl -X GET
```

```
"https://www.ebi.ac.uk/ena/clearinghouse/api/curations/SAMN02469391?attributeName=geographic%20location" -H "accept: */*"
```

Status: 200 OK

```
{
  "curations": [{
    "id": "3051ca86-8584-4acb-8f49-c8cc75eeec4b",
    "recordType": "sample",
    "recordId": "SAMN02469391",
    "attributePre": "geographic location",
    "valuePre": "Norway: Trondheim fjord",
    "attributePost": "geographic location",
    "valuePost": "Norway: Lysefjord fjord",
    "attributeDelete": false,
    "assertionMethod": "manual assertion",
    "assertionEvidences": [{"identifier": "ECO:0006019", "shortForm": "ECO_0006019", "label": "inference based on individual clinical experience used in manual assertion"}],
    "assertionSource":
      "https://mmp.sfb.uit.no/databases/marref/#/records/MMP02469391",
    "assertionAdditionalInfo": "Curator spotted error during project review",
    "providerName": "MarRef",
    "providerUrl":
      "https://mmp.sfb.uit.no/databases/marref/#/records/MMP02469391",
    "archiveProcessing": [{
      "archiveName": "biosamples",
      "archiveProcessingStatus": "on_hold",
      "archiveProcessingNotes": "notes for biosamples",
      "timestamp": "2019-02-15T13:23:41.414+0000"
    }, {
      "archiveName": "string",
      "archiveProcessingStatus": "processed",
      "archiveProcessingNotes": "string",

```

```

        "timestamp": "2019-02-18T16:27:41.643+0000"
      }, {
        "archiveName": "biosamples",
        "archiveProcessingStatus": "processed",
        "archiveProcessingNotes": "sample accepted",
        "timestamp": "2019-02-22T14:29:59.175+0000"
      }],
      "submittedTimestamp": "2019-02-15T13:09:26.847+0000",
      "updatedTimestamp": "2019-02-22T14:36:12.065+0000",
      "suppressed": true
    }, {
      "id": "0020f090-a081-4be7-ae51-0743a3a120f1",
      "recordType": "sample",
      "recordId": "SAMN02469391",
      "attributePre": "isolation source",
      "attributePost": "geographic location",
      "attributeDelete": true,
      "assertionMethod": "manual assertion",
      "assertionEvidences": [{"identifier": "ECO:0006019", "shortForm": "ECO_0006019", "label": "inference based on individual clinical experience used in manual assertion"}],
      "assertionSource":
        "https://mmp.sfb.uit.no/databases/marref/#/records/MMP02469391",
      "assertionAdditionalInfo": "Curator spotted error during project review",
      "providerName": "MarRef",
      "providerUrl":
        "https://mmp.sfb.uit.no/databases/marref/#/records/MMP02469391",
      "archiveProcessing": [{
        "archiveName": "biosamples",
        "archiveProcessingStatus": "processed",
        "archiveProcessingNotes": "test notes for biosamples",
        "timestamp": "2019-02-18T09:51:28.833+0000"
      }, {
        "archiveName": "biosamples",
        "archiveProcessingStatus": "processed",
        "archiveProcessingNotes": "test notes for biosamples 1",
        "timestamp": "2019-02-18T09:51:34.824+0000"
      }, {
        "archiveName": "biosamples",
        "archiveProcessingStatus": "processed",
        "archiveProcessingNotes": "test notes for biosamples 2",
        "timestamp": "2019-02-18T09:51:39.589+0000"
      }, {
        "archiveName": "biosamples",
        "archiveProcessingStatus": "processed",
        "archiveProcessingNotes": "test notes for biosamples 3",
        "timestamp": "2019-02-18T10:16:42.055+0000"
      }],
      "submittedTimestamp": "2019-02-18T09:50:13.485+0000",

```

