

# Cite Exchange Format (CEX)

Christopher Blackwell, Furman University  
Thomas Köntges, University of Leipzig  
Neel Smith, The College of the Holy Cross

## Sharing text libraries and data collections for teaching and research in the humanities

Source text collections and other complex datasets can be very difficult to share and reuse, and especially difficult to aggregate and disaggregate.

CEX, CITE Exchange is a **plain-text, self-documenting, technology-agnostic format** for capturing citable texts, data collections, and arbitrary relationships among citable data at any level of granularity.

CEX is based on the CITE/CTS architecture and it positions itself as an alternative and complement to TEI XML and relational database schemas. Managing and sharing text collections can be

cumbersome, especially if you only want to share a collection of excerpts based on hundreds of individual TEI XML files.

When teaching text-heavy humanities disciplines, such as history, literature or classics, scholars are constantly faced with the problem of creating a source-text collections (that is, a corpus of excerpts of a bigger corpus that is deemed a representative sample able to answer a scholarly investigation), and the challenge of easily sharing this newly generated collection with students and colleagues. Based on current forms of data exchange, scholars and their students facing this task needed to have intimate knowledge of either database solutions like eXistDB or of API-calls.

CEX circumvents this problem by simplifying the format of exchanging texts and related objects following the OCHO2 principles laid out in the CITE/CTS architecture.

Likewise, data collections (coins, geo-spatial data, manuscript folios, etc.) are efficiently served intact by relational databases. Extracting subsets, sharing datasets in whole or in part, and aggregating disparate collections with schemas can be very difficult. CEX, as an exchange format, simplifies this.

## How a CEX File Works

- It is a plain-text file, with filename-suffix **.cex**.
- It may contain a header.
- It consists of **blocks** for containing different categories of data.

## Data Blocks

- Catalogs of Texts
- Text Content
- Descriptions of Data Collections
- Collection Data
- Extensions to Collections (e.g. Binary Image Collections)
- Relations among citable objects
- Data models formally specifying further aggregation of primitives

## Descripción general

<https://github.com/cite-architecture/cex>

### CEX Header

```
#!cexversion
3.0
#!citelibrary
name#Fragment from Herodotus' Histories, Book VIII on Papyrus Oxyrhynchus 2099, dated to
early 2nd century AD.
urn:urn:cite2:cex:fufolio.2018a:P0xy2099
license#CC Share Alike. For details, see more info.
```

### CEX Text Content Block

```
#!ctscatalog
urn:citationScheme#groupName#workTitle#versionLabel#exemplarLabel#online#lang
urn:cts:greekLit:tlg0012.tlg001.fuGrc:#book/line#Homeric Epic#Iliad#Greek. Allen, ed. FU
Greek ed. Creative Commons Attribution 3.0 License#true#grec
```

```
#!ctsdata
urn:cts:greekLit:tlg0012.tlg001.fuGrc:1.1#Μῆνιν ἔτειδε θεὰ Πηληϊάδεω ἀχιλῆος
urn:cts:greekLit:tlg0012.tlg001.fuGrc:1.2#οὐλομένην, ἥ μωρί 'Αχαιοῖς ἄλγε' ξέθκε,
urn:cts:greekLit:tlg0012.tlg001.fuGrc:1.3#πολλάδες δ' ιφθίμους ψυχᾶς Αἰδήι προίσαν
urn:cts:greekLit:tlg0012.tlg001.fuGrc:1.4#ἡρώων, αύτούς δε ἐλύτια τεῦχε κύνεσσιν
urn:cts:greekLit:tlg0012.tlg001.fuGrc:1.5#οινοιοί τε πάσι, Διός δ' ἔτελειτο
```

