WRITING FRAGIDS AN INTRODUCTION

Version: 2021-09-03

BACKGROUND RDF

~ Resource Description Framework

- ~ Resource Description Framework
- ~ Name: jargon, obscurity

- ~ Resource Description Framework
- ~ Name: jargon, obscurity
- ~ Essentially a framework for making claims or assertions

- ~ Resource Description Framework
- ~ Name: jargon, obscurity
- ~ Essentially a framework for making claims or assertions
- ~ Semantically interoperable

~ Graph structure

- ~ Graph structure
- ~ Every datum is a triple: subject, predicate (!), object

- ~ Graph structure
- ~ Every datum is a triple: subject, predicate (!), object
- ~ Subject: URI

- ~ Graph structure
- ~ Every datum is a triple: subject, predicate (!), object
- ~ Subject: URI
- ~ Predicate: URI

- ~ Graph structure
- ~ Every datum is a triple: subject, predicate (!), object
- ~ Subject: URI
- ~ Predicate: URI
- ~ Object: URI or literal (data)

- ~ Graph structure
- ~ Every datum is a triple: subject, predicate (!), object
- ~ Subject: URI
- ~ Predicate: URI
- ~ Object: URI or literal (data)
- ~w3.org/TR/rdf11-primer

BACKGROUND IRI/URI

~ International / Universal Resource Identifier

- ~ International / Universal Resource Identifier
- ~ RFC 3986-87

- ~ International / Universal Resource Identifier
- ~ RFC 3986-87
- ~ Superset of URNs, URLs

- ~ International / Universal Resource Identifier
- ~ RFC 3986-87
- ~ Superset of URNs, URLs
- ~scheme ":" hier-part ["?" query] ["#"
 fragment]

URI EXAMPLES

```
~ foo://example.com:8042/over/there?name=ferret#nose
~ ftp://ftp.is.co.za/rfc/rfc1808.txt
~ http://www.ietf.org/rfc/rfc2396.txt
~ ldap://[2001:db8::7]/c=GB?objectClass?one
~ mailto:John.Doe@example.com
~ news:comp.infosystems.www.servers.unix
~ tel:+1-816-555-1212
~ telnet://192.0.2.16:80/
~ urn:oasis:names:specification:docbook:dtd:xml:4.1.2
```

~ Persistence

- ~ Persistence
- ~ Uniqueness

- ~ Persistence
- ~ Uniqueness
- ~ Interoperability

BACKGROUND TEXTS

We refer to texts

We refer to texts either abstractly

We refer to texts either abstractly or concretely

We refer to texts either abstractly or concretely

"The Iliad says..."

We refer to texts either abstractly or concretely

"The *Iliad* says..."

"This particular edition of the *Iliad* says..."

Conceptual

Material

Conceptual

Material

terms

	Conceptual	Material
terms	work version	

	Conceptual	Material
terms	work	text bearing object
	version	= scriptum

	Conceptual	Material
terms	work	text bearing object
	version	= scriptum

examples

TEXTS

	Conceptual	Material
terms	work	text bearing object
	version	= scriptum
	 The Iliad Alexander's translation of The Iliad Frankenstein 	
examples	 Frankenstein Mel Brooks's Young Frankenstein	

TEXTS

	Conceptual	Material
terms	work	text bearing object
	version	= scriptum
examples	 The Iliad Alexander's translation of The Iliad Frankenstein Mel Brooks's Young Frankenstein 	 this particular manuscript (one-item set) this book (many-item set) this digital file (many-item set)

THE TASK WRITE RDF TRIPLES ABOUT TEXTS

~ Non-digital texts

- ~ Non-digital texts
- ~ Works and scripta

- ~ Non-digital texts
- ~ Works and scripta
- ~ Tokens, characters

- ~ Non-digital texts
- ~ Works and scripta
- ~ Tokens, characters
- ~ Discontinuous phrases

- ~ Non-digital texts
- ~ Works and scripta
- ~ Tokens, characters
- ~ Discontinuous phrases
- ~ RDF claims must generally applicable

WHAT WE CAN DO NOW

Shakespeare wrote Henry V.

Shakespeare wrote Henry V.

Plato's Republic quotes from Homer's Iliad.

