DATA EXHAUST API(s)

Accessing various DIKSHA datasets via data exhaust APIs

Version 1.0

SEPTEMBER 2020

INTRODUCTION

Diksha Datasets under consideration - to be accessed via APIs/ Diksha portal UI

Dataset	For	Description	API Access	UI Access (Diksha Portal)
Course Progress Dataset	Course Creator/ Mentor	Indicates progress of users taking a specific course - including progress of nested courses, and assessment scores, if any.	Yes	Course Dashboard Page
Course User Info Dataset	Course Creator/ Mentor	Lists PII information as declared by users taking a specific course - this is made available only if the user consents to sharing their data.	Yes	Course Dashboard Page
Course assessment response info	Course Creator/ Mentor	Lists users' responses to all questions in a course	Yes	Course Dashboard Page

CONSENT & DATA ACCESS

User consent and various levels of data access (See Appendix for full list of fields in each Dataset)

Dataset	Accessed by	User Consent	Data Access level
Course Progress Dataset	Content Creator/Mentor	N.A.	Course and user progress data with only Name, State & District of the user
Course User Info Dataset	Content Creator/Mentor	No	Course and user progress data with only Name, State & District of the user
		Yes	Course data with complete PII information as provided by the user (Mobile phone and Email ID are shared only for users who have consented to share their details at the time of joining the course)
Course assessment response info	Content Creator/Mentor	N.A.	Users' responses to all questions in a course/ assessment

API Details

Introduction

As the exhaust files are pretty large and need to aggregate data from various sources, they are run as a set of scheduled jobs with a turn around time of 24 hours. There are the following set of APIs involved in making an exhaust request.

1. **Submit Request API** - API to submit a job request. The job status will be in the submitted state.
2. **Get Request API** - API to get the request status and details. If the request is processed the API would return the files available for download.
3. **List Request API** - API to list the last 10 requests made. This is a rarely used API to fetch and download historical exhaust files.

A regular scheduled exhaust pull from any state would involve only the first two APIs - one to submit a request and another to check the request status and download files.

API Endpoints

Following are the endpoints of the 3 APIs

1. **Submit Request API -** *<baseurl>/api/dataset/v1/request/submit*
2. **Get Request API -** *<baseurl>/api/dataset/v1/request/read/{tag}?requestId={requestId}*
3. **List Request API -** *<baseurl>/api/dataset/v1/request/list/{tag}*

Base URL

The base url for DIKSHA preprod will be *https://preprod.ntp.net.in*

The base url for DIKSHA production will be *https://diksha.gov.in*

API Access

An authentication token is required to access the DIKSHA APIs. Please reach out to support via FreshDesk with the following information:

- State name
- State org admin email ID/Mobile number
- Purpose of this request
    - For example., new request to integrate state course progress information to

be integrated with my state portal. token not working, forgot my token etc

API Invocation

Every API follows a standard request and response structure with the authentication token passed as header parameter.

Sample Request

```
{
  "params": {
    "msgid": "4f04da60-1e24-4d31-aa7b-1daf91c46341"
  },
  "request": <RequestData>
}
```

Sample Response

```
{
  "params": {
    "msgid": "4f04da60-1e24-4d31-aa7b-1daf91c46341"
    "status": "successful"
  },
  "responseCode": "OK",
  "result": <ResponseData>
}
```

Sample API Invocation using Curl

```
curl --request POST 'https://diksha.gov.in/api/dataset/v1/request/submit \
--header 'Authorization: Bearer <Pass Authentication token here>' \
--header 'Content-Type: application/json' \
--data-raw '{
  "params": {
    "msgid": "4f04da60-1e24-4d31-aa7b-1daf91c46341"
  }
  "request": <RequestData>
}'
```

Submit Request API

API Endpoint: POST <baseurl>/api/dataset/v1/request/submit

Request Headers (Required)

Header	Value	Description
Authorization	Bearer <auth token>	Header to pass the API authentication token

Content-Type	application/json	Defaults to JSON
X-Channel-Id	<organization id>	The organization id of the tenant/state

Request Data

Key	Data Type	Mandatory	Description
tag	String	Yes	If you are creating one exhaust request per batch, you can use the batchId as the tag instead of org Id. If you are creating one exhaust request for all batches, you can use the orgId as the tag (or) any random text for tag.
dataset	String	Yes	The job/dataset id the request is for. See jobIds section for the jobIds for all datasets
datasetConfig	json	Yes	Job configuration. See the example for more info
encryptionKey	String	Yes	Encryption key for the data exhaust file