### CEX Collections Block

```
// List collections
#!citecollections
URN#Description#Labelling property#Ordering property#License
urn:cite2:hmt:msA.v1:#Pages of the Venetus A
manuscript#urn:cite2:hmt:msA.v1.label:#urn:cite2:hmt:msA.v1.sequence:#CC-attribution-share-alike

// Define a collection's citable typed properties
#!citeproperties
Property#Label#Type#Authority list
urn:cite2:hmt:msA.v1.urn:#URN#Cite2Urn#
urn:cite2:hmt:msA.v1.label:#Label#String#
urn:cite2:hmt:msA.v1.siglum:#Manuscript siglum#String#
urn:cite2:hmt:msA.v1.sequence:#Page sequence#Number#
urn:cite2:hmt:msA.v1.rv:#Recto or Verso#String#recto,verso

// Capture a collection's data
#!citedata
siglum#sequence#urn:rv#label
msA#urn:cite2:hmt:msA.v1:1r#recto#Marcianus Graecus Z. 454 (= 822) (Venetus A) folio 1r
msA#2#urn:cite2:hmt:msA.v1:1v#verso#Marcianus Graecus Z. 454 (= 822) (Venetus A) folio 1v
msA#3#urn:cite2:hmt:msA.v1:2r#recto#Marcianus Graecus Z. 454 (= 822) (Venetus A) folio 2r
msA#4#urn:cite2:hmt:msA.v1:2v#verso#Marcianus Graecus Z. 454 (= 822) (Venetus A) folio 2v
```

## Una guía para la arquitectura CITE

<https://cite-architecture.github.io>

"CEX is unsuited to large-scale projects."  
— Anonymous Reviewer

### Homer Multitext 2018d Data Release:

<https://github.com/homermultitext/hmt-archive/blob/master/releases-cex/hmt-2018d.cex>