Plato's Republic quotes from Homer's Iliad.

```
@prefix db: <http://dbpedia.org/resource/> .
@prefix cito: <http://purl.org/spar/cito/> .
db:Republic_(Plato) cito:cites db:Iliad
```

Plato's Republic quotes from Homer's Iliad.

```
@prefix db: <http://dbpedia.org/resource/> .
@prefix cito: <http://purl.org/spar/cito/> .
db:Republic_(Plato) cito:cites db:Iliad
```

Thanks a lot.

DESIDERATUM: SPECIFICITY

We already create and consume

We already create and consume specific, human-readable references

We already create and consume specific, human-readable references that are persistent, unique, interoperable.

We already create and consume specific, human-readable references that are persistent, unique, interoperable. How to convert this convention to URIs?

In Shakespeare's *Henry VI*, part 2, 1.4.32, 'Henry' is the grammatical object.

In Shakespeare's *Henry VI*, part 2, 1.4.32, 'Henry' is the grammatical object.

"The duke yet lives that Henry shall depose"

In Shakespeare's *Henry VI*, part 2, 1.4.32, 'Henry' is the grammatical object.

"The duke yet lives that Henry shall depose"

```
Oprefix db: <http://dbpedia.org/resource/> .
Oprefix *la: <http://example.org/linguistic-annotation/> .
Oprefix olia: <http://purl.org/olia/olia.owl#> .

db:Henry_VI,_Part_2 *la:hasFeature olia:DirectObject
```

Plato Resp. 328e6 quotes Il. 6.211

Plato Resp. 328e6 quotes Il. 6.211

"the threshold of old age"

Plato Resp. 328e6 quotes Il. 6.211

"the threshold of old age"

```
@prefix db: <http://dbpedia.org/resource/> .
@prefix cito: <http://purl.org/spar/cito/> .
db:Republic_(Plato) cito:cites db:Iliad
```

At Psalm 31.2 (31.1; LXX 30.2) הַטֵּה is translated κλῖνον by the Septuagint

At Psalm 31.2 (31.1; LXX 30.2) הַטֵּה is translated κλῖνον by the Septuagint

At Psalm 31.2 (31.1; LXX 30.2) הַטֵּה is translated κλῖνον by the Septuagint

"incline"

@prefix saws: <http://purl.org/saws/ontology#> .

?????? saws:isDirectTranslationOf ??????

~ URI for a specific portion of text?

- ~ URI for a specific portion of text?
- ~ Texts with multiple reference systems

- ~ URI for a specific portion of text?
- ~ Texts with multiple reference systems
- ~ Texts with ambiguous reference systems

CHALLENGES

- ~ URI for a specific portion of text?
- ~ Texts with multiple reference systems
- ~ Texts with ambiguous reference systems
- ~ How to conceptualize/approach works?

PREVIOUS APPROACHES

GENERAL

GENERAL

Canonical Text Services (CTS)

GENERAL

- Canonical Text Services (CTS)
- ~???

~ Sharing Ancient WisdomS (SAWS) (CTS-inspired)

- ~ Sharing Ancient WisdomS (SAWS) (CTS-inspired)
- ~ Perseus (CTS-inspired)

- ~ Sharing Ancient WisdomS (SAWS) (CTS-inspired)
- ~ Perseus (CTS-inspired)
- ~ Distributed Text Services (various)

~ URN schemes not registered with IANA

- ~ URN schemes not registered with IANA
- ~ (In praise of tag URNs.)

- ~ URN schemes not registered with IANA
- ~ (In praise of tag URNs.)
- ~ Administration of namespaces, subnamespaces

- ~ URN schemes not registered with IANA
- ~ (In praise of tag URNs.)
- ~ Administration of namespaces, subnamespaces
- ~ URIs exposed to collision (uniqueness in jeopardy)

- ~ URN schemes not registered with IANA
- ~ (In praise of tag URNs.)
- ~ Administration of namespaces, subnamespaces
- ~ URIs exposed to collision (uniqueness in jeopardy)
- ~ How to find URIs?

