

A minimal computing DTS API for TEI XML documents

Robert Casties 17.2.2022

What is DTS

The Distributed Text Services (DTS) Specification defines an API for working with collections of text as machine-actionable data.

Publishers of digital text collections can use the DTS API to help them make their textual data Findable, Accessible, Interoperable and Reusable (FAIR).

DTS enables machine-consumption of digital text collections, and can be used by consumers of these collections in a variety of ways, such as for data analysis and the development of user-interfaces, tools and services.

(<https://w3id.org/dts>)

DTS API

DTS specifies 3 distinct operation endpoints:

- Navigation across texts is supported by the **Collections Endpoint**
- Navigation within a text is supported by the **Navigation Endpoint**
- Retrieval of complete or partial texts is supported by the **Documents Endpoint**

The Collections and Navigation endpoints are specified to return **LD+JSON** adhering to the **W3C Hydra** standard. The Documents endpoint is specified to return **TEI/XML** of the requested text or fragment.

(<https://w3id.org/dts>)

DTS capabilities

The DTS API provides the following core capabilities to clients:

- Retrieve lists of collection members
- Retrieve metadata about individual collection items
- Retrieve lists of citeable passages within a text
- Retrieve lists of citeable passages within a text as groups of client-defined sizes (e.g. groups of 10 lines)
- Retrieve metadata about the citation structure of a document
- Retrieve a single text passage at any level of the citation hierarchy
- Retrieve a range of text passages with a clearly defined start and end passage
- Retrieve an entire text

(<https://w3id.org/dts>)

A minimal computing implementation

- Create static TEI-XML (for documents endpoint) and JSON-LD (for navigation) file structure for all (sensible) parameter combinations
 - <https://github.com/robcast/simple-tei2dtsflat>
- Create minimal server mapping API parameters to static file access
 - <https://github.com/robcast/dtsflat-server>

DTS navigation endpoint

Query Parameters

name	description	methods
id	the unique identifier (normally a URN) for the Resource being navigated	GET
ref	(NOT used with <code>start</code> and <code>end</code>) a single passage identifier providing the point of reference for the Navigation request within the Resource. Such identifiers should be unique within a given Resource.	GET
start	(NOT used if a <code>ref</code> is specified, requires <code>end</code> as well) Start of the range passages to serve as the reference point for the Navigation request. This parameter is inclusive, so the supplied reference is considered part of the specified range.	GET
end	(NOT used if a <code>ref</code> is specified, requires <code>start</code> as well) End of the range of passages to serve as the reference point for the Navigation request. This parameter is inclusive, so the supplied reference is considered part of the specified range.	GET
level	the depth (as a number) for reference identifiers to be retrieved, relative to the specified <code>ref</code> or <code>start / end</code> values. <i>E.g.</i> , if a request should return the children of the passage "1.2", then the <code>ref</code> parameter should be "1.2" and the <code>level</code> parameter should be 2. This parameter does not indicate the absolute depth of the references in the Resource's citation tree.	GET
groupBy	Retrieve passages in groups of this size instead of single units. This would normally mean that the <code>member</code> list returned would be a list of ranges, each of which contains this number of passages.	GET
max	Allows for limiting the number of results and getting pagination	GET
exclude	Exclude keys in members' object such as <code>exclude=dts:extensions</code>	GET

DTS document endpoint

Name	Description	Methods
id	(Required) Identifier for a document. Where possible this should be a URI	GET, POST, PUT, DELETE
ref	Passage identifier (used together with <code>id</code> ; can't be used with <code>start</code> and <code>end</code>)	GET, PUT, DELETE
start	(For range) Start of a range of passages (can't be used with <code>ref</code>)	GET, PUT, DELETE
end	(For range) End of a range of passages (requires <code>start</code> and no <code>ref</code>)	GET, PUT, DELETE
after	(Optional) Passage after which the new segment should be inserted	POST
before	(Optional) Passage after which the new segment should be inserted	POST
token	(May be required by implementation) Authentication token for access control	POST, PUT, DELETE
format	(Optional) Specifies a data format for response/request body other than the default	GET, POST, PUT, DELETE