File Size (CEX): 11 megabytes

Images: 1,649

Citable Passages of Text: 27,625

Data Objects: 19,536

Relations: 19,819

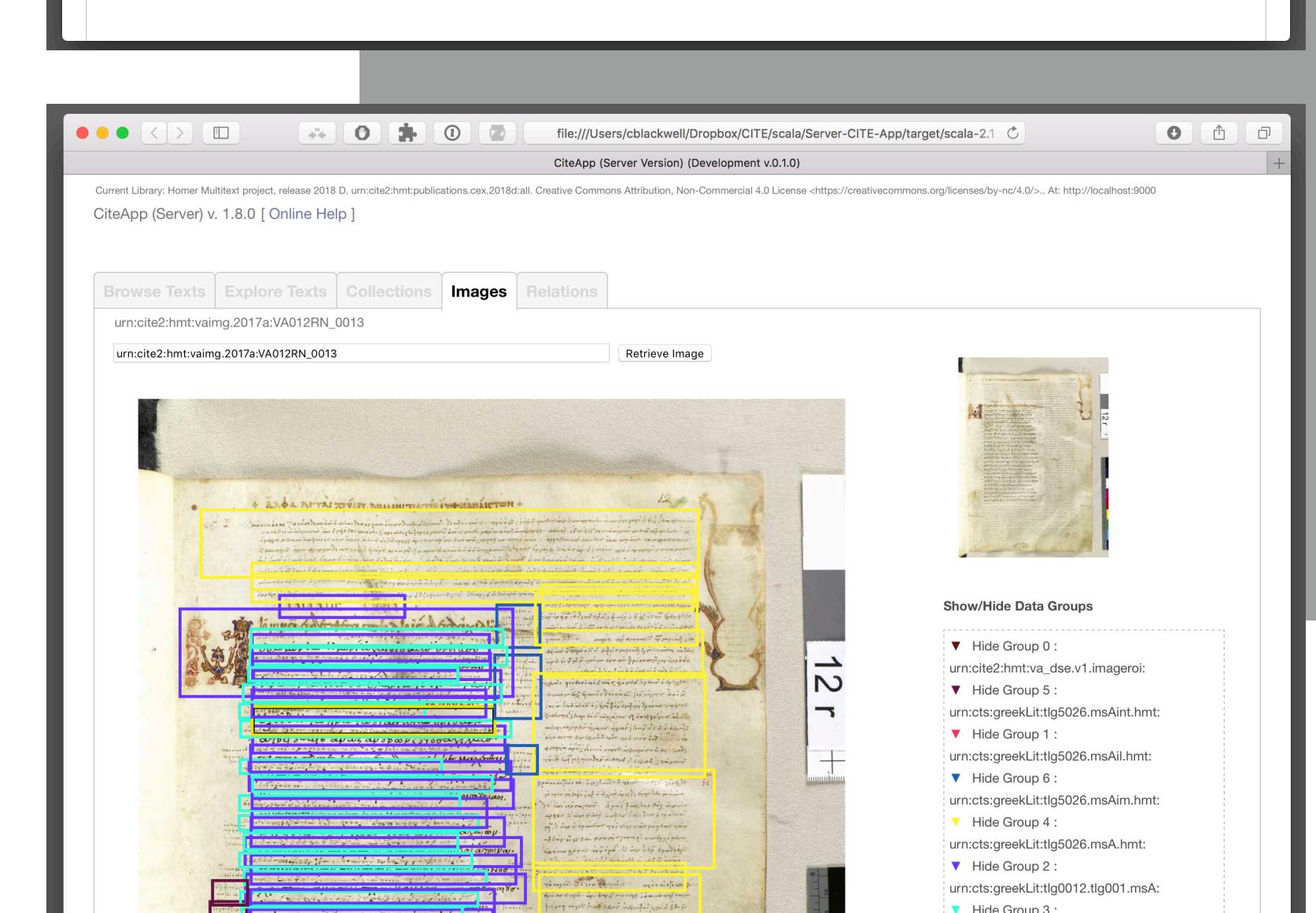
