

## Centro DSP Audience Data Transfer v2.2

2017-04-05

### Overview

The following document outlines the technical process for integrating with Centro DSP to transmit audience data segments in the form of cookie IDs or mobile IFAs (identifier for advertising – Google Advertising ID or Apple IDFA).

Centro can receive audience data via:

- Real-time streaming S2S calls
- Batch file (preferred)

In either case, for cookie audiences, a user sync must be established. The data provider must store the user ID mapping, and transmit audiences to us in terms of our user ID.

Before data is transmitted, the segment taxonomy must be sent to Centro and confirmed loaded; we will discard data for segment IDs we do not recognize. Please send taxonomy to [dsp.newdata@centro.net](mailto:dsp.newdata@centro.net).

Centro DSP supports the ability to remove users from an audience if they were previously assigned to that audience. If audience removal is not used then the user will remain as part of the assigned audience(s) for a minimum of 30 days.

For engineering or operational assistance during integration please contact [dataintegration@centro.net](mailto:dataintegration@centro.net) and [rtbops@centro.net](mailto:rtbops@centro.net).

### User Sync

Centro will provide a pixel URL like this:

<http://pixel.sitescout.com/connectors/dataprovider/usersync?redir=>

The data provider should append a URL for us to redirect to, in URL encoded form. Include our {userId}” macro in the appropriate place and we will substitute this with our user ID. (Support for other macro strings can be provided if necessary, with advance notice.)

### Real-time streaming S2S (server to server) calls

Server-to-server audience data transfer uses HTTP(S) as the protocol for transferring data. Centro will provide a user data URL like this:

<http://pixel-a.sitescout.com/connectors/dataprovider/userdata>

Send HTTP POST requests to this URL with a JSON object with the following structure:

Key	Value	Description
users	JSON array containing JSON objects	Contains an array of user data or mobile-device IDs
userID	String	Centro DSP unique ID for this user provided by user matching process
idfa	String (hyphenated uppercase)	Apple identifier for advertising
googleAdID	String (hyphenated lowercase)	Google advertising ID
audiences	JSON array containing strings	Array of segment IDs this user is to be added to (as represented using the data provider's segment IDs)
removed	JSON array containing strings	Array of segment IDs this user is to be removed from

Note that, for each user object, only one field from `userID`, `idfa`, or `googleAdID` can be present. The other two fields must be absent. The `audiences` and `removed` fields can be present in the same user object.

#### Example (without audience removal)

```
{
  "users": [
    {
      "userID":
      "bb64ba25-01b0-421f-b1a7-3266c3913c74",
      "audiences": [
        "segment1",
        "segment2",
        "segmentN"
      ]
    },
    {
      "idfa": "9D02646A-3CB7-47FB-C5A7-1383DEC33292",
      "audiences": [
        "segment1",
        "segment2",
        "segmentN"
      ]
    }
  ]
}
```

### Example (with audience removal)

```
{
  "users": [
    {
      "userID":
      "bb64ba25-01b0-421f-b1a7-3266c3913c74",
      "audiences": [
        "segment1",
        "segment2",
        "segmentN"
      ]
      "removed": [
        "segment3",
        "segment4",
        "segmentN"
      ]
    }
  ]
}
```

### Rate Limits

Please limit transmission to a maximum of 100 users per request, 5 simultaneous connections, and no more than 3000 QPS total.

## Batch File (Preferred)

In this method, the data provider supplies file(s) to a drop box that will be collected and processed periodically by Centro. Batches will be in the format of comma-separated (CSV) or tab-separated (TSV) files, optionally compressed with gzip.

### File Format

The files will be named according to this pattern:

```
data_{centro-providerId}_{dateUTC(YYYYMMDDHHMMSS)}_part_{#}_of_{#}.{ext}{.gz}
```

The extension will be “csv” or “tsv” according to the chosen format, or “.csv.gz” / “.tsv.gz” if gzipped. An example of a file name:

```
data_15_20150608030413_part_1_of_12.tsv
```

Each file will be no greater than 50MB. If a batch will exceed this, it must be split into multiple files. Batch files may be full refresh (each file contains all users), or incremental (each file contains new

users since last batch file generation). We prefer incremental updates. For users on which we already hold a record, we will append any new audiences to the existing record.

