Digital Hermeneutics for Digital and Non-Digital Humanists

Federico Boschetti
CoPhiLab, Istituto di Linguistica Computazionale “A. Zampolli”, CNR
& VeDPH, DSU, Ca’ Foscari University of Venice
federico.boschetti@ilc.cnr.it

Master in Digital Humanities, DSU - Università Ca’ Foscari

Venezia, 17th April 2020

Introduction

Plans and actions to fill the gap between Digital and Non-Digital Humanities in the next decade require a reflexion about

- Communities
- Methods
- Data
- Instruments
- Infrastructures
- Knowledge

Communities

DH Community and Subcommunities

The community of Digital Humanists is highly heterogeneous, because its members have

- different **backgrounds**
- different **goals**
- different **methodologies**

We can identify at least three subcommunities:

- Humanists (What?)
- Computer scientists, Computer Engineers, Developers (How?)
- DH natives (What and How, in many flavors)

Humanists

Academic humanists are **specialists** in the subfields of the Humanities

Specialization guarantees the advancement of the disciplines, the increase of knowledge

BUT

Fragmentation is an obstacle to the systemic vision and the share of knowledge

Computer scientists, engineers and developers

COMPUTER SCIENTISTS

- Data Models
- Algorithms
- Complexity

COMPUTER ENGINEERS

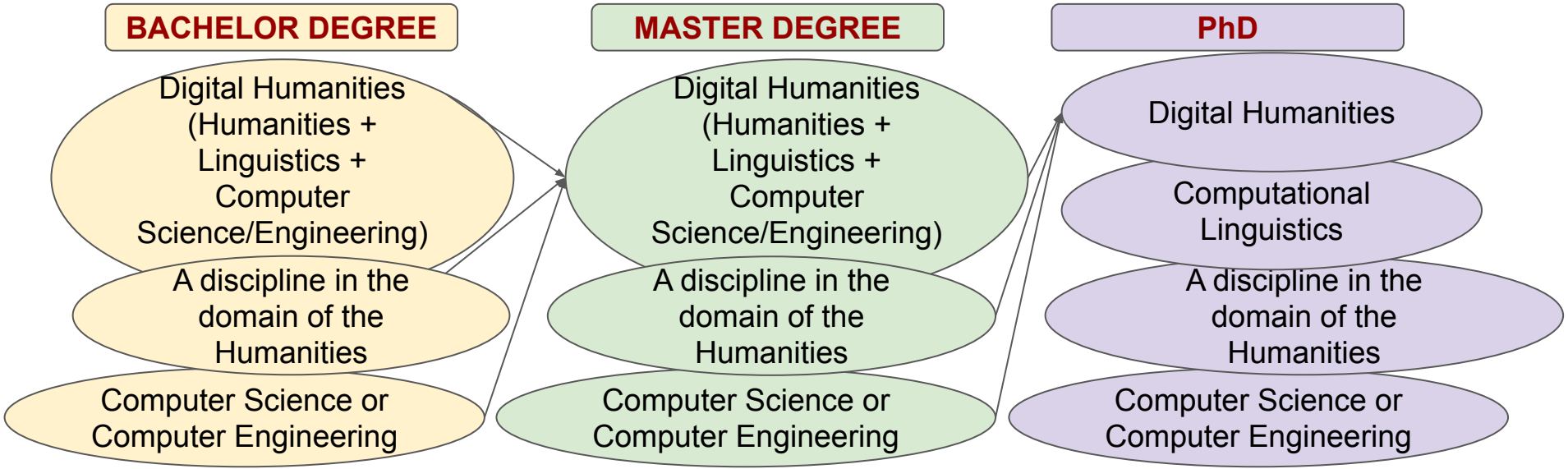
- Design
- Architectures
- Optimization

SOFTWARE DEVELOPERS

- Implementation

DH natives

Possible compositions (among others)



Philology or Philologies?

The **edition** of texts (ecdotics) and their **interpretation** (hermeneutics) are the two pillars of the philological activity, but the discipline is highly specialized in

- Classical Philology
- Biblical Philology
- Romance Philology
- German Philology
- ...

Unshared traditions of studies and methods

Very often, specialists of a single author (e.g. Homer or Dante) or period (e.g. Middle Age) are not in contact with specialists of other authors or periods.
Independent **traditions of studies** arise, with unshared methods

- Lachmann's method
- Bédier's method
- Textual bibliography
- Genetic Philology

Interaction between philologists and “programmers”

The direct communication between philologists and computer scientists, computer engineers and/or developers is challenging:

- no common vocabulary
- bad formalization of requirements and specifications
- too big expectations
- mutual mistrust

Best practices … for whom?

Who are the recipients of "**best practices**"?

- **Digital humanists** suggest XML-TEI encoding
- **Computer engineers** create visual tools

BUT

specialists in the domain of the Humanities suffer a severe **cognitive stress** in XML-TEI encoding and have a sub-optimal user experience with visual tools for iterative tasks

Mediation

Mediation is necessary among

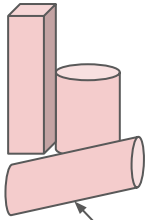
- the subcommunities of philologists (classical, biblical, german …)
- the subcommunities of other humanists (e.g. epigraphists, paleographers, etc.)
- the community of computer scientists
- the community of computer engineers
- the community of developers
- the subcommunities of Digital Humanists

Methods

Methodological pillars

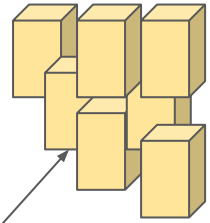
- Irreducibility of the historical-critical method to the scientific experimental method
- Hermeneutic circle

Irreducibility of the historical-critical method to the experimental-scientific method



interesting as a **specific exemplar** in its historical context, meaningful per se

object under obser-vation



interesting as a **generic sample** of a wider phenomenon, interchangeable with other samples

Historical-critical method

The historical-critcal method tends to be **idiographic**, i.e. interested in specific, unique phenomena (according to the neo-Kantian philosopher W. Windelband)

Primary sources often are **rare** and their interpretation require **qualitative analyses**

Secondary literature in its historical perspective confirm the **intrinsic value** of the literary or documentary primary source

