

Aligning HuCit with CIDOC-CRM and FRBRoo

Matteo Romanello^{1,2} Michele Pasin²

²Department of Digital Humanities
King's College London

¹Deutsches Archäologisches Institut
Berlin

CIDOC-CRM SIG meeting
Amsterdam
17 November 2011

Summary

Introduction

Hucit Ontology

Wrap-up

Some Context

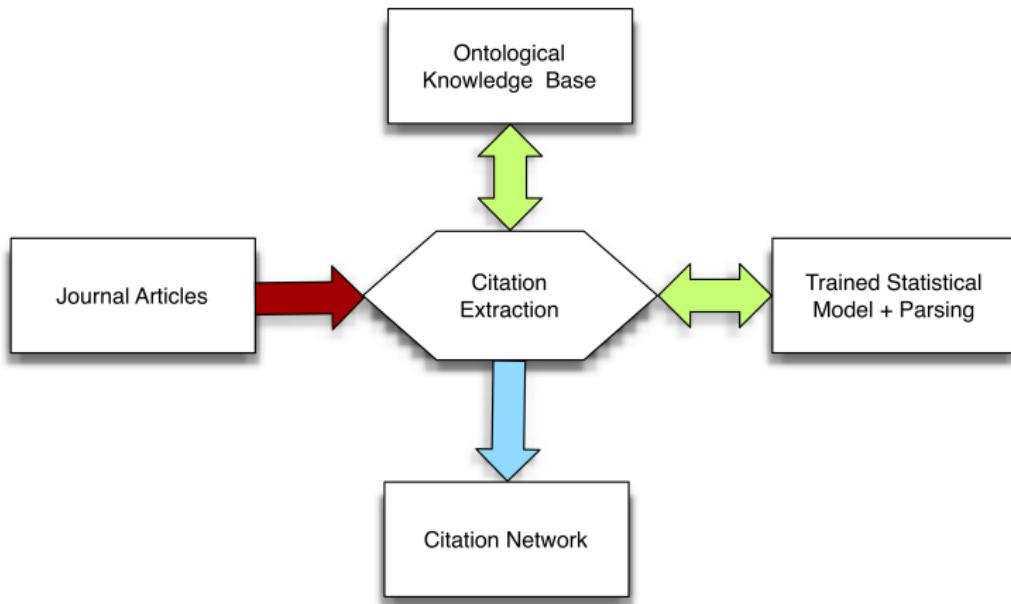
HuCit

- ▶ **Humanities Citation Ontology**
- ▶ DDH internal seminar on Knowledge Representation
- ▶ DH 2011 <http://dh2011abstracts.stanford.edu/xtf/view?docId=tei/ab-143.xml>
- ▶ **browse:** <http://www.michelepasin.org/demos/ontoview/?uri=https://purl.org/net/hucit>
- ▶ **code:** <https://bitbucket.org/56k/hucit>

My PhD at the DDH, KCL

- ▶ part of a domain ontology for Classics
 - ▶ ontology for a domain-specific Knowledge Base (KB)
 - ▶ KB supports tasks of information extraction

The Big(ger) Picture



Related Work

Ontology Engineering

- ▶ Guarino / Mizoguchi
- ▶ Semantic Publishing and Referencing ontologies (SPAR)
 - ▶ CiTO, Citation Typing Ontology
 - ▶ Fabio ontology
- ▶ SWAN Ecosystem, esp. Scientific Discourse Ontology
- ▶ Open Annotation Collaboration (OAC)
- ▶ OpenCyc

Canonical Citations

- ▶ Canonical Text Services (CTS), Harvard Univ.
- ▶ Classical Works Knowledge Base (CWKB), Cornell Univ.

What are Canonical Citations?