### CSV Format

CSV format supports audience insertion only. The CSV file will have the following fields:

Field #	Type	Description
1	String	Identifier – i.e. Centro DSP user ID, IDFA or Google Advertising ID
2	Integer	ID type, where: 0 = Centro DSP user ID 1 = iOS IDFA (uppercase with hyphens) 2 = SHA1 hashed Android ID 3 = Google Advertising ID (lowercase with hyphens)
3, 4, 5, etc.	String	One or more audiences that this user ID is a member of, represented using the data provider's segment IDs.

No quotes will be used around each value. Each line represents a device and the audiences it belongs to. Unix-style newlines (LF) will be used. No additional spaces will be present before or after the separator.

An example of a line:

```
AEBE52E7-03EE-455A-B3C4-E57283966239,1,104,105,604,435,523
```

(This is an iOS IDFA that is in audiences 104, 105, 604, 435 and 523.)

### TSV Format

TSV format supports audience insertion and deletion. The TSV file will have the following fields:

Field #	Type	Description
1	String	Identifier – i.e. Centro DSP user ID, IDFA or Google Advertising ID
2	Integer	ID type, where: 0 = Centro DSP user ID 1 = iOS IDFA (uppercase with hyphens) 2 = SHA1 hashed Android ID 3 = Google Advertising ID (lowercase with hyphens)
3	String	One or more audiences that this user ID is a member of represented using the data provider's segment IDs. Each audience is separated by a comma.
4	String	One or more audiences that this user ID is to be removed from. Each audience is separated by a comma. This field is optional.

No quotes will be used around each value. Each line represents a device and the audiences it belongs to. Unix-style newlines (LF) will be used. No additional spaces will be present before or after the separator.

An example of a line with audience insertion only. For clarity, the tab character has been represented by "<TAB>":

AEBE52E7-03EE-455A-B3C4-E57283966239<TAB>1<TAB>104,105,604,435,523

(This is an iOS IDFA that is in audiences 104, 105, 604, 435, and 523.)

An example of a line with audience insertion and removal:

01327a89-07ec-4d63-bb5e-9d06e82d202f<TAB>0<TAB>456,234,565<TAB>104,105

(This is a cookie that is in audiences 456,234,565 but is no longer in audiences 104,105)

An example of a line with audience removal only:

01327a89-07ec-4d63-bb5e-9d06e82d202f<TAB>0<TAB><TAB>104,105

(This is a cookie that is no longer in audiences 104,105)

## Delivery and Schedule

Either the data provider will permission Centro to access an Amazon S3 bucket, or Centro will provide a bucket. The data provider will dump batch files into this bucket periodically. Each file will be represented by an S3 object.

Files must be retained in this bucket for at least 7 days, after which the provider may assume the file has been collected and they may safely delete it. If the S3 bucket is provided by Centro, we will auto-expire files after 7 days.

Centro will periodically check the S3 bucket for new files, at least once per day but as often as once per minute. We will download all new files since last processing (based on S3 timestamp) and queue them for insertion into our user profile store.

## Failures, Delivery Delays, Etc.

In the event of processing failures, network issues, delivery delays, or other unforeseen events, we may not collect new files on a given day, however we will fetch them at the next scheduled processing (the following day).

In the event that Centro is unable to parse a file (i.e. malformed data), we will skip the file and notify the data provider of the error by email. The data provide can correct the error and re-upload it to the S3 bucket with a new filename according to the above format. We will process it at the next scheduled processing time.

## Initial Load

If incremental updates are used, since each batch represents the changes from the previous batch, it will be necessary for the data provider to supply a full dump during initial onboarding. The data provider may either:

- Place files in the S3 bucket, or
- In the event that the S3 bucket is shared with multiple partners, and placing these files would interfere with their operations, special arrangements can be made to receive the initial load which will be manually processed

## Change History

Version	Change
2.2	<ul style="list-style-type: none"> <li>• Clarified case of googleAdID</li> </ul>
2.1	<ul style="list-style-type: none"> <li>• Revised user ID macro to our standard (“{userId}”)</li> <li>• Added gzip support</li> </ul>
2.0	<ul style="list-style-type: none"> <li>• Introduced TSV format with audience removal support</li> </ul>