Response Data

Key	Data Type	Description
tag	String	Unique identifier to identify the request source passed in the request
dataset	String	The dataset id the request is for. See the “Dataset Details” section for more info
datasetConfig	Json	Dataset configuration passed in the request. See the “Dataset Details” section info
requestId	String	Unique request id for the request. This id is used to query for status back using the Get Request API
requestedChannel	String	The channel requesting the data
status	String	Status of the request. Will be “SUBMITTED” when the request is made

Dataset Details

Following are the dataset ids for all datasets

1. Course Progress Dataset - **progress-exhaust**
2. Course User Info Dataset - **userinfo-exhaust**

3. Course Assessment Response Dataset - **response-exhaust**

All jobs have the following standard configuration that can be passed.

```
{
  "batchId": String,
  "batchFilter": Array[String],
  "searchFilter": SearchCriteria
}
```

One of “batchId” or “batchFilter” or “searchFilter” is required else the request would fail with an “Invalid Request” status message.

SearchCriteria Structure

The SearchCriteria is used to choose the trackable objects and identify the batch ids for processing the dataset request.

SearchCriteria with Trackable Object Identifiers	SearchCriteria based on limit

Format:
`{`
`  "request": {`
`    "filters": {`
`      "contentType": String,`
`      "channel": String,`
`      "identifier": Array[String]`
`    },`
`    "limit": Int`
`  }`
`}`
	Format:
	`{`
	`  "request": {`
	`    "filters": {`
	`      "contentType": String,`
	`      "channel": String`
	`    },`
	`    "limit": Int`
	`  }`
	`}`

• All the properties are required. • “contentType” is always “Course”. • Uses the identifiers specified in the SearchCriteria • When we use identifiers in the SearchCriteria, we can submit multiple requests per dataset.	• All the properties are required. • “contentType” is always “Course”. • Randomly select the objects with the given filters based on the limit value. • “limit” should be less than 50. • Only one request per dataset we can submit with this SearchCriteria.
`{`	`{`

```
  "request": {
    "filters": {
      "contentType":"Course",
      "channel": "0123456789",
      "identifier": ["course1"]
    },
    "limit": 50
  }
}
```

```
  "request": {
    "filters": {
      "contentType":"Course",
      "channel": "0123456789"
    },
    "limit": 50
  }
}
```

Note: *When you use SearchCriteria to submit a request, we recommend to use the* ***"SearchCriteria with trackable object identifiers"****.*

[IMPORTANT] The batchFilter field should be populated only with valid batch IDs from the system

Sample Request

```
curl --request POST 'https://diksha.gov.in/api/dataset/v1/request/submit \
--header 'Authorization: Bearer {{api_key}}' \
--header 'Content-Type: application/json' \
--header 'X-Channel-Id: 0126796199493140480' \
--data-raw '{
    "request": {
    "tag": "013109114254794752101111202O",
    "dataset": "progress-exhaust",
    "datasetConfig": {
      "batchFilter": ["01310911425479475210",”batch2”,”batch3”]
    },
    "encryptionKey": "uKW)Afn9D5"
  }
}'
```

Sample Response

```
{
  "params": {
    "msgid": "4f04da60-1e24-4d31-aa7b-1daf91c46341"
    "status": "successful"
  },
  "responseCode": "OK",
  "result": {
    "tag": "013109114254794752101111202O",
    "dataset": "progress-exhaust",
    "datasetConfig": {
      "batchFilter": ["01310911425479475210",”batch2”,”batch3”]
    },
    "requestId": "462CDD1241226D5CA2E777DA522691EF",
    "requestedChannel": "0126796199493140480",
    "status": "SUBMITTED"
```

```
  }
}
```

Sample Request with SearchCriteria

```
curl --request POST 'https://diksha.gov.in/api/dataset/v1/request/submit \
--header 'Authorization: Bearer {{api_key}}' \
--header 'Content-Type: application/json' \
--header 'X-Channel-Id: 0126796199493140480' \
--data-raw '{
    "request": {
    "tag": "013109114254794752101111 2020",
    "dataset": "progress-exhaust",
    "datasetConfig": {
      “searchFilter”: {
        "request": {
          "filters": {
            "contentType":"Course",
            "channel": "0126796199493140480",
            "identifier": ["course1"]
          }
        }
      }
    },
    "encryptionKey": "uKW)Afn9D5"
  }
}'
```