It is not, in general, permissible to alter a text to suit historical or geographical fact, unless there is something in the language itself which also demands a change : a principle which justifies, for example, the insertion of a negative in *Thuc. VIII. 101.* and the obelizing of *νεώτερος* in *Ar. Const. Ath. 24. 3.* But there are also cases in which the error is certain and the author most unlikely to have committed it; if both these conditions are present, we may suspect the text. It is generally agreed that these last two conditions are present in this passage of Herodotus ; but it has not been noticed (I believe) that the first is present also. Blakesley indeed accepted the text at its face value, and considered it a triumphant proof that the boasts of Athens about freeing the Greeks of Asia from the barbarian were all false—a fact that seems to have escaped her contemporary critics: and others have toyed with this explanation. But if the direct statement of Thucydides (VIII. 5.5) and the terms of the treaty between Persia and Sparta (VIII. 18) are not sufficient to refute it, consider these two facts: (i.) Erythrae had an Athenian garrison, from about 460, and the Athenian commander had some control even over her internal affairs ; she could not restore exiles hostile to Athens : did she pay tribute to Persia ? (ii.) In the quarrel between Miletus and Samos in 441, it is Miletus, the city on the mainland, that appeals to Athens ; and Pissuthnes and the Samian oligarchs prepare to march against her. Why, if the Persians were strong enough to exact tribute ? Macan argues that many new cities became tributary to Athens for the first time after Erymedon, and many were lost again and became tributary to Persia after the Egyptian disaster. But neither of these changes affected Ionia ; and if they had, it would make Herodotus' statement even more untrue, for he asserts that there was no change. Macan also adduces *Thuc. I. 138. 5.*, the gift by Xerxes to Themistocles of Magnesia, Myus, and Lampsacus. But note that it was only in Magnesia that he

Canonical Citations

Motivation

- ▶ refer to the very **objects of research**, i.e. texts
- ▶ **precision** of references to texts
- ▶ **persistency** of citations
- ▶ **interoperability** of citations across editions

Characteristics

- ▶ are not tight to a specific edition of a text (loosely coupled)
- ▶ canonical: based on agreement within a community of practice

Examples

- ▶ Hom. *Il.* 1.1
- ▶ Athen. *Deipn*, X 412a
- ▶ Arist. *Poetics* 1451a35-b6

Citations as Links

Thucydides 4.8.3-9

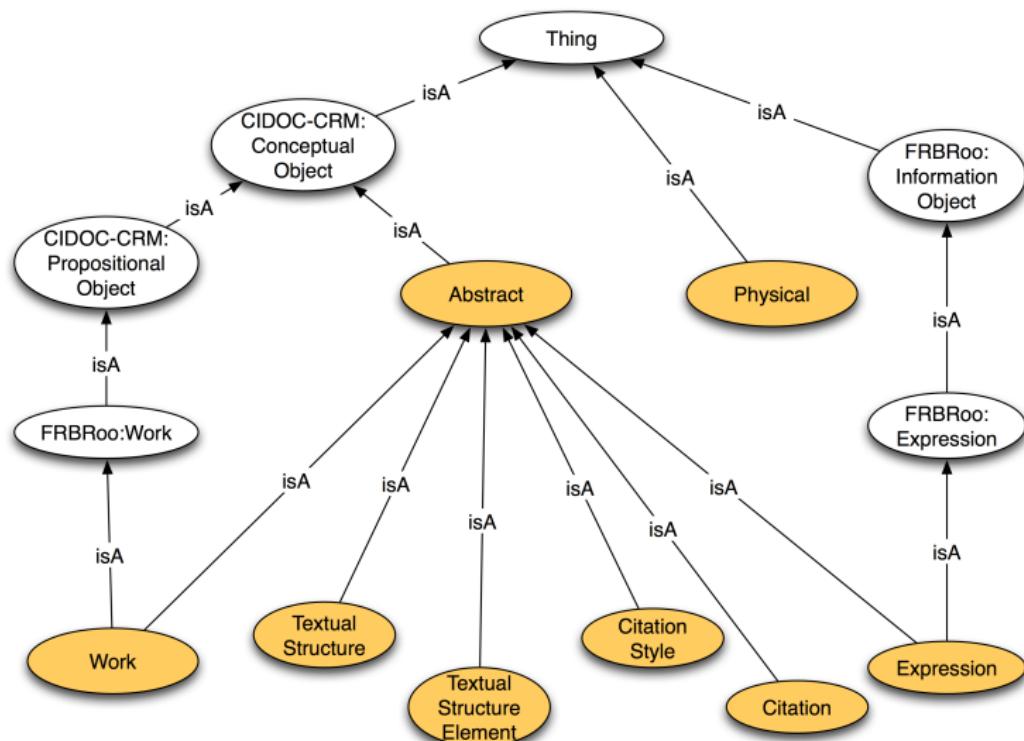
- ▶ Classical Works Knowledge Base (CWKB) – CornellU
- ▶ <http://cwkb.org/>
 - ▶ http://cwkb.org/resolver?ctx_ver=z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:canonical_cit&rft.workid=http://cwkb.org/workid/tlg:0003.001&rft.au=Thucydides&rft.slevel1=4&rft.slevel2=8&rft.slevel3=3&rft.elevel3=9&rfr_id=info:sid/aph%22
- ▶ Canonical Text Services (CTS) – HarvardU
- ▶ <http://chs75.chs.harvard.edu/projects/diginc/specs/cts>
 - ▶ <urn:cts:greekLit:tlg0003.tlg001:4.8.3-4.8.9>