- ~ URN schemes not registered with IANA
- ~ (In praise of tag URNs.)
- ~ Administration of namespaces, subnamespaces
- ~ URIs exposed to collision (uniqueness in jeopardy)
- ~ How to find URIs?
- ~ Domain model lacking

- ~ URN schemes not registered with IANA
- ~ (In praise of tag URNs.)
- ~ Administration of namespaces, subnamespaces
- ~ URIs exposed to collision (uniqueness in jeopardy)
- ~ How to find URIs?
- ~ Domain model lacking
- ~ Works tethered to messy, inconsistent catalogs

- ~ URN schemes not registered with IANA
- ~ (In praise of tag URNs.)
- ~ Administration of namespaces, subnamespaces
- ~ URIs exposed to collision (uniqueness in jeopardy)
- ~ How to find URIs?
- ~ Domain model lacking
- ~ Works tethered to messy, inconsistent catalogs
- ~ "Canonical" references?? (uniqueness in jeopardy)

- ~ URN schemes not registered with IANA
- ~ (In praise of tag URNs.)
- ~ Administration of namespaces, subnamespaces
- ~ URIs exposed to collision (uniqueness in jeopardy)
- ~ How to find URIs?
- ~ Domain model lacking
- ~ Works tethered to messy, inconsistent catalogs
- ~ "Canonical" references?? (uniqueness in jeopardy)
- ~ Dependent upon specific digital corpora (persistence in jeopardy)

STEP BACK

~ Don't invent a new URI system

- ~ Don't invent a new URI system
- ~ Coopt URIs already in use

- ~ Don't invent a new URI system
- ~ Coopt URIs already in use
- ~ Adopt the persective of the URI writer, not the data provider/curator

- ~ Don't invent a new URI system
- ~ Coopt URIs already in use
- ~ Adopt the persective of the URI writer, not the data provider/curator
- ~ Stipulate in detail a domain model (what are texts?)

- ~ Don't invent a new URI system
- ~ Coopt URIs already in use
- ~ Adopt the persective of the URI writer, not the data provider/curator
- ~ Stipulate in detail a domain model (what are texts?)
- ~ Start with references that are tractable

- ~ Don't invent a new URI system
- ~ Coopt URIs already in use
- ~ Adopt the persective of the URI writer, not the data provider/curator
- ~ Stipulate in detail a domain model (what are texts?)
- ~ Start with references that are tractable
- ~ Define a rigorous syntax

A PROPOSAL WRITING FRAGIDS

~ Fragment identifier, e.g., http://example.org/index.html#hello

- ~ Fragment identifier, e.g.,
- http://example.org/index.html#hello
- ~ W3C Draft: Best Practices for Fragment Identifiers and
- Media Type Definitions

- ~ Fragment identifier, e.g.,
- http://example.org/index.html#hello
- ~ W3C Draft: Best Practices for Fragment Identifiers and Media Type Definitions
- ~ Most (all?) fragid schemes target digital media...

- ~ Fragment identifier, e.g., http://example.org/index.html#hello
- ~ W3C Draft: Best Practices for Fragment Identifiers and Media Type Definitions
- ~ Most (all?) fragid schemes target digital media...
- ~...but no reason they couldn't be applied to the nondigital world

THE BASIC IDEA

THE BASIC IDEA

Start with a base URI for a work or scriptum.

THE BASIC IDEA

Start with a base URI for a work or scriptum. Specify how you're interpreting the URI.

THE BASIC IDEA

Start with a base URI for a work or scriptum. Specify how you're interpreting the URI. Declare the reference system.

THE BASIC IDEA

Start with a base URI for a work or scriptum. Specify how you're interpreting the URI.

Declare the reference system.

Drill down with the reference system.

THE BASIC IDEA

Start with a base URI for a work or scriptum.

Specify how you're interpreting the URI.

Declare the reference system.

Drill down with the reference system.

Qualify with tokens/characters.

EXAMPLES

http://www.worldcat.org/oclc/860129739#\$wf0:a=w;t=m;r=.;14:2\$

1. \$wf0: — version number

- 1. \$wf0: version number
- 2. a=w; treat the base URI as a work (not a scriptum)

- 1. \$wf0: version number
- 2. a=w; treat the base URI as a work (not a scriptum)
- 3. t=m; ref. system type is material (not logical)

- 1. \$wf0: version number
- 2. a=w; treat the base URI as a work (not a scriptum)
- 3. t=m; ref. system type is material (not logical)
- 4. r=.; the reference scriptum is the base URI, qua scriptum