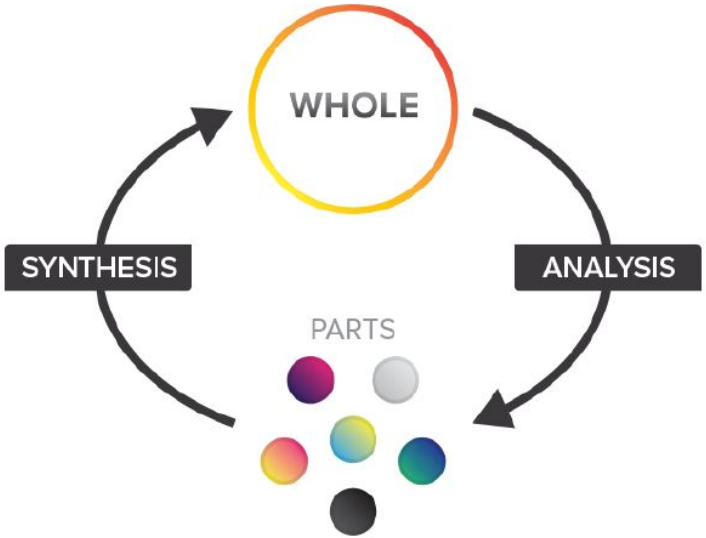
Experimental-scientific method

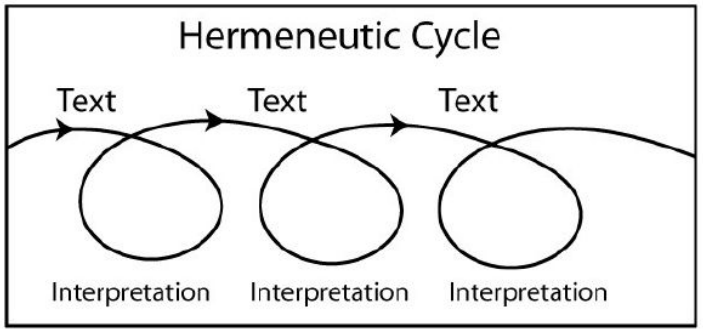
the experimental-scientific method tends to be **nomothetic**, i.e. interested in general, repeatable phenomena (Windelband)

The object under observation is just a **sample**, a **frequent** manifestation of one or more phenomena, suitable for **quantitative analyses**

The scientific literature is not focused in the peculiar details of each sample

Hermeneutic circle





(image sources: `http://bit.ly/1KkmhBU` and `http://bit.ly/2zv71lT`)

Research question

The research question should be **relevant** in the domain of the Humanities.

Accordingly, methods should be **understandable** (and **reviewed**) by domain experts and results should be **valuable** for a specific field of the humanities (e.g. classical philology)

Domain-Specific Languages (DSLs)

Domain-Specific Languages (DSLs) are computer languages optimized for a particular domain of application or domain of knowledge

DSLs are **concise** and **familiar** to domain experts

For instance, SQL is a DSL focused on relational databases

The definition of the grammar for a Domain-Specific Language follows a top-down approach, which can be easily managed by the domain expert, possibly with the help of a software designer

Open categories and incremental adjustments

The knowledge of the whole (e.g. a large corpus of literary texts) increases during the annotation process. The adjustment of the categories used to annotate texts is not an accident but it is an integral part of the annotating process.

For this reason categories that we use to annotate texts must be open, according to the annotation procedures promoted by the CATMA Project (http://catma.de)

Standards + residuals

Standards are **vital**:

- we must reuse resources created by others
- we must provide the resources created by us

BUT

- we can adopt standard formats for import and export of our data and use a domain-optimized format inside our applications
- apply standards, but accept to have **residuals**, if necessary
- participate to the standardization scientific boards

Foundational ontologies + domain ontologies

ONTOLOGICAL ORGANIZATION OF YOUR ANNOTATIONS

We suggest a mixed top-down and bottom-up approach: the initial tagset, established *a priori*, is adjusted *a posteriori* with the aid of expert ontologists

FOUNDATIONAL ONTOLOGIES

The upper-level concepts are organized in a foundational ontology (e.g. the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) developed by N. Guarino

DOMAIN ONTOLOGIES

The domain-specific concepts (e.g. the concepts involved in the representation of rituals in ancient tragedies) are organized in a domain ontology

Data

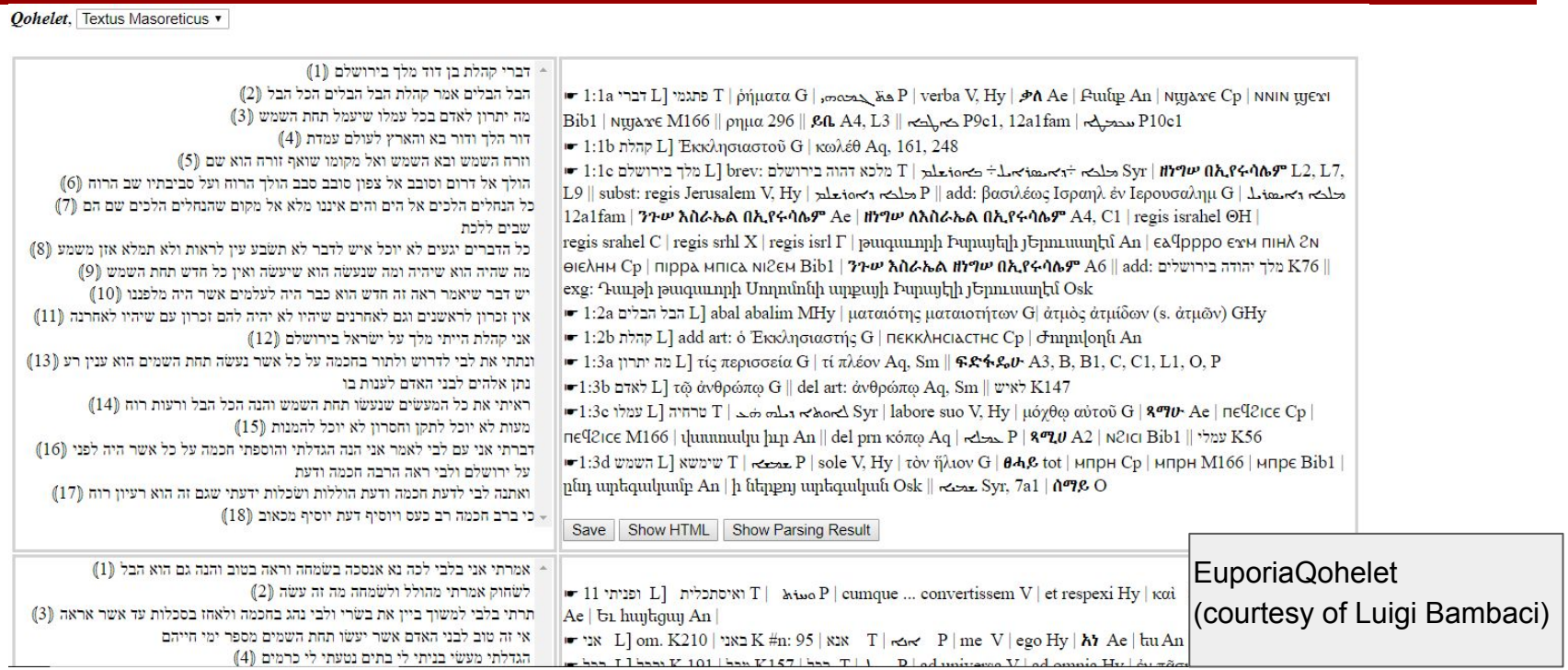
Digital Ecdotics

Best practices in digital ecdotics are based on XML-TEI encoding. But the verbosity of XML distracts specialists from their focus.