```
    "updatedTimestamp": "2019-02-18T10:16:42.055+0000",
    "suppressed": false
  }, {
    "id": "7f26c676-cb4a-496a-abb6-3449ace2a987",
    "recordType": "sample",
    "recordId": "SAMN02469391",
    "attributePre": "geographic location",
    "valuePre": "Norway: Trondheim fjord",
    "attributePost": "geographic location",
    "valuePost": "Norway: Lysefjord fjord",
    "attributeDelete": false,
    "assertionMethod": "manual assertion",
    "assertionEvidences": [{"identifier": "ECO:0006019", "shortForm": "
ECO_0006019", "label": "inference based on individual clinical experience used in manual
assertion"}],
    "assertionSource":
    "https://mmp.sfb.uit.no/databases/marref/#/records/MMP02469391",
    "assertionAdditionalInfo": "Curator spotted error during project review",
    "providerName": "MarRef",
    "providerUrl":
    "https://mmp.sfb.uit.no/databases/marref/#/records/MMP02469391",
    "submittedTimestamp": "2019-02-22T14:02:20.691+0000",
    "updatedTimestamp": "2019-02-22T14:02:20.691+0000",
    "suppressed": false
  }],
  "totalAttributes": 3
}
```

Status: 404 NOT FOUND

```
{
  "success": false,
  "timestamp": "2019-01-29T16:31:46.467+0000",
  "errors": [
    {
      "message": "Record Not Found"
    }
  ]
}
```

Status: 400 BAD REQUEST