Simple tei2dtsflat

```
usage: tei2dtsflat.py [-h] [--version] [-l {INFO,DEBUG,ERROR}] [-b BASEDIR] [-i DOCID]
                    [--gen-id-prefix GENID_PREFIX] [-u URL_PREFIX] [--document-prefix DOC_PREFIX]
                    [--navigation-prefix NAV_PREFIX] [-m {div}]
                    inputfile
```

Create DTSflat file structure from TEI XML.

positional arguments:

inputfile TEI XML input file.

optional arguments:

-h, --help show this help message and exit
--version show program's version number and exit
-l {INFO,DEBUG,ERROR}, --log {INFO,DEBUG,ERROR} Log level.
-b BASEDIR, --base-dir BASEDIR DTSflat output base directory.
-i DOCID, --document-id DOCID DTS main document id (default: inputfile).
--gen-id-prefix GENID_PREFIX Prefix for generated xml-ids.
-u URL_PREFIX, --url-prefix URL_PREFIX DTS API base URL prefix.
--document-prefix DOC_PREFIX DTS document endpoint URL prefix (below base URL).
--navigation-prefix NAV_PREFIX DTS navigation endpoint URL prefix (below base URL).
-m {div}, --navigation-mode {div} Type of navigation structure: div=by tei:div.

TEI...

```
<div xml:id="echoid-div3" type="toc">
  <head rend="italics" xml:id="echoid-head6" xml:space="preserve">TRACTATVS QVI IN HOC <lb />volumine continentur.</head>
  <head xml:id="echoid-head7" xml:space="preserve">Theoremata Arithmetica.</head>
  <head xml:id="echoid-head8" xml:space="preserve">Derationibus operationum perfpctiuæ.</head>
  <head xml:id="echoid-head9" xml:space="preserve">De Mechanicis.</head>
  <head xml:id="echoid-head10" xml:space="preserve">Disputationes de quibusdam placitis Arift.</head>
  <head xml:id="echoid-head11" xml:space="preserve">In quintum Euclidis librum.</head>
  <head xml:id="echoid-head12" xml:space="preserve">Phyfica, & Mathematica refponfa per Epiftolas.</head>
  <pb facs="#p0007" />
</div>
<div xml:id="echoid-div4" type="dedication">
  <head xml:id="echoid-head13" xml:space="preserve">SERENISSIMO <lb />CAROLO EMANVELI <lb />Sabaudia Ducis, <choice><ex>&c.</ex><am>&c.</am></choice>
  <p rend="italics">
    <s xml:id="echoid-s1" xml:space="preserve"><hi rend="small caps italics">AGitvr</hi> nonusdecimus annus ex quo litte-<lb />ris Sereniſſimi patris
    <s xml:id="echoid-s2" xml:space="preserve">Is aduenientem tam humanè <lb />excepit, tanta deinde liberalitate fuit com-<lb />plexus ego viciſſim
    <s xml:id="echoid-s3" xml:space="preserve">Cuius in me benignitas, mea <lb />in illum obferuantia mirum in modum mutuo vfu, & confue-<lb />t
    <s xml:id="echoid-s4" xml:space="preserve">quo quidem tempo-<lb />re de Mathematicis ſcientijs mecum agebat, in quibus perdi-<lb />fcendis mea o
    <s xml:id="echoid-s5" xml:space="preserve">Cui vt quod in me eſet fatifacerem, acrius <lb />quàm antea ſtudia Cadaætamen ſemper fui propen
    <s xml:id="echoid-s6" xml:space="preserve"><choice><ex>Illiusque</ex><am>Illiusq̃</am></choice> imitatione (vt ferècæteri Principum <lb />ſtudia
    <s xml:id="echoid-s7" xml:space="preserve">Cùmque ego nunquam laborem amicorum cauſa defugerim, <lb />euenit vt post tot annorum curricula, mea
    <s xml:id="echoid-s8" xml:space="preserve">Quas, cùm rationibus in epiftola ſub-<lb />ſequenti allatis edere conſtituiſſem, non ſub cuiuſque alt
    <s xml:id="echoid-s9" xml:space="preserve">tum quòd patri debitum libellum filio reddere par erat, tum
    <pb facs="#p0008" />
    quòd in tuæ Celfitudine paternam in me fouendo, & augendo<unclear reason="illegible" /> <lb />benignit atem ineſſe ſemper ſum expertus, tum a
    <s xml:id="echoid-s10" xml:space="preserve">Acceſſit, quod ego ſemper in <lb />his dedic ationibus ſpectandum put qui, tuam Celfitudinem tunclea
    <s xml:id="echoid-s11" xml:space="preserve">Quare, & veterum Per-<lb />farum Regum gloriã æquauit, & nos veluti in ſpem certam <lb />fæ
    <s xml:id="echoid-s12" xml:space="preserve">Tua igitur celfi-<lb />tudo libellum tot ei nominibus debitum, <lb />ea qua ſolet humanitate accipe-
    <s xml:id="echoid-s13" xml:space="preserve">Deus tuas <lb />omnes cogitationes, <lb />& conatus ad <lb />fæliciſſi-<lb />mos <lb />ſemper exi
  </p>
  <pb facs="#p0009" />
</div>
```