Sample Response

```
{
  "params": {
    "msgid": "4f04da60-1e24-4d31-aa7b-1daf91c46341"
    "status": "successful"
  },
  "responseCode": "OK",
  "result": {
    "tag": "013109114254794752101111 2020",
    "dataset": "progress-exhaust",
    "datasetConfig": {
      "batchFilter": ["01310911425479475210",”batch2”,”batch3”]
    },
    "requestId": "462CDD1241226D5CA2E777DA522691EF",
    "requestedChannel": "0126796199493140480",
    "status": "SUBMITTED"
  }
}
```

[IMPORTANT] Input errors to look out for: At times, input might be wrongly submitted with “batchId” as array or “batchFilter” as string data. In order to verify the input, kindly use following Link to verify schema https://www.jsonschemavalidator.net/s/SmDRMnsM

Get Request API

API Endpoint: GET <baseurl>/api/dataset/v1/request/read/{tag}?requestId={requestId}

Request Headers (Required)

Header	Value	Description
Authorization	Bearer <auth token>	Header to pass the API authentication token
Content-Type	application/json	Defaults to JSON
X-Channel-Id	<organization id>	The organization id of the tenant/state

Request Parameters

Key	Data Type	Mandatory	Description
tag	String	Yes	Unique identifier to identify the request source. **Use the same value passed in the "tag" field of the submit API**
requestId	String	Yes	The request_id field received from the submit API

Response Data (Key fields)

Key	Data Type	Description
tag	String	Unique identifier to identify the request source passed in the request
dataset	String	The dataset id the request is for.
datasetConfig	Json	Dataset configuration passed in the request
requestId	String	Unique request id for the request. This id is used to query for status back using the Get Request API
requestedChannel	String	The channel requesting the data
status	String	Status of the request. See the "Job Statuses" section for all available statuses
lastUpdated	DateTime	The timestamp denoting when the job was last updated

downloadUrls	List	List of download links for the encrypted files
expiresAt	Long	System readable timestamp by when the links would expire.
statusMessage	String	Error message if any. Only applicable when the job status is FAILED.

Note:

- Each GET API request for the {{tag}} and {{requestId}} will provide an updated “downloadUrls” link with corresponding updated “expiresAt” time in the response. Current link expiry set at 30 mins.
- statusMessage “No data found” implies that batchId submitted is not found.
- statusMessage “Invalid input data” implies that the request submitted is in wrong format.

Job Statuses

Following are the job statuses

1. SUBMITTED - Initial status when the job was submitted
2. PROCESSING - Status when the job was picked up for processing
3. SUCCESS - Status when the job run was successful
4. FAILED - Status if the job has failed.

Sample Request

```
curl --request GET
'https://diksha.gov.in/api/dataset/v1/request/read/0131091142547947521011112020?reques
tId=462CDD1241226D5CA2E777DA522691EF' \
--header 'Authorization: Bearer {{api_key}}' \
--header 'Content-Type: application/json' \
--header 'X-Channel-Id: 0126796199493140480'
```

Sample Response

```
{
  "params": {
    "msgid": "4f04da60-1e24-4d31-aa7b-1daf91c46341"
    "status": "successful"
  },
  "responseCode": "OK",
  "result": {
```

```
    "tag": "0126796199493140480",
    "dataset": "progress-exhaust",
    "datasetConfig": {
      "batchId": "01310911425479475210"
    },
    "requestId": "462CDD1241226D5CA2E777DA522691EF",
    "requestedChannel": "0126796199493140480",
    "status": "SUCCESS"
    "downloadUrls": ["https://<blobstore>/01310911425479475210_progress_20200920.zip"]
    "expiresAt": 1600805630000
  }
}
```

List Request API

API Endpoint: GET <baseurl>/api/dataset/v1/request/list/{tag}

Request Headers (Required)

Header	Value	Description
Authorization	Bearer <auth token>	Header to pass the API authentication token
Content-Type	application/json	Defaults to JSON
X-Channel-Id	<organization id>	The organization id of the tenant/state

Request Parameters

Key	Data Type	Mandatory	Description
tag	String	Yes	Unique identifier to identify the request source. **Use the same value passed in the "tag" field of the submit API**

Response Data (Key fields) - List of Requests

Key	Data Type	Description
tag	String	Unique identifier to identify the request source passed in the request
dataset	String	The job/dataset id the request is for.
datasetConfig	Json	Job configuration passed in the request
requestId	String	Unique request id for the request. This id is used to query for status back using the Get Request API

requestedChannel	String	The channel requesting the data
status	String	Status of the request.
lastUpdated	DateTime	The timestamp denoting when the job was last updated
downloadUrls	List	List of download links for the encrypted files
expiresAt	Long	System readable timestamp by when the links would expire
statusMessage	String	Error message if any. Only applicable when the job status is FAILED.