Introduction

Hucit Ontology

Wrap-up

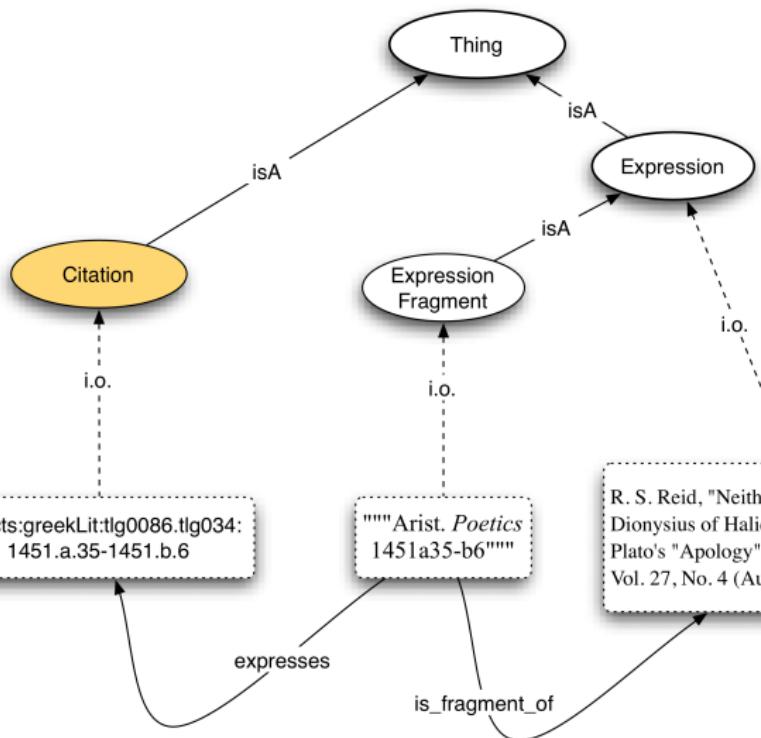
Hucit's Overview



What is a Citation?