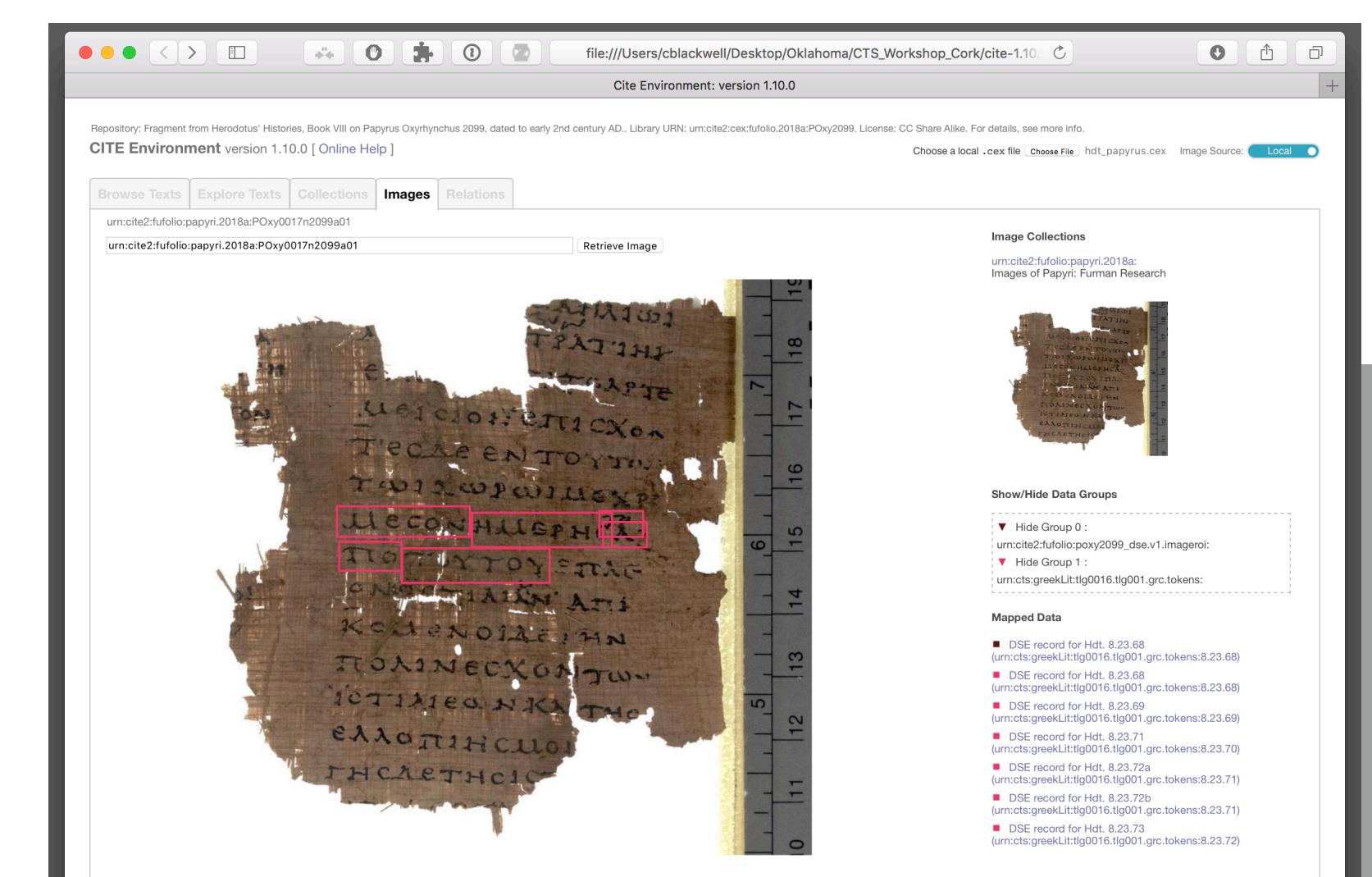
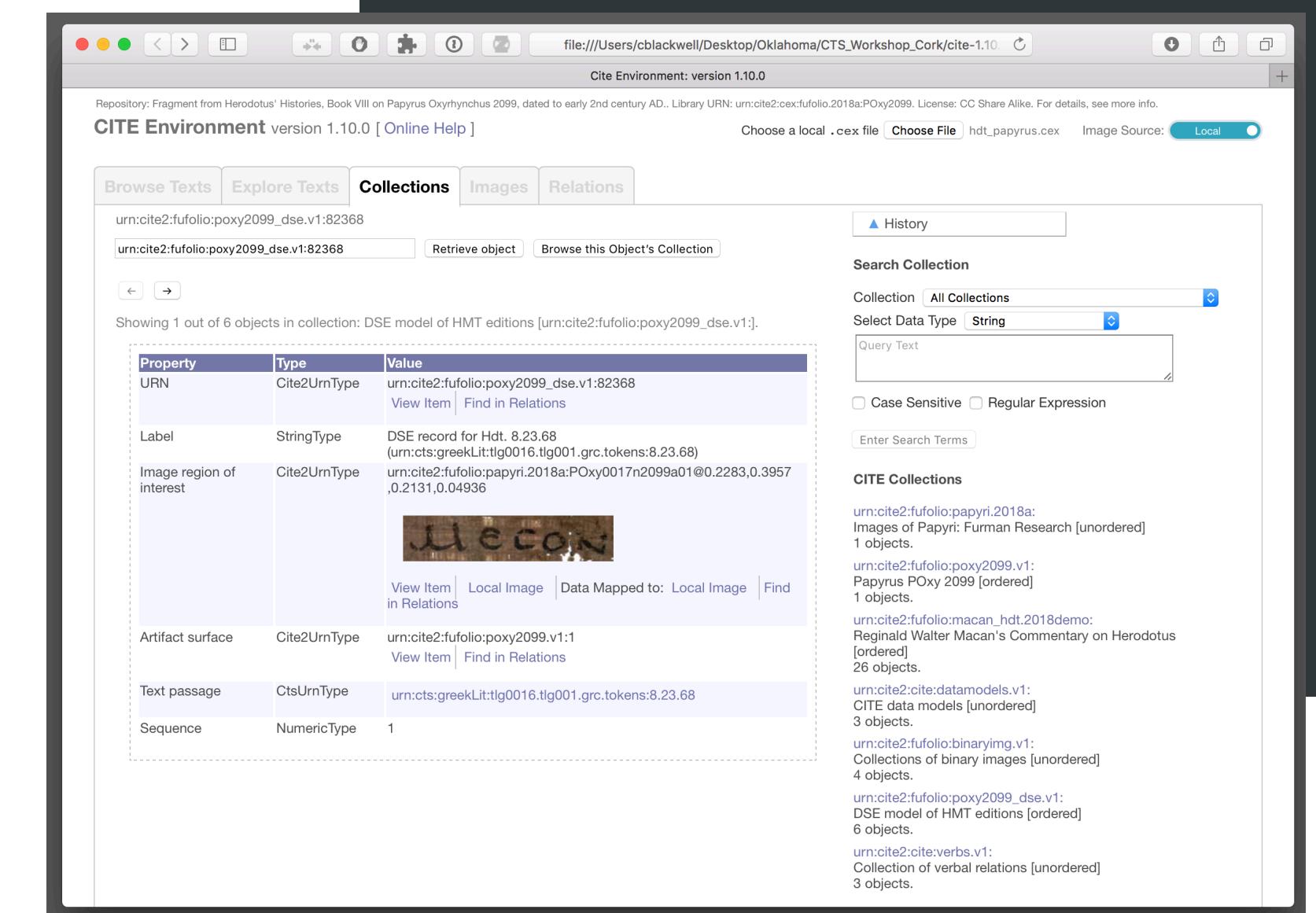
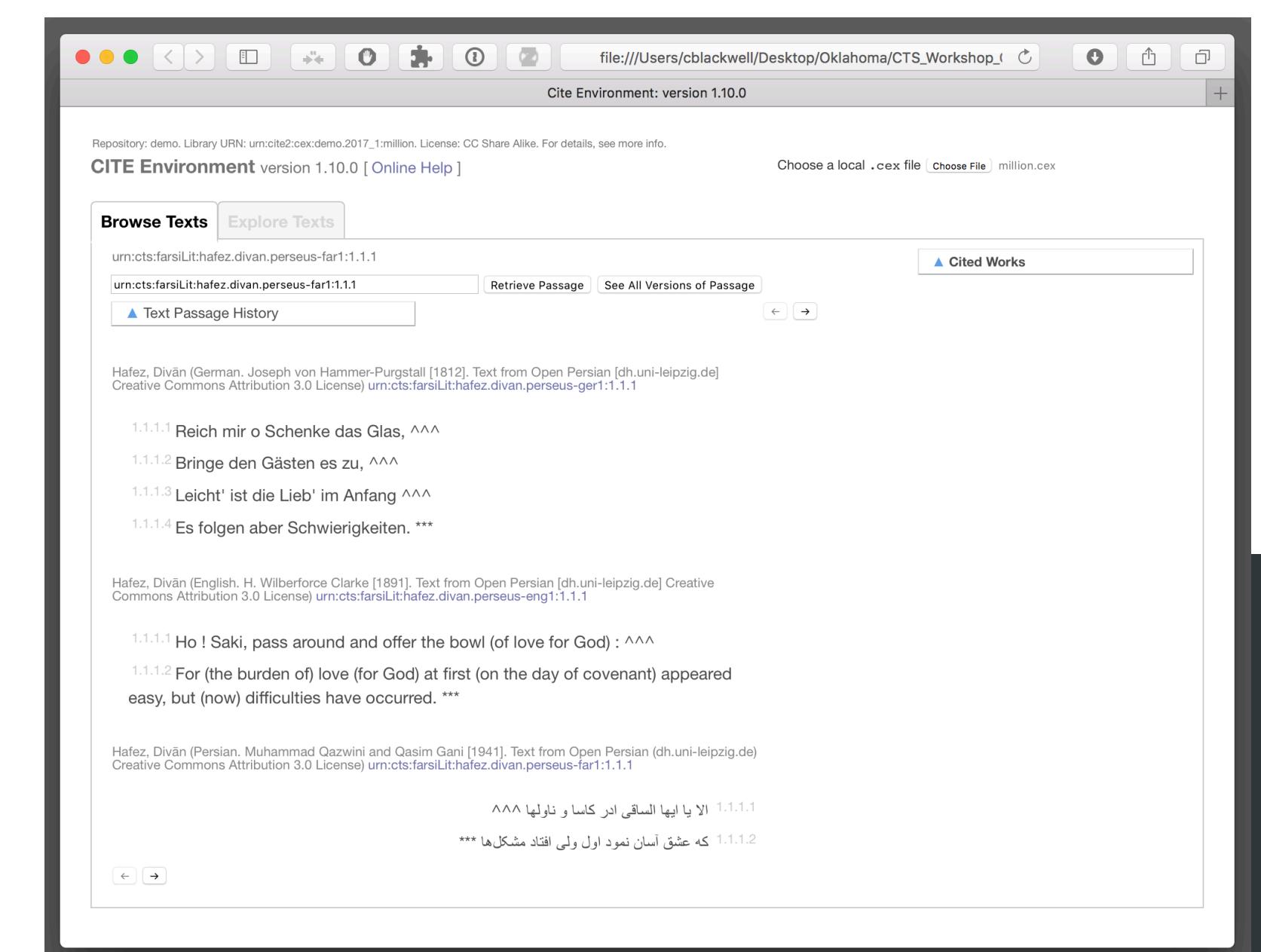
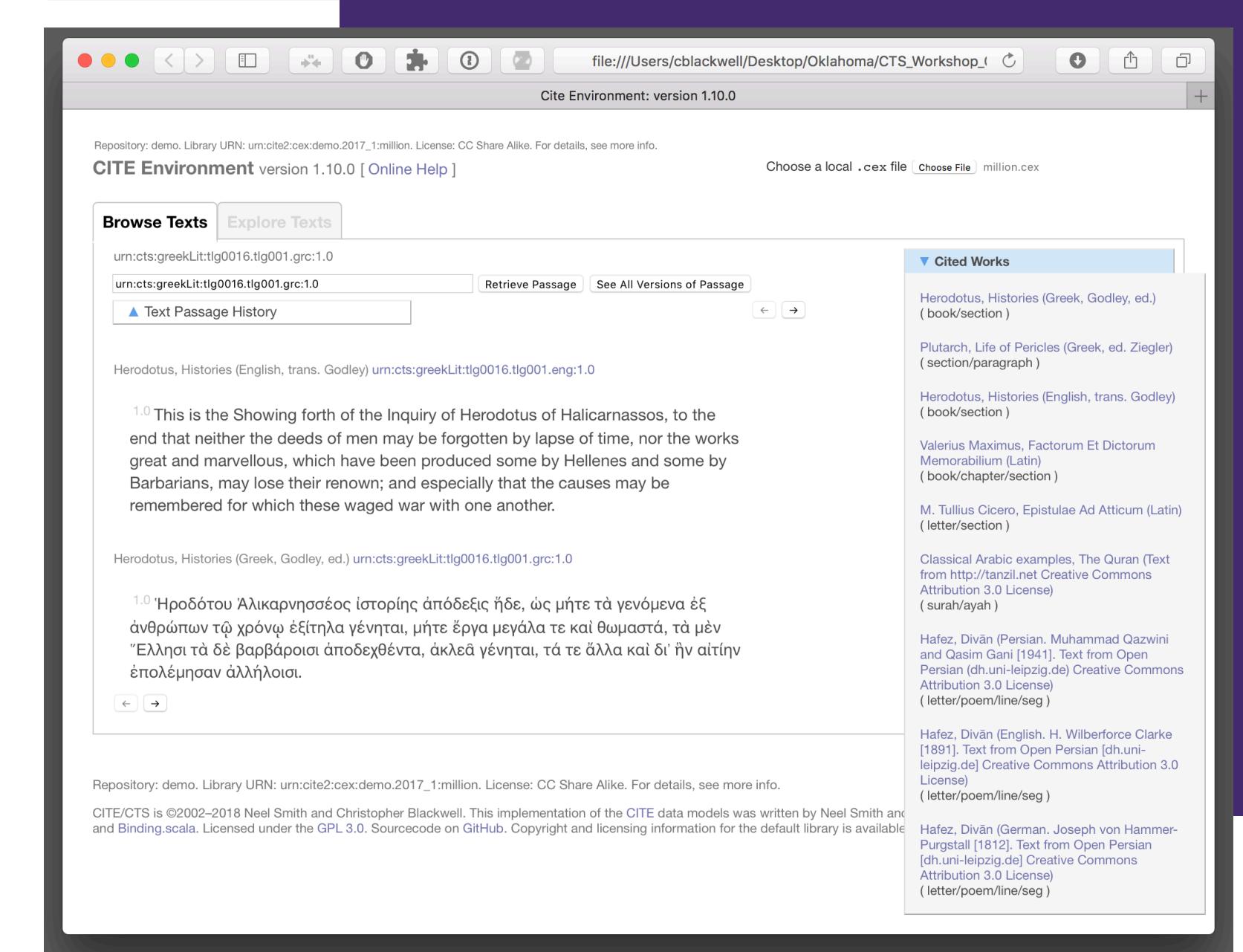