- 1. \$wf0: version number
- 2. a=w; treat the base URI as a work (not a scriptum)
- 3. t=m; ref. system type is material (not logical)
- 4. r=.; the reference scriptum is the base URI, qua scriptum
- 5. **14:2** material unit 14, subunit 2

http://dbpedia.org/resource/Republic_(Plato)#\$wf0:a=w;t=m;r=http://www.worldcat.org/oclc/1688

1. \$wf0: — version number

- 1. \$wf0: version number
- 2. a=w; treat the base URI as a work (not a scriptum)

- 1. \$wf0: version number
- 2. a=w; treat the base URI as a work (not a scriptum)
- 3. t=m; ref. system type is material (not logical)

- 1. \$wf0: version number
- 2. a=w; treat the base URI as a work (not a scriptum)
- 3. t=m; ref. system type is material (not logical)
- 4. r=http://www.worldcat.org/oclc/1688842; the reference scriptum is Stephanus's edition

- 1. \$wf0: version number
- 2. a=w; treat the base URI as a work (not a scriptum)
- 3. t=m; ref. system type is material (not logical)
- 4. r=http://www.worldcat.org/oclc/1688842; the reference scriptum is Stephanus's edition
- 5. 2.328.5.6 material units 2 (volume), 328 (page), 5 (subpage), 6 (line)

- 1. \$wf0: version number
- 2. a=w; treat the base URI as a work (not a scriptum)
- 3. t=m; ref. system type is material (not logical)
- 4. r=http://www.worldcat.org/oclc/1688842; the reference scriptum is Stephanus's edition
- 5. 2.328.5.6 material units 2 (volume), 328 (page), 5 (subpage), 6 (line)
- Problems: volume number, line discrepancies, modern editions with different material

http://dbpedia.org/resource/Henry_VI,_Part_2#\$wf0:a=w;t=1;r=http://www.worldcat.org/oclc/9352

1. \$wf0: — version number

- 1. \$wf0: version number
- 2. a=w; treat the base URI as a work (not a scriptum)

- 1. \$wf0: version number
- 2. a=w; treat the base URI as a work (not a scriptum)
- 3. t=1; ref. system type is logical (not material)

- 1. \$wf0: version number
- 2. a=w; treat the base URI as a work (not a scriptum)
- 3. t=1; ref. system type is logical (not material)
- 4. r=http://www.worldcat.org/oclc/935285574; the reference scriptum is Warren's edition

- 1. \$wf0: version number
- 2. a=w; treat the base URI as a work (not a scriptum)
- 3. t=1; ref. system type is logical (not material)
- 4. r=http://www.worldcat.org/oclc/935285574; the reference scriptum is Warren's edition
- 5. 1.4.32 logical units 1 (act), 4 (scene), 32 (line)

- 1. \$wf0: version number
- 2. a=w; treat the base URI as a work (not a scriptum)
- 3. t=1; ref. system type is logical (not material)
- 4. r=http://www.worldcat.org/oclc/935285574; the reference scriptum is Warren's edition
- 5. 1. 4. 32 logical units 1 (act), 4 (scene), 32 (line)
- 6. :: Henry[1] first instance of 'Henry'

- 1. \$wf0: version number
- 2. a=w; treat the base URI as a work (not a scriptum)
- 3. t=1; ref. system type is logical (not material)
- 4. r=http://www.worldcat.org/oclc/935285574; the reference scriptum is Warren's edition
- 5. 1.4.32 logical units 1 (act), 4 (scene), 32 (line)
- 6. ::Henry[1] first instance of 'Henry'
 Problems: unqualified works cannot take tokens

WF DESIGN

~ Texts: works or scripta

- ~ Texts: works or scripta
- ~ Works
 - work-versions
 - conceptual
 - more like Bibframe than FRBR
 - can nest, overlap

- ~ Texts: works or scripta
- ~ Works
 - work-versions
 - conceptual
 - more like Bibframe than FRBR
 - can nest, overlap
- ~ Scripta
 - class of material objects
 - may be constrained

- ~ Texts: works or scripta
- ~ Works
 - work-versions
 - conceptual
 - more like Bibframe than FRBR
 - can nest, overlap
- ~ Scripta
 - class of material objects
 - may be constrained
- ~ Scriptum readers