We want

- not only **machine actionable** digital scholarly editions (a need in Digital Philology),
- but also **compact**, **truly human readable** digital and **printable** editions (a need in the traditional subfields of philology)

Multilingual critical apparatus



EuporiaQohelet
(courtesy of Luigi Bambaci)

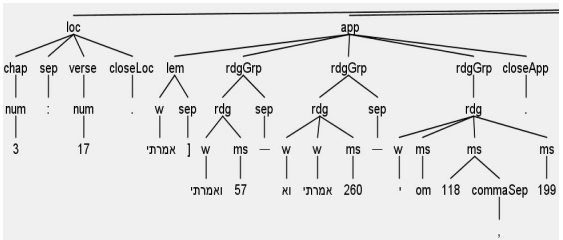
Critical apparatus as DSL

XSL-T

from TEI back to DSL

1. לכל bis 210 — כל 147. ועת — ו^ 152, 201. השמים sup. raſ. 180 — השמש 147, primo 77 ; forte 1.
2. ועת 1°— ו^ 56. עת לשעת^ 56. לעקר 2, 19, 89, 107, 141, 239, 248, 270 — להקור 275 — primo לעקות 244.
3. להרג 1, 2, 4, 14, 17, 18, 30, 50, 56, 57, 82, 93, 107, 111, 118, 170, 172, 173, 176, 177, 187, 188, 196, 212, 213, 218, 225, 240, 252, 260, 279, 284, 286, 680. לרפא 1, 4, 14, 17, 18, 19, 30, 56, 57, 77, 82, 93, 95, 108, 121, 136, 151, 167, 172, 173, 187, 188, 212, 213, 218, 224, 239, 240, 248, 252, 253, 384, 680. עת 2° — ועת 95. לפרץ 4, 19, 30, 56, 82, 93, 95, 119, 151, 157, 172, 181, 188, 212, 218, 225, 235, 236, 239, 244, 260 — לנתוץ 384. נ — לבנות forte ב 2.
4. עת לבכות bis 17. לבכות^ 151 — ו^ 18. ועת forte — nunc רפוד 178. ספד 82 ועד 14. לשחק 4, 18, 30, 107, 170, 177, 248, 680. רקד 30.
5. ערת 1° forte על 101. אבנים^ 19, 152. כנס 224 — בצתם כנוס sup. raſ. 31. עת 2° — עת* 674. כנס כנוס 117 — לחבק 14, 17, 18, 30, 56, 57, 80, 93, 94, 100, 102, 107, 117, 119, 141, 153, 166, 177, 191, 218, 224, 237, 240, 244, 251, 253, 260, 384. ועת לרחק^ 157. לרחוק 14, 57, 76, 77, 80, 95, 99, 101, 107, 108, 117, 118, 119, 121, 152, 153, 166, 173, 177, 178, 188, 192, 212, 213, 218, 223, 224, 225, 226, 227, 235, 236, 237, 240, 245, 249, 252, 259, 384, 680, 693. מחבק^ 152.

Parsing Kennicott's apparatus
(courtesy of Luigi Bambaci)




```
1  <app>
2    <lem wit="#L">אמרתי</lem>
3    <rdgGrp>
4       <rdg wit="#T">אמרית</rdg>
5       <rdg wit="#GV">εἶπον</rdg>
6    </rdgGrp>
7    <rdgGrp>
8       <rdg wit="#GB">καὶ εἶπον</rdg>
9       <rdg wit="#P">ܐܡܪܬ</rdg>
10      <rdg wit="#V">et dixi</rdg>
11   </rdgGrp>
12    <rdgGrp>
13      <rdg wit="#GA">ἐκεῖ εἶπον</rdg>
14   </rdgGrp>
15 </app>
```

from DSL to Parser

CFG

from Parser to TEI

VISITOR

Digital Hermeneutics

Interpretation is **potentially** an **infinite** process (Schleiermacher), but it is **actually limited** by the hermeneutic activity historically determined (Rastier).

Treebanks for morphosyntactic analyses, semantic annotations, named entity recognition should integrate the information extracted from ancient, modern and contemporary commentaries.

Evaluating variants

13 D'altra parte è indubbio che nel tradurre nel suo francese neogotico il latino di Adso, il Vallet abbia introdotto ~~di suo~~ varie licenze, e non ~~sempre~~ soltanto stilistiche. Per esempio i personaggi parlano talora delle virtù delle erbe rifacendosi chiaramente a quel libro dei segreti attribuito ad Alberto Magno che ebbe nei secoli infiniti rifacimenti. È certo che Adso lo conoscesse, ma rimane il fatto che egli ne cita dei brani che riecheggiano troppo letteralmente vuoi ricette di Paracelso vuoi chiare interpolazioni di un'edizione dell'Alberto di sicura epoca Tudor. D'altra parte ho appurato in seguito che ai tempi in cui il Vallet trascriveva (?) il manoscritto di Adso ~~, circolava~~ circolavano a Parigi ~~un'edizione settecentesca~~ edizioni settecentesche del Grand e del Petit Albert ormai irrimediabilmente ~~inquinata~~ inquinate. Tuttavia, come essere sicuri che il testo a cui si rifacevano Adso o i monaci di cui egli annotava i discorsi, non contenesse, tra glosse, scolii e appendici varie, anche annotazioni che poi avrebbero nutrito la cultura posteriore?

```
* [0] @narratoreEco §manoscritti
* [1] {di suo} = §riduzione/min
* [2] {sempre} = §riduzione/par || {sempre} = #FO
* [3] {, circolava} : <circolavano> = #del : #concordanza/numero
* [4] {un'edizione settecentesca} : <edizioni settecentesche> = #del : #concordanza/numero
* [5] {inquinata} : <inquinate> = #del : #concordanza/numero
```