DTSflat



DTSflat server

☰ README.md ✎

DTSflat-server

Minimal computing server for a [DTS API](#) from a DTSflat file structure.

The server supports the DTS Navigation Endpoint and Documents Endpoint.

The DTSflat files can be generated from a TEI XML file using <https://github.com/robcast/simple-tei2dtsflat>

The server uses a simple [Nginx](#) configuration to serve pre-generated files without additional active components.

Required

- [Docker](#)
- [docker-compose](#)

Alternatively you can use the Nginx config with an existing Nginx server without Docker.

Configuration

```
cp .env.template .env
```

Edit the `.env` file to add your server hostname in `VIRTUAL_HOST` and the path to the directory with the DTSflat files in `DTS_DATA_DIR`.

Run

```
docker-compose up -d
```

The DTS API will be accessible through your hostname:

- <http://your.host.name/dts/navigation/?id=your-document-id>
- <http://your.host.name/dts/documents/?id=your-document-id>

nginx/templates/default.conf.template

```
13     # base for all try_files
14     root /var/www/dts-data;
15
16     # map DTS documents endpoint to files
17     location ${API_BASE_PATH}${DOCUMENT_API_PATH} {
18         default_type text/xml;
19         # second try works with empty $arg_ref but fails with invalid $arg_ref
20         try_files /$arg_id/$arg_ref/tei-frag.xml /$arg_id$arg_ref/tei-full.xml =400;
21     }
22
23     # map DTS navigation endpoint to files
24     location ${API_BASE_PATH}${NAVIGATION_API_PATH} {
25         default_type application/json;
26         try_files /$arg_id$ref_path$level_path/dts-nav.json =400;
27     }
28 }
29
30 # map ref and level arguments to add a slash when set
31
32 map $arg_ref $ref_path {
33     default "";
34     ~(.+) /$1;
35 }
36
37 map $arg_level $level_path {
38     default "";
39     ~(.+) /$1;
40 }
```


My motivation

Hack Idea: Explore integration of IIIF APIs for images together with DTS API for text #18

Edit

New issue

Closed robcast opened this issue on Sep 14, 2021 · 14 comments



robcast commented on Sep 14, 2021



Proponent / Mentor

Robert Casties (@robcast)

Brief Description

The International Image Interoperability Framework (IIIF, <https://iiif.io>) is a set of specifications and a community revolutionizing access to images on the internet. IIIF viewers are used by many cultural heritage institutions to present scanned books and manuscripts but the integration of full texts e.g. for searchable overlays, text+image views or richer interactive editions is still rather weak. It would be great if the DTS API could fill this void and become the IIIF analog for text use in the internet.

I would like to look at the integration points of the IIIF API with respect to full-text and the DTS API with respect to images on the specification level, collect concrete use cases in both communities, evaluate problem areas in concrete application use, and develop extensions to the specifications if necessary. This could be mostly done at the specification/paper level but I would also like to look at concrete applications like full-text extensions to the Mirador IIIF viewer.

Coding skills needed

Interest in reading and discussing specification prose would be very welcome. Coding skills would not be needed unless somebody would offer to create a prototype using Mirador (Javascript/React) or other IIIF viewer.

Team members (if applicable)

The hack is open to any and all interested in bringing IIIF and DTS together.

Assignees

No one assigned

Labels

Hackathon

Projects

None yet

Milestone

No milestone

Linked pull requests

Successfully merging a pull request may close this issue.

None yet

Notifications

Customize

Unsubscribe

You're receiving notifications because you were mentioned.

4 participants