Sample Request

```
curl --request GET
'https://diksha.gov.in/api/dataset/v1/request/list/0126796199493140480' \
--header 'Authorization: Bearer {{api_key}}' \
--header 'Content-Type: application/json' \
--header 'X-Channel-Id: 0126796199493140480'
```

Sample Response

```
{
  "params": {
    "msgid": "4f04da60-1e24-4d31-aa7b-1daf91c46341"
    "status": "successful"
  },
  "responseCode": "OK",
  "result": [{
    "tag": "0126796199493140480",
    "dataset": "progress-exhaust",
    "datasetConfig": {
      "batchId": "01310911425479475210"
    },
    "requestId": "462CDD1241226D5CA2E777DA522691EF",
    "requestedChannel": "0126796199493140480",
    "status": "SUCCESS"
    "downloadUrls": ["https://<blobstore>/01310911425479475210_progress_20200920.zip"]
    "expiresAt": 1600805630000
  },{
    "tag": "0126796199493140480",
    "dataset": "userinfo-exhaust",
    "datasetConfig": {
      "batchId": "01310911425479475210"
    },
    "requestId": "462CDD1241226D5CA2E777DA522691EG",
    "requestedChannel": "0126796199493140480",
    "status": "SUCCESS"
    "downloadUrls": ["https://<blobstore>/01310911425479475210_userinfo_20200920.zip"]
```

```
    "expiresAt": 1600805630000
  },{
    "tag": "0126796199493140480",
    "dataset": "response-exhaust",
    "datasetConfig": {
      "batchId": "01310911425479475211"
    },
    "requestId": "462CDD1241226D5CA2E777DA522691EG",
    "requestedChannel": "0126796199493140480",
    "status": "FAILED"
    "statusMessage": "Requested batch not found/Batch has expired"
  }]
}
```

Release 3.3 Deliverables (01 Oct 2020)

Below data exhausts will be made available as on-demand reports to the course creator/mentor. In addition, the state tech team can also automatically pull this information from the APIs by following the instructions provided in this document.

- Course progress dataset
- Course user info dataset
- Course assessment response info

APPENDIX :: DATASET FIELDS

(I) COURSE PROGRESS DATASET

Column Label	Column Type	Data Type	Description

Collection Id	Static	String	Id of the collection.
Collection Name	Static	String	Collection Title
Batch Id	Static	String	Batch Id
Batch Name	Static	String	Batch Title
User UUID	Static	String	The system generated DIKSHA unique user ID
State	Static	String	User declared state for self signed up users. If the user belongs to a state tenant, then the state as passed from state SSO or derived from school ID.
District	Static	String	User declared district for self signed up users. If the user belongs to a state tenant, then the district as passed from state SSO or derived from school ID.
Org Name	Static	String	Name of user org - DIKSHA Custodian for self signed up users and respective tenant names if the user belongs to a state tenant
School Id	Static	String	If user belongs to the state tenant, then the school ID mapped to this user. If user is self declared user then the

			user declared school ID.
School Name	Static	String	If user belongs to a state tenant, then the school name mapped to this user. If user is self declared user then the user declared org/school name.
Block Name	Static	String	The Block, if selected by the user during on-boarding/ from their profile.
Declared Board	Static	String	The board selected during user on-boarding.
Cluster	Static	String	The Cluster, if selected by the user during on-boarding/ from their profile
Usertype	Static	String	One of the four ‘type’ values that the user can choose from - Teacher, Student, Parent, Officials
Usersubtype	Static	String	Sub type value, if available for the chosen Usertype - eg. HM, SPD, CRP etc.
Declared Org	Static	String	If the user is a self signed up user then this is the value filled by the user in the 'With' part of the self signed up declaration. For SSO users this will be blank.