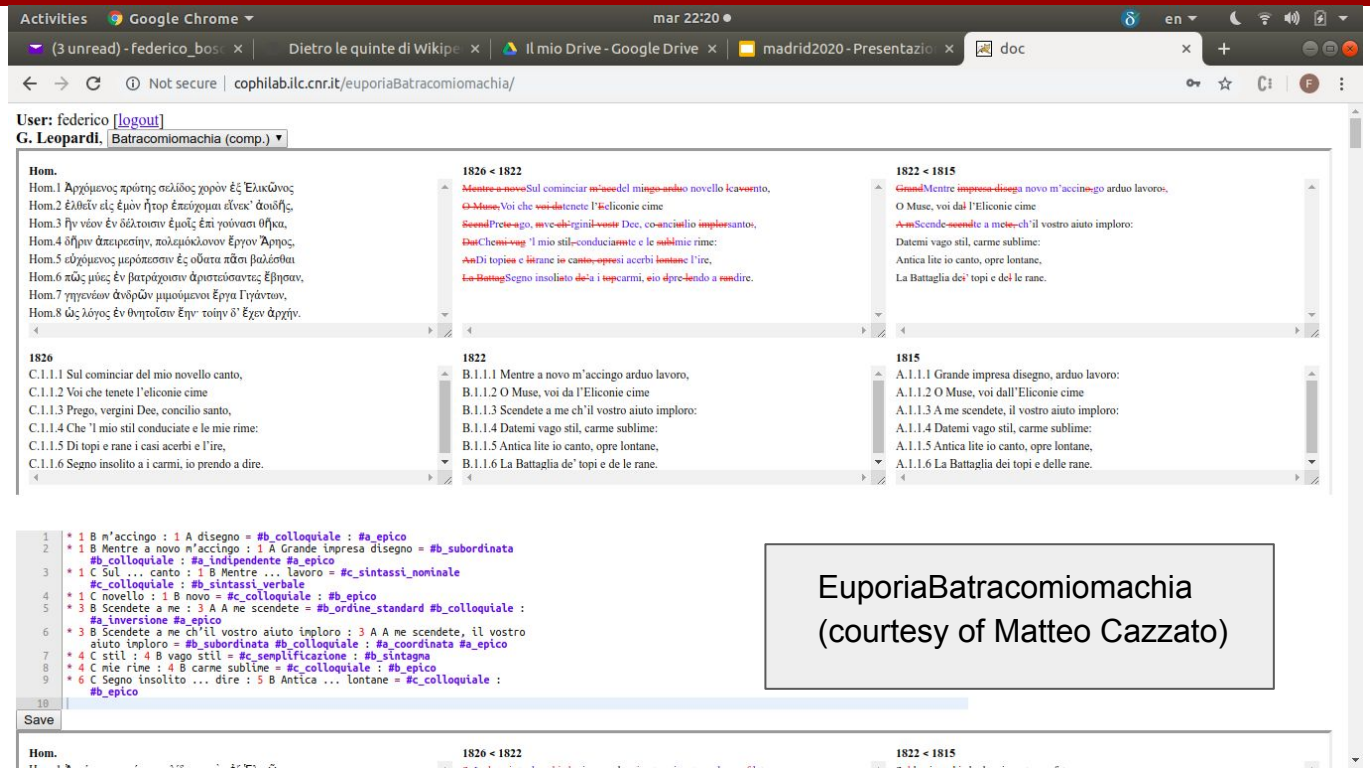
Saved

14 Infine, dovevo conservare in latino i passaggi che lo stesso abate Vallet non ritenne opportuno tradurre, forse per conservare l'aria del tempo? Non v'erano giustificazioni precise per farlo, se non un senso, forse malinteso, di fedeltà alla mia fonte... Ho eliminato il soverchio, ma qualcosa ho lasciato. E temo di aver fatto come i cattivi romanzieri che, mettendo in scena un personaggio francese, gli fanno dire — parbleu! — e — la femme, ah! la femme! — .

Saved

EuporiaEco
(courtesy of Christian D'Agata)

Lexical investigations



EuporiaBatracomiomachia
(courtesy of Matteo Cazzato)

Semantic and thematic annotation

Aeschylus, *Agamemnon*

60 οὕτω δ' Ἀτρέως παῖδας ὁ κρείσσων
61 ἐπ' Ἀλεξάνδρῳ πέμπει ξένιος
62 Ζεὺς πολυάνορος ἀμφὶ γυναικὸς
63 πολλὰ παλαίσματα καὶ γυιοβαρῆ
64 γόνατος κονίαισιν ἐρειδομένου
65 διακναιομένης τ' ἐν προτελείοις
66 κάμακος θήσων Δαναοῖσι
67 Τρωσί θ' ὁμοίως. ἔστι δ' ὅπη νῦν
68 ἔστι· τελεῖται δ' ἐς τὸ πεπρωμένον·
69 οὔθ' ὑποκαίων οὔθ' ὑπολείβων
70 οὔτε δακρύων ἀπύρων ἱερῶν
71 ὀργὰς ἀτενεῖς παραθέλξει.
72 ἡμεῖς δ' ἀτίται σαρκὶ παλαιᾷ
73 τῆς τότ' ἀρωγῆς ὑπολειφθέντες
74 μίμνομεν ἰσχὺν
75 ἰσόπαιδα νέμοντες ἐπὶ σκήπτροις.
76 ὅ τε γὰρ νεαρὸς μυελὸς στέρνων
77 ἐντὸς ἀνᾴσσων
78 ἰσόπρεσβυς, Ἄρης δ' οὐκ ἔνι χώρᾳ,
79 τό θ' ὑπέργηρων φυλλάδος ἤδη
80 κατακαρφομένης τρίποδας μὲν ὁδοὺς
81 στείχει, παιδὸς δ' οὐδὲν ἀρείων
82 ὄναρ ἡμερόφαντον ἀλαίνει.
83 σὺ δέ, Τυνδάρεω
84 θύγατερ, βασίλεια Κλυταιμήστρα,
85 τί χρέος; τί νέον; τί δ' ἐπαισθομένη,
86 τίνος ἀγγελίας
87 πειθοῖ περίπεμπτα θυοσκεῖς;
88 πάντων δὲ θεῶν τῶν ἀστυνόμων,
89 ὑπάτων, χθονίων,

☛ [69 οὔθ' ὑποκαίων] #victimam_ardere ☚
☛ [69 ὑπολείβων] #libatio ☚
☛ [69 ὑπολείβων] @vl:69_1 ἀπολείβων Bothe ☚
☛ [69 ὑπολείβων] @vl:69_2 ἐπιλείβων Schütz ☚
☛ [70 οὔτε δακρύων] @vl:70_1 om. Bamberger ☚
☛ [69 οὔθ' ὑπολείβων _ἀπύρων ἱερῶν] {@vl:69_2} {@vi:70_2} #in_oblationem_libare☚
☛ [69 οὔθ' ὑπολείβων] {@vl:69_2} #in_sacrificium_libare Fraenkel ☚
☛ [69 οὔθ' ὑπολείβων_ ἀπύρων ἱερῶν] {@vl:69_2} {@vi:70_1a} #in_sacrificium_libare Wecklein ☚
☛ [69 οὔθ' ὑπολείβων_70 ἀπύρων ἱερῶν] {@vl:69_1} {!@vi:70_1} #libatio ☚
☛ [70 ἀπύρων ἱερῶν] @vi:70_1 #sacrificium #sine_igne ☚
☛ [70 ἀπύρων ἱερῶν] @vi:70_2 #oblatio_incruenta #sine_igne ☚
☛ [70 ἀπύρων ἱερῶν / 71 ὀργὰς ἀτενεῖς] {@vi:70_1} @vi:70_1a #sacrificium_reicere ☚
☛ [70 ἀπύρων ἱερῶν] {@vi:70_1} @vi:70_1b #hominem_sacrificare Bollack ☚
☛ [71 ὀργὰς ... παραθέλξει] #ritus_propositum #deos_placare☚
☛ [71 ὀργὰς ἀτενεῖς] #ira_deorum ☚
☛ [75 νέμοντες ἐπὶ σκήπτροις] #!h #s #sceptrum ☚

EuporiaRAGT/Editor
(courtesy of Giulia Re)