- ▶ modern \neq canonical citations
- ▶ citation as first class object
 - ▶ **entity not property** => common approach: act of citing as “cites” property
 - ▶ the portion of text containing a citation is not directly captured
 - ▶ canonical citations have a structure to be captured
 - ▶ canonical citations have a *greater granularity*
- ▶ citation has
 1. **appearance (style)** => variation in the set of signs used to express the same citation
 2. **structure** => a structure that allows one to “parse” (read) a citation
 3. **content** => the *something* it refers to

Citation

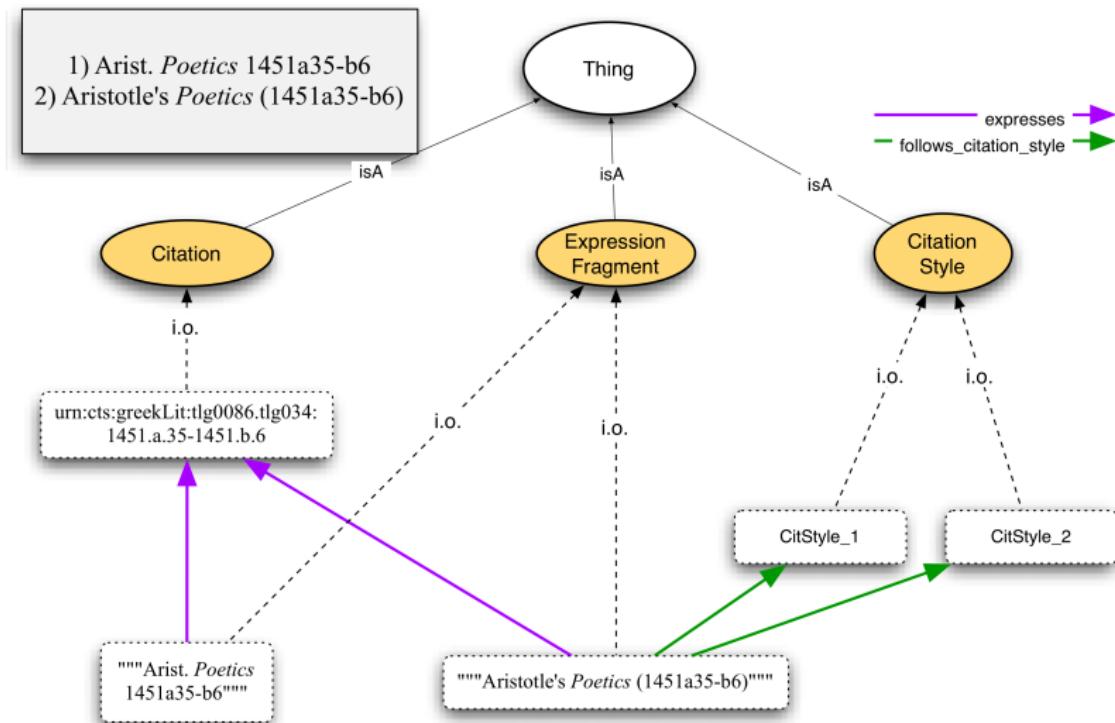


Citation Style

Factors causing variations in style

- ▶ publication venue
 - ▶ journal/collection level style (cf. APA, Harvard, Chicago for modern references)
 - ▶ language (Latin vs English abbreviations)
- ▶ audience specialisation
 - ▶ the higher the specialisation, the shorter the references
 - ▶ **Example:** Hom. II. 11 ≡ α 1
- ▶ diachronic variation
 - ▶ practices evolve
 - ▶ standardisation process

Citation Style



Citing Aristotle

Bekker numbers, the standard form of reference to works in the Corpus Aristotelicum, are based on the page numbers used in the [...] edition of the complete works of Aristotle [...]. They take their name from the editor of that edition, the classical philologist August Immanuel Bekker (1785-1871).[...]

All modern editions or translations of Aristotle intended for scholarly readers use Bekker numbers [...].