```
{
  "success": false,
  "timestamp": "2019-02-25T14:30:22.469+0000",
  "errors": [{
    "message": "Curation/Record ID is invalid."
  }]
}
```

Duplication Handling

Duplication of curation data during the creation has been handled based on the below scenarios.

- Each new curation record will be compared with the existing record available based on the recordId and provider name.
- If the existing record is available, then compare each value of the record with existing record except curation id, created timestamp and updated timestamp.
- If all the values are matching, then new curation record cannot be created, and the existing curation id will be returned in API.
- If some or none of the values are matching, then new curation record can be created.

Error Response messages

1. **Record Not found** - If no records are found from the curations data store.
2. **Forbidden** - If the user doesn't have permission to perform the actions.
E.g. Only Archive users can change the status of the Curations data
3. **Unauthorized** - If the user is not authorized by the AAP.
E.g. If the user token is invalid or expired
4. **<Field name> must not be null** - If any of the mandatory fields are missed
5. **Curation ID is invalid** - For updated operations (HTTP PATCH), if the Id provided does not match the valid curation ID pattern (UUID).
6. **Curation ID/Record ID is invalid** - For read operation (HTTP GET), if the Id provided does not match the valid curation ID pattern or a supported Accession pattern.
7. **No data identifier specified for <datatype> datatype** - If data type is given and data identifier is missing

Database fields

Curation record

| Field | Example | Type | Requirement | Notes |
|----------------|--------------------------------------|---|--|---|
| _id | dee11d4e-63c6-4d90-983c-5c9f1e79e96c | System set uuid | System set | System set uuid. Set by MongoDB UUID() command. |
| recordType | 'sample'/'sequence' | String. controlled list (sample sequence) | Mandatory | Target type to determine which archives will be interested. API will throw error if not value from controlled list. |
| recordId | SAMEXXXX XXX/ERRXX XXX | free text string | Mandatory | archive record identifier. Initially will be free text string with no validation. Future versions of CDCH may utilise dependent archive accession validation code/APIs. |
| dataType | covid19 | String. controlled list (covid19) | Optional | Target type to determine the data type of the record. API will throw error if not value from controlled list. |
| dataIdentifier | COVIDI26 | free text string | Optional (Mandatory only if datatype is mentioned) | Data identifier for the given dataType. |
| attributePre | sex/location/breed | free text string | Optional (at least one of | Preexisting attribute to |

| | | | | |
|-------------------|--|--|--|---|
| | | | attribute_pre or attribute_post is required) | correct. See use cases section for appropriate usage. |
| valuePre | male/UK/Ram bouillet | free text string | Optional | Pre-existing value to curate. See use cases section for appropriate usage. |
| attributePost | sex/location/breed | free text string | Optional (at least one of attribute_pre or attribute_post is required) | Replace current attribute with this one. See use cases section for appropriate usage. |
| valuePost | male/UK/Ram bouillet | free text string | Optional | Replace current value with this one. See use cases section for appropriate usage. |
| attributeDelete | true | boolean | Optional | Flag for deleting the attribute listed in attribute_pre. Must not specify attribute_post or value_post |
| assertionMethod | 'automatic assertion'/'manual assertion' | Valid text or synonym that is ontology child of ECO_0000217 assertion method | Mandatory | Whether curation was determined manually or automatically taken from Evidence & Conclusion Ontology (ECO) |
| assertionEvidence | AssertionEvidence | List. See below. | Mandatory | Supporting evidence for the assertion taken from Evidence & Conclusion Ontology (ECO) |
| assertionSource | | Valid URI | Optional (could make mandatory for some values of | Valid URI link either to the supporting publication or external reference |

| | | | | |
|-------------------------|---|------------------|--|---|
| | | | assertion_evidence) | database that supports the stated assertion |
| assertionAdditionalInfo | | free text | Optional | Additional free text info to support assertion |
| aapAccount | User | DBRef | System captured from authorisation token | Record AAP Details of provider. |
| providerName | MarRef/EMG/Silva | free text | Mandatory | name of provider |
| providerUrl | https://mmp.sfb.uit.no/databases/marref/ | Valid URI | Optional | URL of the provider |
| submittedTimestamp | 2018-11-30T09:15:15.353Z | ISODATE | System captured | Timestamp always saves in UTC datetime. Set first time record is created. |
| updatedTimestamp | 2018-11-30T09:15:15.353Z | ISODATE | System captured | Timestamp always saves in UTC datetime. Set every time record is updated. |
| suppressed | true/false | boolean | Optional | Curation data can be suppressed only by the record owner |
| archiveProcessing | ArchiveProcessing | List. See below. | Optional | can only be edited by authenticated archive |

Assertion Evidence

| Field | Example | Type | Requirement | Notes |
|------------|----------------------------------|---|---|---|
| identifier | ECO_0006019 Or ECO:0006019 | Valid Identifier that is ontology child of ECO_0000000 evidence | Optional (at least identifier or label is required) | Supporting evidence for the assertion taken from Evidence & Conclusion Ontology |

| | | | | |
|-------|--|--|---|--|
| | | | | (ECO). It will also update the according assertion text in the curation record |
| label | author statement from published clinical study | Valid text or synonym that is ontology child of ECO_0000000 evidence | Optional (at least identifier or label is required) | Supporting evidence for the assertion taken from Evidence & Conclusion Ontology (ECO). It will also update the according ECO Identifier in the curation record |

NOTE: If both ECO Identifier and assertion text has been passed, then ECO Identifier takes the precedence and overwrite its assertion text accordingly

Archive Processing

| Field | Example | Type | Requirement | Notes |
|-------------------------|---|-----------------|-----------------|--|
| archiveName | biosamples | Free text | Mandatory | can only be edited by authenticated archive |
| archiveProcessingStatus | processed/on_hold/rejected | Controlled list | Mandatory | can only be edited by authenticated archive |
| archiveProcessingNotes | Accepted but formatted to meet archive convention | free text | Optional | can only be edited by authenticated archive |
| timestamp | 2018-11-30T09:15:15.353Z | ISODATE | System captured | Timestamp always saves in UTC datetime. Set every time archive data are added. |

User

| Field | Example | Type | Requirement | Notes |
|----------|--|-------------------|--|---|
| _id | usr-c29baadb-3d75-4b49-8888-d42fea25732b | valid AAP account | System captured from authorisation token | record AAP ID of provider. Obtained from AAP, userReference field |
| userName | Joe Bloggs | valid AAP account | System captured from authorisation token | AAP name field. This field is restricted and will not be in public read API. Will only be available to approved authenticated archives. Subject to GDPR. |
| email | person@someplace.com | valid AAP account | System captured from authorisation token | AAP email field. This field is restricted and will not be in public read API. Will only be available to approved authenticated archives. Subject to GDPR. |

Appendix - A

Table -1 RecordId Validation Patterns

| RecordType | Pattern Identifier | Validation Pattern |
|------------|--------------------|---|
| Sequence | SEQUENCE_REGEX | ([A-Z][0-9]{5,6}(\.\.[0-9]+) ([A-Z]{2}[0-9]{5,6}(\.\.[0-9]+)) ([a-zA-Z]{4}[0-9]{2}S[0-9]{6,8}(\.\.[0-9]+))) |
| | RUN_REGEX | ([ESDR]RR[0-9]{6,}) |
| Sample | SAMPLE_REGEX | ([EDSR]RS[0-9]{6,}) |
| | BIO_SAMPLE_REGEX | (SAME[A]?[0-9]{6,} SAM[ND][0-9]{8,}) |
| | RUN_REGEX | ([ESDR]RR[0-9]{6,}) |
| Study | PROJECT_REGEX | (PRJ[A-Z]{2}[0-9]+) |
| | STUDY_REGEX | ([ESDR]RP[0-9]{6,}) |