Adapting Turtle

Tragica, Hippolytus

Ἀφροδίτη
1 Πολλὴ μὲν ἐν βροτοῖσι κοὐκ ἀνώνυμος
2 θεὰ κέκλημαι Κύπρις οὐρανοῦ τ' ἔσω·
3 ὅσοι τε Πόντου τερμόνων τ' Ἀτλαντικῶν
4 ναίουσιν εἴσω, φῶς ὁρῶντες ἡλίου,
5 τοὺς μὲν σέβοντας τἀμὰ πρεσβεύω κράτη,
6 σφάλλω δ' ὅσοι φρονοῦσιν εἰς ἡμᾶς μέγα.
7 ἔνεστι γὰρ δὴ κἀν θεῶν γένει τόδε·
8 τιμώμενοι χαίρουσιν ἀνθρώπων ὕπο.
9 δείξω δὲ μύθων τῶνδ' ἀλήθειαν τάχα·
10 ὁ γάρ με Θησέως παῖς, Ἀμαζόνος τόκος,
11 Ἱππόλυτος, ἁγνοῦ Πιτθέως παιδεύματα,
12 μόνος πολιτῶν τῆσδε γῆς Τροζηνίας
13 λέγει κακίστην δαιμόνων πεφυκέναι·
14 ἀναίνεται δὲ λέκτρα κοὐ ψαύει γάμων,
15 Φοίβου δ' ἀδελφὴν Ἄρτεμιν, Διὸς κόρην,
16 τιμᾷ, μεγίστην δαιμόνων ἡγούμενος,
17 χλωρὰν δ' ἀν' ὕλην παρθένῳ ξυνὼν ἀεὶ
18 κυσὶν ταχείαις θῆρας ἐξαιρεῖ χθονός,
19 μείζω βροτείας προσπεσὼν ὁμιλίας.
20 τούτοισι μέν νυν οὐ φθονῶ· τί γάρ με δεῖ;

```
1  * 2 οὐρανοῦ : #ontology sky is_sacred_to Aphrodites @cfr:59 .
2  * 3 Πόντου ... Ἀτλαντικῶν : #ontology sea is_sacred_to Aphrodites .
3  * 17 χλωρὰν ... 18 χθονός : this_wood is_associated_to Hippolytus .
     #ontology hunt is_sacred_to Artemis; is_associated_to ephebia .
4  * 17  χλωρὰν δ' ἀν' ὕλην : this_wood is_sacred_to Artemis;
     is_associated_to hunt.
5  * 18 θῆρας ἐξαιρεῖ χθονός : Hippolytus is_associated_to hunt .
     #ontology hunt is_sacred_to Artemis; is_associated_to ephebia .
6  * 36 ναυστολεῖ : sea is_associated_to Theseus . #ontology sail
     is_sacred_to Aphrodites @bibl:Demetiou2010 .
7  * 44 ὁ πόντιος 45 ἅναξ : πόντιος is_a epithete; is_associated_to sea,
     Poseidon, Aphrodite.
8
```

Save

EuporiaHippolytus
(courtesy of Giulia Re)

The involvement of students

42 ἁβροδιαίτων δ' ἕπεται Λυδῶν
43 ὄχλος, οἵτ' ἐπίπαν ἠπειρογενὲς
44 κατέχουσιν ἔθνος, τοὺς Μιτρογαθὴς
45 Ἀρκτεύς τ' ἀγαθός, βασιλῆς δίοποι,

Tr. di Fraccaroli

41 Dietro i fastosi battaglier superbi
42 Vedova Lidia affretta i preghi e piange.
43 Nullo, se cuore ebbe da forte e nerbi,
44 Fe' voti invan che il fato aspro si cange:
45 Due re li spingon fieramente acerbi,
46 Artèo, che vale ei solo una falange,
47 E Mitrogàte, a cui su l' auree bende
48 Serpe lucida biscia e al sol s' accende.

Nuova traduzione

39 Incede con eleganza la schiera dei Lidi,
40 abitanti della regione più interna.
41 Li seguono con molti carri, a due o tre gioghi,
42 Arteo il valoroso e Metrogate, re entrambi,

Salvato

☛ [43 ἠπειρογενὲς ~ 44 ἔθνος] #FR: iperbato☚
☛ [42-43 Λυδῶν / ὄχλος] #FR: enjambement ☚
☛ [42 Λυδῶν] #NE: popolo ☚
☛ [43-44 ἠπειρογενὲς/ ἔθνος] #FR: enjambement ☚
☛ [46 ei solo una falange] #TR: non viene riscontrata nessuna corrispondenza col testo greco ☚
☛ [47-48 a cui su l' auree bende ... / al sol s' accende] #TR: non viene riscontrata nessuna corrispondenza col testo greco ☚
☛ [42 Λυδῶν] #NE: popolo #OL: il popolo dei Lidi, discendente dalla popolazione Ittita, fu governata da abili governatori ☚
☛ [42 ἁβροδιαίτων ~ Λυδῶν] #FR: iperbato☚
☛ [42 ἁβροδιαίτων] #CO: ἁβρός δίαιτα, di vita raffinata☚
☛ [44 Μιτρογαθὴς] #NE: persona. ☚
☛ [45 Ἀρκτεύς] #NE: persona. #FR: hapax legomenon. ☚
☛ [] ☚☛ [] ☚

Salvato

EuporiaPersae
(courtesy of Liceo Gargallo, Siracusa)

The involvement of citizens

===TITOLO===
Sommeil Interrompu

===DESCRIZIONE===
* immagine: Una donna interrompe il sonno di un uomo che si riposa vicino ad un albero.
* note: Sono presenti tre timbri e un francobollo.

===TESTO===
* data: 24/07/1913
* luogo: Ravenna
* corpo: A Ravenna piove sempre: è una gioia per chi non è andato ai bagni. Per ora niente di nuovo, tutti bene. Stasera qui c'è musica, come ieri che fu S.Apollinare: ad Arona, niente
* commiato: Saluti affettuosi,
* firma: Giuseppe

===DESTINATARIO===
* nome: Signorina Oliva Turtura
* indirizzo: Via Cavour 12
* località: Arona (Lago Maggiore)

===NOTE===
La cartolina è leggermente danneggiata nella parte inferiore sinistra

EuporiaCartoline
(courtesy of Francesco Melighetti)

Instruments

Make your (simple) tools

Digital Humanists cannot be limited by technologies created in other domains, often very distant from their interests (e.g. business applications), sometimes just apparently distant (e.g. bioinformatics, which provides align algorithms userful in stemmatology) and sometimes very close but not totally satisfying (e.g. computational linguistics)

Digital Humanists must make their tools according to their research question

Working prototypes, no mock-ups

- Digital Humanists cannot be expert only in data encoding, but they cannot be skilled at the same level of computer engineers
- In order to be autonomous, they should be able to create working prototypes, not just mock-ups to be implemented by developers
- A blend of javascript, python, xquery, XML, json, html, css should be the **minimal** programming competence of a Digital Humanist, but also java, SQL, XSLT, R should become familiar