While Bekker numbers are the dominant method used to refer to the works of Aristotle, Catholic or Thomist scholars often use the medieval method of reference by book, chapter, and sentence [...]

http://en.wikipedia.org/wiki/Corpus_Aristotelicum#Bekker_numbers

Text/Citation Structure

Citing Aristotle

- ▶ according to medieval system
 - ▶ book / chapter / sentence
 - ▶ e.g. Arist. *Poetics* VIII 4 - IX 3
- ▶ according to Bekker numbers
 - ▶ page / column / line
 - ▶ e.g. Arist. *Poetics* 1451a35

Textual Structure

- ▶ Reference System
- ▶ abstract but derives from an Expression; superimposed on the text
- ▶ loose metaphor = geo coordinate system
- ▶ allows us precisely to *locate* a passage within a text

34

VIII. 3—IX. 5. 1451 a 27—1451 b 12

τῷ ἀγερμῷ, ὃν οὐδὲν θατέρον γενομένου ἀναγκαῖον ἦν
ἢ εἰκὸς θάτερον γενέσθαι, ἀλλὰ περὶ μίαν πρᾶξιν οἴων
λέγομεν τὴν Ὀδύσσειαν συνέπτησεν, ὅμοιος δὲ καὶ τὴν
30 Ἰλιάδα. χρὴ οὖν καθάπερ καὶ ἐν ταῖς ἀλλαῖς μυμητικαῖς ἡ μία 4
μύμησις ἐνός ἔστιν οὕτω καὶ τὸν μῦθον, ἐπεὶ πράξεως μύμησις
ἐστι, μᾶς τε εἶναι καὶ ταύτης ὅλης καὶ τὰ μέρη συνεστά-
ναι τοι τηραγμούν οὕτως ὥστε μετατιθεμένου τινὸς μέρους
ἢ ἀφαιρουμένου διαφέρεσθαι καὶ κινεῖσθαι τὸ δόλον. ὃ γὰρ
35 προστὸν ἡ μῆτρα προστὸν μηδὲν ποιεῖ ἐπιδήλον, οὐδὲν μόριον τοῦ
ὅλου ἔστιν.

Φανερὸν δὲ ἐκ τῶν εἰρημένων καὶ ὅτι οὐ τὸ τὰ γενόμενα λέγει, τοῦτο ποιητοῦ ἔργου ἐστίν, ἀλλ' οὐλα ὡν γένοιτο καὶ τὰ δυνατὰ κατὰ τὸ εἰκὸς ἢ τὸ ἀναγκαῖον. ὁ γάρ 2 στορικὸς καὶ ὁ ποιητὴς οὐ τῷ ἢ ἔμμετρα λέγειν ἢ ἔμμετρα διαφέρουσιν (εἴη γάρ ἀν τὰ Ἡροδότου εἰς μέτρα τεθῆναι, αἱ οὐδὲν ἦπτοι ἀν εἴη ἴστορία τις μετὰ μέτρου ἢ ἀνεν μέτρων). Ἰλλὰ τούτῳ διαφέρει, τῷ τὸν μὲν τὰ γενόμενα λέγειν, 5 ὃν δὲ οὐλα ἀν γένοιτο. διὸ καὶ φιλοσοφώτερον καὶ 3 την πουδαίωτερον ποίησις ἴστορίας ἐστίν. ἡ μὲν γὰρ ποίησις μᾶλλον τὰ καθόλου, ἡ δὲ ἴστορία τὰ καθ' ἔκαστον λέγει. Ἱστιν δὲ καθόλου μέν, τῷ ποιή τὰ ποια ἀττα συμβαίνει 4 λέγειν ἡ πράττειν κατὰ τὸ εἰκὸς ἢ τὸ ἀναγκαῖον, οὐ στο- 10 ρικός γάρ εσται ἡ ποίησις ὄντωμα ἐπιτιθεμένη· τὸ δὲ καθ' ἔκαστον, τὸ Ἀλκιβιάδης ἔπραξεν ἢ τὸ ἔπαθεν. ἐπὶ μὲν οὖν τῇ 5 τωμοδίᾳ ἡδη τοῦτο δῆλον γέγονεν· συστήσατες γάρ τὸν

184

ΠΕΡΙ ΣΟΦΙΣΤΙΚΩΝ ΕΑΓΓΥΩΝ

ΦΥΣΙΚΗΣ ΑΚΡΟΑΣΕΩΣ A.