Enrolment Date	Static	Date	Collection enrolment date (for nested courses/collections it will be the parent collection enrolment date)
Completion Date	Static	Date	Collection completion date (for nested courses/collections it will be the parent collection completion date)
Progress	Static	Number	Collection progress (for nested courses/collections this will be the parent collection progress)
Certificate Status	Static	String	Issued - if the certificate is issued. Blank - if it is not issued and. Failed - if issue has failed
Total Score	Static	Number	Total Score received by the user across all assessments within the collection with category type as “SelfAssess”
<nested_collection_id> - Progress	Dynamic	Number	User’s progress at a nested collection level. This is a dynamic column. For ex: If there are 3 nested trackable collections within the parent collection there will be 3 corresponding columns
<assessment_id> - Score	Dynamic	Number	User’s best attempt for the given assessment id. This is a dynamic column. For ex: If there are 3 assessments within the parent collection there will be 3 corresponding

			columns

(II) COURSE USER INFO DATASET

Column Label	Column Type	Data Type	Description
Collection Id	Static	String	Id of the collection.
Collection Name	Static	String	Collection Title
Batch Id	Static	String	Batch Id
Batch Name	Static	String	Batch Title
User UUID	Static	String	The system generated DIKSHA unique user ID
User Name	Static	String	Name of the user

(On user consent)			
State	Static	String	User declared state for self signed up users. If the user belongs to a state tenant, then the state as passed from state SSO or derived from school ID.
District	Static	String	User declared district for self signed up users. If the user belongs to a state tenant, then the district as passed from state SSO or derived from school ID.
Org Name	Static	String	Name of user org - DIKSHA Custodian for self signed up users and respective tenant names if the user belongs to a state tenant
Mobile number (On user consent)	Static	String	Mobile number of the user, as per the user profile. This unmasked mobile number is provided only if the users have consented to share

			their details at the time of course enrolment.
Email ID (On user consent)	Static	String	Email ID of the user, as per the user profile.. This unmasked email ID is provided only if the users have consented to share their details at the time of course enrolment.
Consent Provided	Static	String	Yes/No. Flag to denote whether user has consented to the data sharing.
Consent Provided Date	Static	Date	Date when the user has consented to share the data
Block Name	Static	String	Block name mapped to the user's org/school id. If user is self declared user then this column will be blank.
Cluster	Static	String	The Cluster, if selected by the user during on-boarding/ from their profile

Usertype	Static	String	One of the four ‘type’ values that the user can choose from - Teacher, Student, Parent, Officials
Usersubtype	Static	String	Sub type value, if available for the chosen Usertype - eg. HM, SPD, CRP etc.
School Id	Static	String	The ID of the school mapped to the user
School Name	Static	String	The name of the school mapped to the user

(III) COURSE ASSESSMENT RESPONSE INFO

Column Label	Colum	Data	Description

	n Type	Type	
Collection Id	Static	String	Id of the collection.
Collection Name	Static	String	Collection Title
Batch Id	Static	String	Batch Id
Batch Name	Static	String	Batch Title
User UUID	Static	String	The system generated DIKSHA unique user ID
QuestionSet Id	Static	String	Id of the question set
QuestionSet Title	Static	String	Title of the question set
Attempt Id	Static	String	Id of the attempt. There can be more than one attempt for the same question set
Attempted On	Static	DateTime	Date on which the last attempt happened for this attempt id

Question Id	Static	String	Question id
Question Type	Static	String	Question type - mcq/mtf/mmcq/ftb
Question Title	Static	String	Title of the question
Question Description	Static	String	Description of the question
Question Duration	Static	Number	Time taken to answer the question
Question Score	Static	Number	Score received for the question
Question Max Score	Static	Number	Max applicable score for the question
Question Options	Static	JSON	Options shown to the user
Question Response (On user	Static	JSON	Responses given by the user

consent for question type registration)			

(IV) ORG LEVEL USER PII DATASET (when the user provides Org level consent)

Columns	
Name	Name of the User
User UUID	System generated DIKSHA unique user ID
State	User-declared State for self signed up users. If the user is a state validated user then the state as passed from state SSO or derived from school ID
District	User-declared District for self signed up users. If the user is a state validated user then the state as passed from state SSO or derived from school ID

Block	Block, as declared by the user during on-boarding/ from their profile
Cluster	Cluster, as declared by the user during on-boarding/ from their profile
School Name	The name of the school mapped to the user
School UDISE ID	The UDISE ID of the school mapped to the user.
External ID	The external ID of the user as specified by the user (optional) while sharing global consent. For SSO users, this value is set when the state system passes the external ID of the user to Diksha.
Profile Email	Email ID of the user as provided in their profile
Profile Phone number	Mobile number of the user, as provided in their profile
Org Org Phone	Additional org-specific phone number (optional) as submitted by the user when they give consent to share their details with

	the Org
Org Email ID	Additional org-specific Email ID (optional) as submitted by the user when they give consent to share their details with the Org
Usertype	One of the four ‘type’ values that the user can choose from - Teacher, Student, Parent, Officials
Usersubtype	Sub type value, if available for the chosen Usertype - eg. HM, SPD, CRP etc.
Root Org of the user	Root Org association of the user