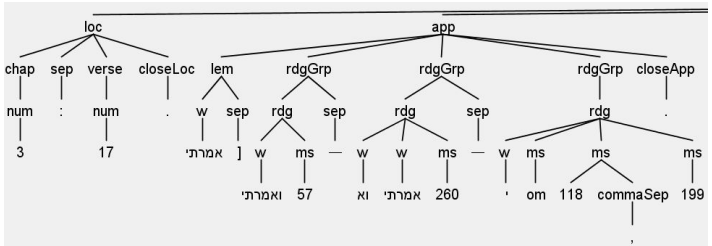
From DSL to XML

```
grammar QoheletEuporia;

app : loc lem;
lem : w+ wit lemSep;
loc : chap + locSep + v?;
chap : NUM;
v : NUM;
locSep : DOUBLE_POINT;
lemSep : R_BRACKET;
wit : ALPHA_SEQ;
w : HEBW ;

NUM : [0-9]+('.'[0-9]+)?;
ALPHA_SEQ : [a-zA-Z]+;
DOUBLE_POINT : ':';
R_BRACKET : ']';
HEBW : [\u0590-\u05ff]+;
```

(courtesy of Luigi Bambaci)

CFG — ANTLR-4 SOFTWARE — VISITOR

PARSER




```
<listApp>
   <loc>
      <chap>
         <num>5</num>
      </chap>
      <sep>:</sep>
      <verse>
         <num>1</num>
      </verse>
      <closeLoc>.</closeLoc>
   </loc>
   <app>
      <lem>
         <w>אל</w>
         <occ>
            <num>1</num>
            <numeroSign>°</numeroSign>
            <sep>—</sep>
         </occ>
      </lem>
      <rdgGrp>
         <rdg>
            <term>primo</term>
            <w>על</w>
            <ms>18</ms>
         </rdg>
      </rdgGrp>
      <closeApp>.</closeApp>
   </app>...
</listApp>
```

xquery, a viable solution

xquery is based on the **FLWOR** (For, Let, Where, Order-by, Return) construct. It is very compact and students in traditional humanities can easily learn it

eXist-db

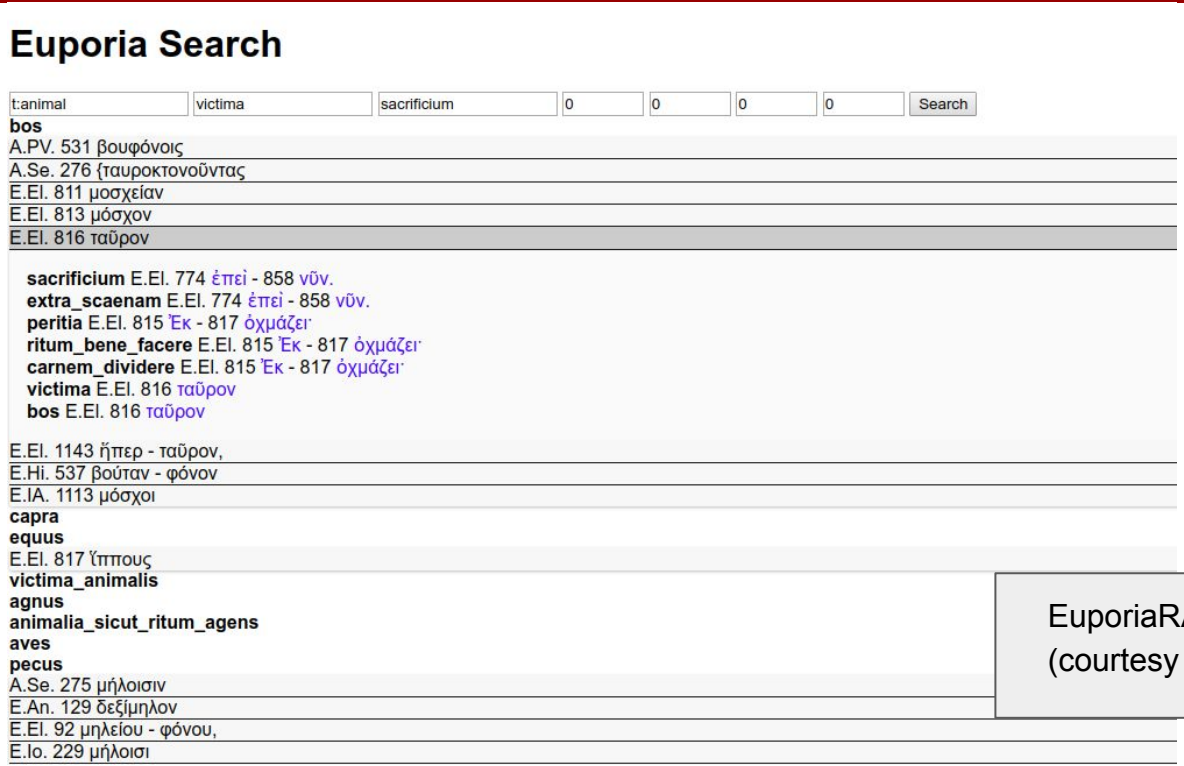
By using eXist-db, few lines of code are enough to create indexes, concordances, CRUD applications, complex layouts etc.

```
xquery version "3.1";
declare namespace output = "http://www.w3.org/2010/xslt-xquery-serialization";
declare boundary-space preserve;
declare option output:method "html";
declare option output:media-type "text/html";

let $selection:=request:get-parameter("selection", ())
let $annoDoc:=doc("/db/apps/matteo/annotation.xml")
return <html><head><meta charset="UTF-8"/><link rel="stylesheet" type="text/css" href="canno.css"/></head><body><table>{
    for $keyword in $annoDoc//keyword
        where $keyword=$selection
        order by $keyword
        let $note:=$keyword/ancestor::note
        let $bAnno:=if ($note//firstWordRef//numRef) then $note//firstWordRef//numRef else $note//singleWordRef//numRef
        let $eAnno:=if ($note//lastWordRef//numRef) then $note//lastWordRef//numRef else $note//firstWordRef//numRef
              let $bWord:=if ($note//firstWordRef//wordRef) then $note//firstWordRef//wordRef else $note//singleWordRef//wordRef
        let $eWord:=if ($note//lastWordRef//wordRef) then $note//lastWordRef//wordRef else $note//firstWordRef//wordRef
        return <tr><td><a href="./search-for-annotation-text.xql#{$bAnno}" target="iftext">{substring($bAnno,3)}</a></td><td>{$bWord}</td><td>{substring($eAnno,3)}</td><td>
}</table></body></html>
```

A new generation of humanist-developers

We need a new generation of humanists, especially of philologists, skilled in computer programming: coding (and not just textual encoding) must be a daily practice comparable to the use of word-processors, search engines online, etc.