A. 700 m D. 16. November 1921.

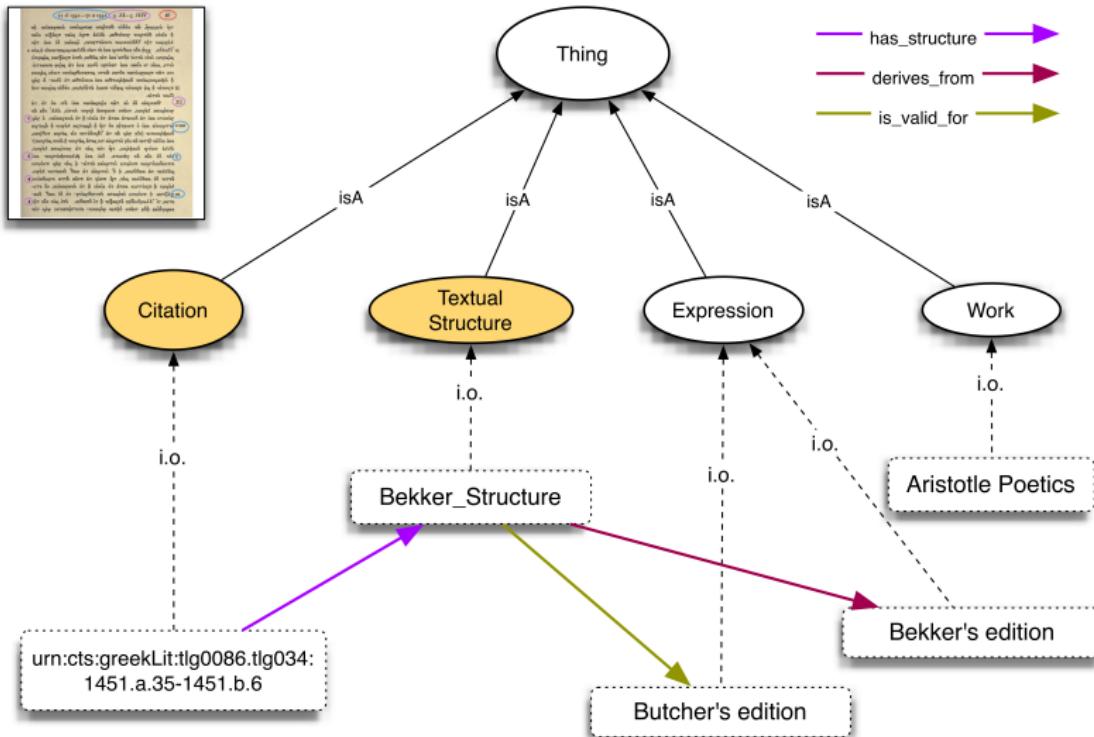
3. $\hat{\text{ba}}\hat{\text{r}} ABD$. 4. $\hat{\text{ba}}\hat{\text{r}} \hat{\text{a}} ABD$.

Geodäsie EFT. Τις διαρκείες αναφέρεται ως ηρή διάρκεια σαν οντότητα L .