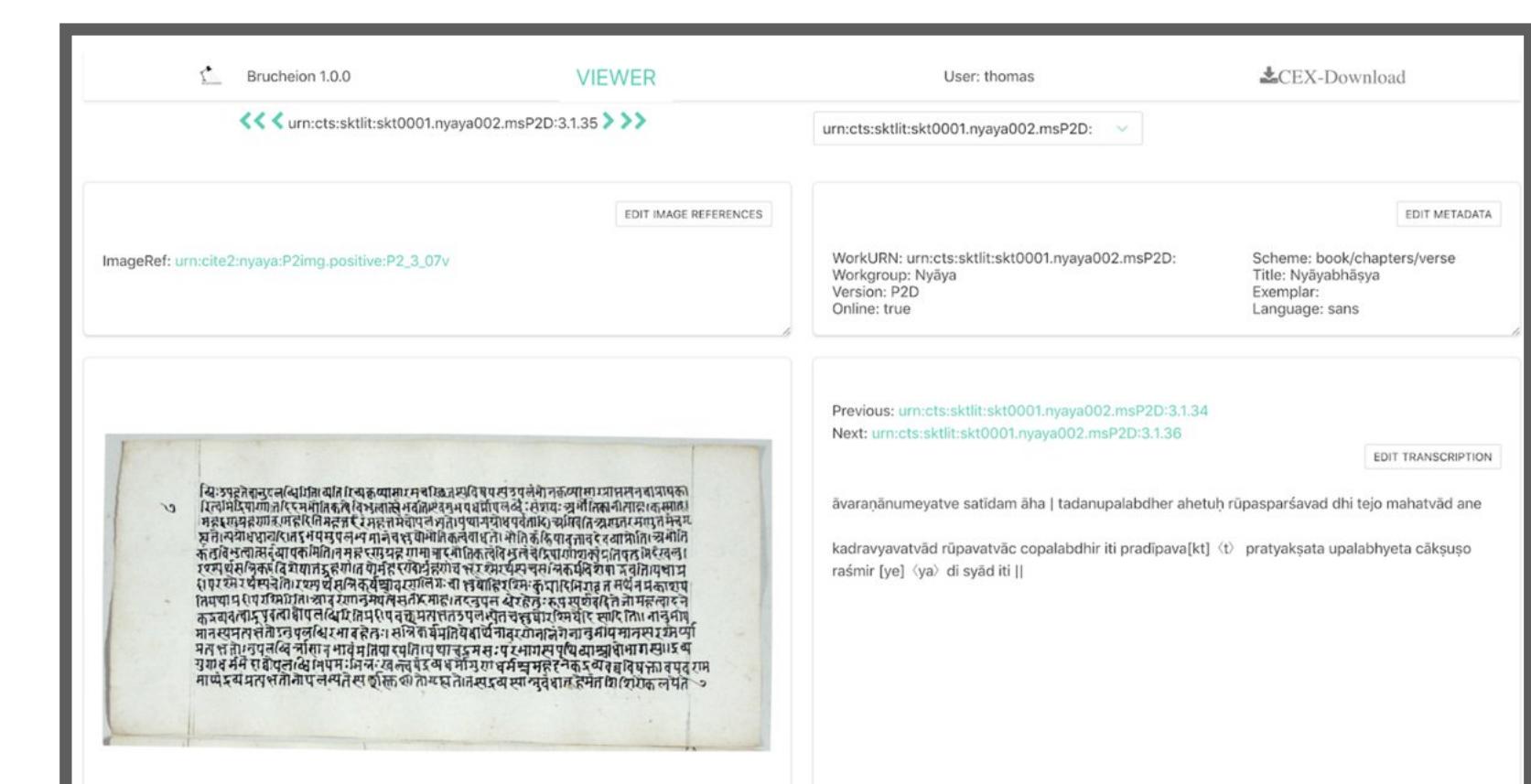
Data Served from CEX Source:

<http://beta.hpcc.uh.edu/hmt/hmt-microservice/>

<http://www.homermultitext.org/cite-app/>

Screen Captures from Brucheon (by Thomas Köntges) and CiteApp (by C. Blackwell and N. Smith), two digital library viewers for interacting with data in CEX format.

**Solicite un dispositivo de memoria USB gratis!**



## Tutoriales y demostraciones

[https://github.com/Eumaeus/dh2018\\_cex](https://github.com/Eumaeus/dh2018_cex)