EuporiaSearch



EuporiaRAGT/Search
(courtesy of Gloria Mugelli)

Engineering the prototypes

Prototypes are project-centered. They are an affordable for pilot project and small project. Well funded projects (e.g. ERCs) and infrastructural projects (e.g. some Horizon 2020 projects) require that prototypes are re-engineered, in order to be modular, scalable, sustainable

User-centered and community-centered approaches

Users

Humanists have **specific research goals** and the instruments they use must be consistent with their research questions

For example, the study of the colometry of a poetic text requires instruments that allow users to properly display how verses are placed on the pages of different manuscripts

Communities

The interaction among the *community of humanists*, the *community of computer scientists, computer engineers* and *developers* and *the community of digital humanists* helps to bring out different **points of view** in the search for optimal solutions

Libraries of components and Web APIs

LIBRARIES OF COMPONENTS

Community-centered projects (e.g. DiXiT, PARTHENOS) are based on the requirement analysis at the community level, not just at the single user level

The creation of libraries of components, with the same APIs and implementations in many programming languages, is a natural consequence

WEB APIs

Web APIs for textual scholarship promote the distribution of textual resources and computational instruments

Infrastructures

Research Infrastructures for the Humanities

CLARIN

Common Language Resources and Technology Infrastructures

DARIAH

Digital Research Infrastructure for the Arts and the Humanities

FAIR Data

- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**eusable

Wilkinson, M. D., et al. (2016). **The FAIR Guiding Principles for scientific data management and stewardship**. *Scientific data*, *3*, 160018. doi:10.1038/sdata.2016.18

Findable

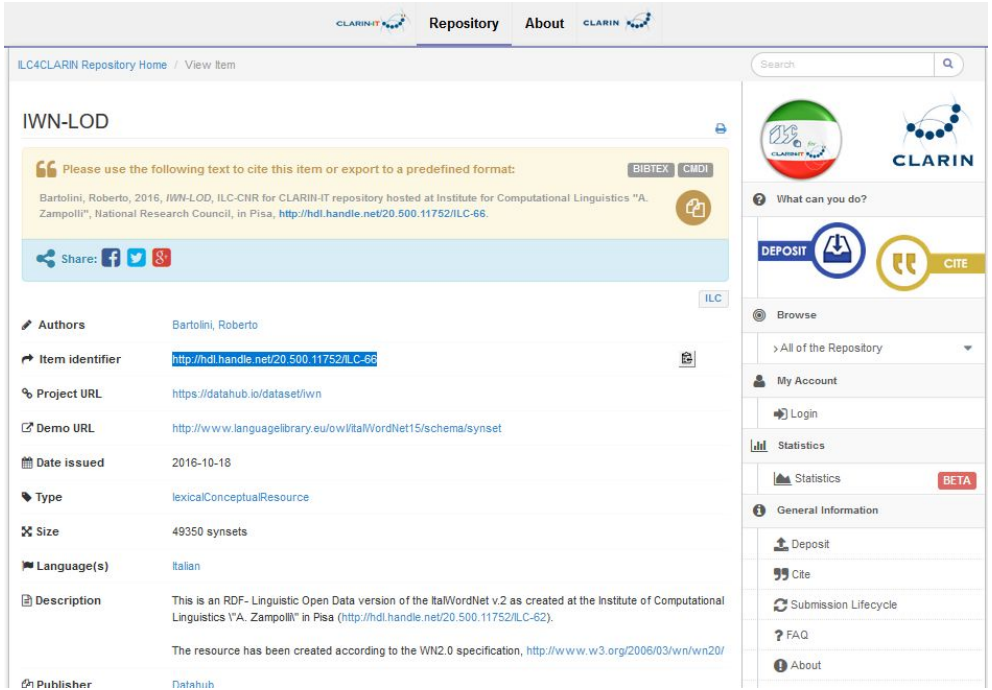
“F1. (meta)data are assigned a **globally unique and persistent identifier**

F2. data are described with **rich metadata** (defined by R1 below)

F3. metadata clearly and explicitly include the identifier of the data it describes

F4. (meta)data are registered or indexed in a searchable resource”

(Wilkinson et al., 2016)



Accessible

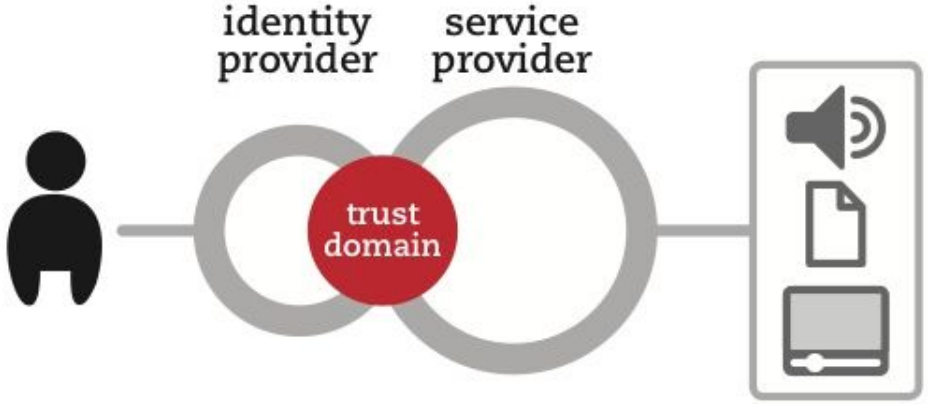
“A1. (meta)data are retrievable by their identifier using a standardized communications protocol

A1.1 the protocol is open, free, and universally implementable

A1.2 the protocol allows for an authentication and authorization procedure, where necessary

A2. **metadata are accessible, even when the data are no longer available**”

(Wilkinson et al., 2016)



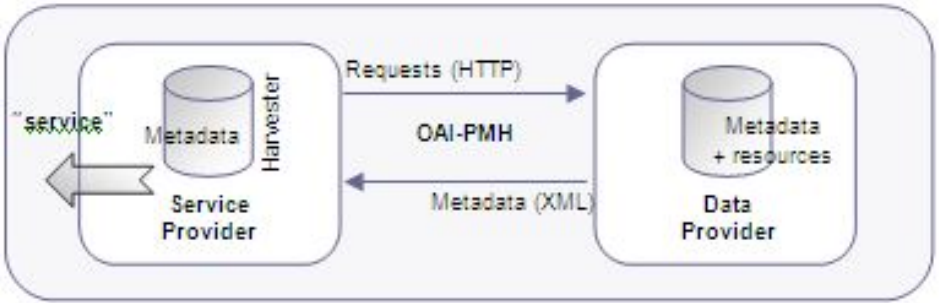
Interoperable

“I1. (meta)data use a **formal**, accessible, shared, and broadly applicable language for knowledge representation

I2. (meta)data use **vocabularies that follow FAIR principles**

I3. (meta)data include **qualified references to other (meta)data**”

(Wilkinson et al., 2016)