2. \exists $j \in I$ et $\text{cor } E_j = 12$, si $\exists i \in I$ $\exists j < i$, tel que $E_i = 14$. Si $\exists i \in I$ $\exists j < i$, tel que $E_i = 16$, alors $FL_i = \text{just } I_i = 21$. Si $\exists i \in I$ $\exists j < i$, tel que $E_i = 23$, alors $FL_i = 26$. Si $\exists i \in I$

Butcher's edition (1922) p. 34 and Bekker's edition (1831) p. 184

Textual Structure and Expressions

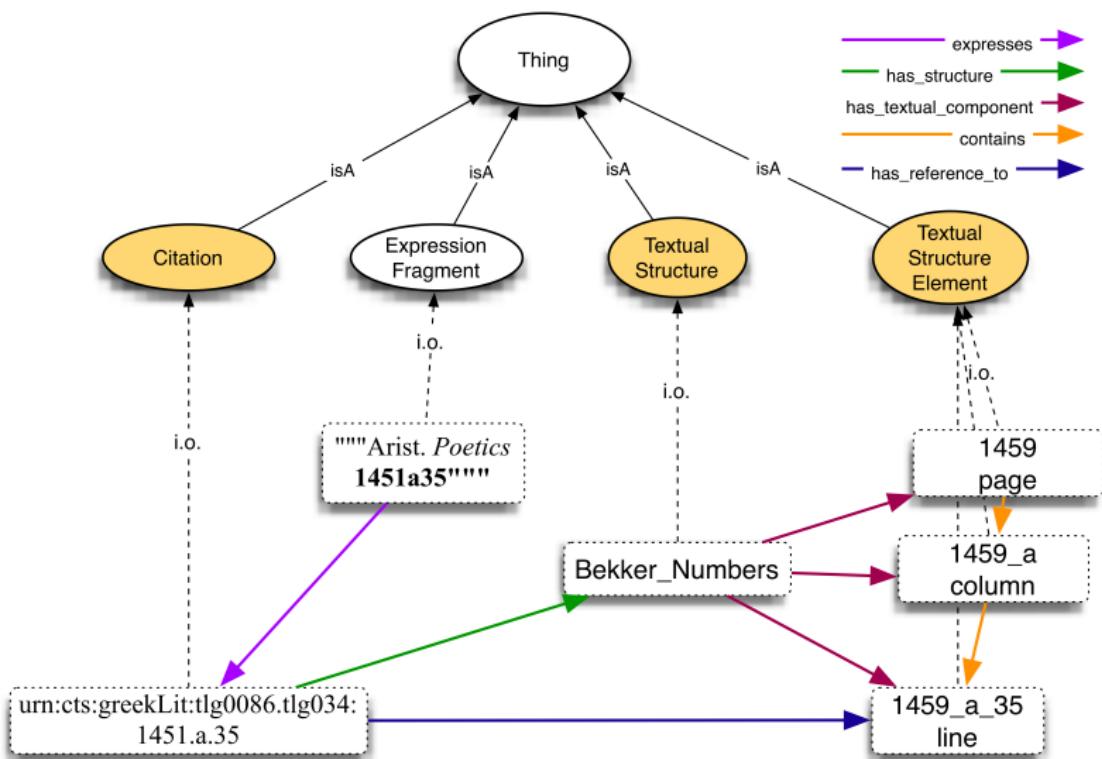


What is a Citation Pointing to?

A Citation

- ▶ is an abstract concept
- ▶ pointing to an element of the hierarchical structure of the text being cited
- ▶ and can be resolved into *many* passages of text, as many as the available expressions
- ▶ citations can follow different citation schemes (bekker, medieval)

Textual Structure as Citation Scheme



So Why?

Application-level

- ▶ enable/support *automatic parsing* of citations
- ▶ validate citations
- ▶ apply citations style to canonical citations, cf. Citation Style Language (CSL)
- ▶ extracting meaning + allow for reasoning

Modelling

- ▶ Provide a logical model of how citations in the humanities work
- ▶ Through ontologies => interoperability between different digital libraries and identifier schemes (CWKB, CTS)

Discussion

- ▶ a class for describing abstract structures in the CRM
- ▶ best way of hooking external ontologies,
CIDOC-CRM/FRBRoo
- ▶ which RDF namespace to use for FRBRoo?

Thanks for your attention!
Questions? Comments?

matteo.romanello@kcl.ac.uk
michele.pasin@kcl.ac.uk