- ~ Texts: works or scripta
- ~ Works
 - work-versions
 - conceptual
 - more like Bibframe than FRBR
 - can nest, overlap
- ~ Scripta
 - class of material objects
 - may be constrained
- ~ Scriptum readers
- ~ Text divisions: material or logical

- ~ Texts: works or scripta
- ~ Works
 - work-versions
 - conceptual
 - more like Bibframe than FRBR
 - can nest, overlap
- ~ Scripta
 - class of material objects
 - may be constrained
- ~ Scriptum readers
- ~ Text divisions: material or logical

Everything above needs thoughtful, in-depth discussion

~Only scriptum readers should attempt to write WFs

- ~Only scriptum readers should attempt to write WFs
- ~Only uniquely enumerable divisions

- ~Only scriptum readers should attempt to write WFs
- ~ Only uniquely enumerable divisions
- ~Only writings with one material reference system, one material logical system

- ~Only scriptum readers should attempt to write WFs
- ~Only uniquely enumerable divisions
- ~Only writings with one material reference system, one material logical system
- ~ Prime candidates: modern articles, books

- ~Only scriptum readers should attempt to write WFs
- ~Only uniquely enumerable divisions
- ~Only writings with one material reference system, one material logical system
- ~ Prime candidates: modern articles, books

Everything above needs thoughtful, in-depth discussion

CHALLENGES

~ Too confusing, theorized?

- ~ Too confusing, theorized?
- ~ Epistemology: scriptum reader?

- ~ Too confusing, theorized?
- ~ Epistemology: scriptum reader?
- ~ Things with too many URIs (scripta)

- ~ Too confusing, theorized?
- ~ Epistemology: scriptum reader?
- ~ Things with too many URIs (scripta)
- ~ Things without any URIs (works, work-versions)

- ~ Too confusing, theorized?
- ~ Epistemology: scriptum reader?
- ~ Things with too many URIs (scripta)
- ~ Things without any URIs (works, work-versions)
- ~ Constraining a works to a work-version

- ~ Too confusing, theorized?
- ~ Epistemology: scriptum reader?
- ~ Things with too many URIs (scripta)
- ~ Things without any URIs (works, work-versions)
- ~ Constraining a works to a work-version
- ~ Syntax conflicts? Query constructor, other fragments.

- ~ Too confusing, theorized?
- ~ Epistemology: scriptum reader?
- ~ Things with too many URIs (scripta)
- ~ Things without any URIs (works, work-versions)
- ~ Constraining a works to a work-version
- ~ Syntax conflicts? Query constructor, other fragments.
- ~ End/footnotes?

- ~ Too confusing, theorized?
- ~ Epistemology: scriptum reader?
- ~ Things with too many URIs (scripta)
- ~ Things without any URIs (works, work-versions)
- ~ Constraining a works to a work-version
- ~ Syntax conflicts? Query constructor, other fragments.
- ~ End/footnotes?
- ~ Tokens based on spaces only?

- ~ Too confusing, theorized?
- ~ Epistemology: scriptum reader?
- ~ Things with too many URIs (scripta)
- ~ Things without any URIs (works, work-versions)
- ~ Constraining a works to a work-version
- ~ Syntax conflicts? Query constructor, other fragments.
- ~ End/footnotes?
- ~ Tokens based on spaces only?
- ~ Error handling

VIABILITY

Breakthrough technology?

Breakthrough technology?
Utter failure?

Breakthrough technology?
Utter failure?
Somewhere in-between?

Breakthrough technology?
Utter failure?
Somewhere in-between?

Failure can be success.

STATUS

~ Currently in conceptual phase

- ~ Currently in conceptual phase
- ~ Conversations will begin soon

- ~ Currently in conceptual phase
- ~ Conversations will begin soon
- ~ Draft specifications, examples, test suites

- ~ Currently in conceptual phase
- ~ Conversations will begin soon
- ~ Draft specifications, examples, test suites
- ~ Includes rudimentary demonstrations of WF-conformance for two media types (TEI, TAN)

- ~ Currently in conceptual phase
- ~ Conversations will begin soon
- ~ Draft specifications, examples, test suites
- ~ Includes rudimentary demonstrations of WF-conformance for two media types (TEI, TAN)
- ~ Part of the work of consolidating Linked Open Data for linguistic annotations

WRITING FRAGIDS JOIN THE CONVERSATION

Joel Kalvesmaki, kalvesmaki@gmail.com