Reusable

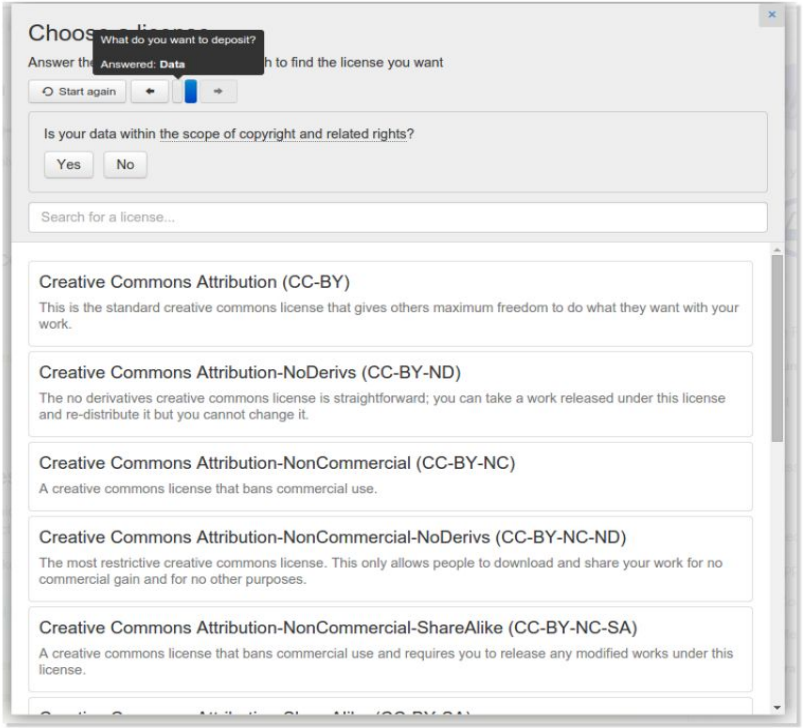
“R1. meta(data) are **richly** described with a **plurality** of **accurate** and relevant **attributes**

R1.1. (meta)data are released with a clear and accessible **data usage license**

R1.2. (meta)data are associated with detailed **provenance**

R1.3. (meta)data meet **domain-relevant community standards** ”

(Wilkinson et al., 2016)



Versioning

- How to ensure that we are talking about the **same thing** under the **same name**?

  PROBLEM

- How to migrate (=**reuse**) annotations from an older to a newer version?

Obsolescence Management

LONG-TERM PRESERVATION ISSUES

"one no longer preserves **tangible physical objects** per se, but views **abstract representations** of such objects that can be reconstructed in an **unpredictable** technological future."

J.P. Chanod, **Will Your Data Still Be Around Tomorrow?**, 2013, http://bit.ly/2tMRP3c)

Sustainable Repositories

"Storing language resources and related datasets is something that requires a sound organization and attention for digital sustainability. After all, one of the important aims of CLARIN is to ensure that digital language resources are made available to a broad community on a **long-term basis**. This is achieved by establishing **data repositories** at the **centres**, which host **digital files** and the **associated metadata**. For reference purposes, these repositories also assign **persistent identifiers** to the resources, so that e.g. a specific dataset can be easily **cited** in a paper." (http://bit.ly/2U3wLjl)

Repeatability, Replicability and Reproducibility

ACM DEFINITIONS

- **Repeatability (Same team, same experimental setup)**
- **Replicability (Different team, same experimental setup)**
- **Reproducibility (Different team, different experimental setup)**

GOODMAN'S DEFINITIONS

- **Methods reproducibility (=ACM replicability):** provide **sufficient detail** about **procedures** and **data** so that the same procedures could be exactly repeated
- **Results reproducibility (=ACM reproducibility):** obtain the same **results** from an **independent** study with procedures as closely matched to the original study as possible
- **Inferential reproducibility:** draw the same **conclusions** from either an **independent replication** of a study or a **reanalysis** of the original study

H. Plesser, **Replicability vs. Reproducibility**, 2018, http://bit.ly/36w2dte

Knowledge

The observer as part of the model

- Digital Humanities need to overcome the naive vision of the modern science, the dogma of objectivity
- Digital Humanists must be aware of the most updated epistemological debates
- In particular, they must take into account that the observer is part of the model: for example, CIDOC-CRM-Inf goes in this direction

Philological Schools

- The neutrality of the point of view is a myth
- Data-driven *versus* hypothesis-driven research is another myth, because also the choice of (big) data, the pre-processing of data is based on hypotheses
- It is better to accept the existence of different digital philological schools, as in the past many different philological schools co-existed, internally consistent and externally opposed (e.g. Boeckh *versus* Hermann)

Freedom (from technological constraints)

As claimed above, Digital Humanists must be guided only by their research questions, and never limited by technological constraints: they must invent their instruments and propose new standards, everytime the current solutions are not sufficient

Recording subjective choices

- Subjectivity is a value, not a bias that must be reduced or eliminated
- Subjective choices (e.g. which variant to insert in the established text and which variants to put in the critical apparatus) are determined by the horizons of belief of the scholars
- We need to model the subjectivity (e.g. by applying the principles of the epistemic and doxastic logic)

Die erschließende Wiedergabe

Edition is die erschließende Wiedergabe historischer Dokumente

“In German, this works quite well. Unfortunately, however, it relies on the central, yet untranslatable, term ***erschließen***, which encompasses any activity that increases the amount of information concerning a specific object and thus enhances its accessibility and usability. Depending on context, words such as ***develop***, ***open up***, ***deduce*** or ***infer*** may be used to render this concept in English.” (P. Sahle, 2016)

Systemic annotations

But the investigations made by Digital Humanists are not impressionistic, they are systemic, because the same phenomena are (or want to be) annotated in the same way

Research questions and metaquestions

- Traditional humanists have research questions that their instruments can answer
- Digital Humanists can bring their investigation at an upper level, the level of metaquestions, possible because we can process large (annotated) textual resources
- A typical traditional questions: find loci paralleli, on a lexical base, to a specific textual passage
- A new metaquestion: find loci paralleli, on a semantic base, of every textual unit

Close - Distant - Close Reading

- The traditional close reading, i.e. the slow and accurate understanding of a text line by line, sentence by sentence, can be now integrated by distant reading, i.e. the application of statistical methods to large textual collections
- But distant reading is not the ultimate goal; it is a mean to explore which are the most interesting regions of an entire collection
- From the exploration of the corpus as a whole we must go back to the understanding of the parts: the never-ending hermeneutic circle

Scientific knowledge is historically determined

- Because the scientific knowledge is historically determined, no scholarly edition is the definitive edition; no treebank can be the unique treebank
- We have to avoid the risk of a new, digital vulgate, of a new *ipse dixit*

Conclusion

PLANS

- Resuming the interrupted dialogue between Digital and Non-Digital Humanists
- Coordinating the roles of the subcommunities inside the community of Digital Humanities

ACTIONS

- Identifying research questions that are relevant for both Digital and Non-Digital Humanists
- Creating textual resources, computational instruments and research infrastructures that are valuable for both Digital and Non-Digital Humanists

ACKNOWLEDGMENTS

Luigi Bambaci, University of Bologna
Andrea Bellandi, CNR-ILC
Matteo Cazzato, University of Pavia
Marilena Crucitti, Liceo Gargallo, Siracusa
Christian D'Agata, University of Catania
Riccardo Del Gratta, CNR-ILC
Angelo Mario Del Grosso, CNR-ILC
Anas Fahad Khan, CNR-ILC
Luis Fernando Lionello Nuñez, University of Bologna
Monica Monachini, CNR-ILC
Gloria Mugelli, University of Pisa
Giulia Re, University of Pisa
Andrea Taddei, University of Pisa