

Paulo maiora canamus

Raccolta di studi per Paolo Mastandrea

a cura di Martina Venuti, Massimo Manca

La galassia *Musisque Deoque*: storia e prospettive

Federico Boschetti

Istituto ILC-CNR, Italia

Angelo Mario Del Grosso

Istituto ILC-CNR, Italia

Linda Spinazzè

Università Ca' Foscari Venezia, Italia

Abstract The resources in support of Latin scholars created under the scientific direction of Paolo Mastandrea are numerous; the undertaking of *Musisque Deoque*, which aims at equipping the entire corpus of ancient Latin poetry with “significant variants”, is the most emblematic effort, open to further developments. Looking at the general history of these projects, we try to trace the future path of the “*Musisque Deoque* galaxy” within Open Science.

Keywords Classical philology. Latin poetry. Intertextuality. Open data. FAIR principles.

Sommario 1 Introduzione. – 2 Le Digital Humanities: molte comunità. – 3 La filologia classica: una disciplina storica da valorizzare nel contesto europeo. – 4 L'importanza dei dati FAIR. – 5 Ripetibilità, replicabilità e riproducibilità. – 6 Evitare il rischio di una nuova *vulgata*. – 7 La galassia *Musisque Deoque*. – 8 Per una breve storia dei progetti della galassia *Musisque Deoque*. – 9 La Galassia *Musisque Deoque* nella direzione FAIR. – 10 Conclusione: un bilancio e una prospettiva futura.

1 Introduzione

L'applicazione dell'informatica alle discipline umanistiche in Italia, per riprendere una bella metafora di Paul Maas (1952) riferita alla tradizione testuale, ha una «costituzione carsica». Non è possibile infatti



Edizioni
Ca' Foscari

Antichistica 32 | Filologia e letteratura 5

e-ISSN 2610-9352 | ISSN 2610-8836

ISBN [ebook] 978-88-6969-557-5 | ISBN [print] 978-88-6969-557-5

Peer review | Open access

Submitted 2021-05-28 | Accepted 2021-07-03 | Published 2021-12-14

© 2021 | Creative Commons 4.0 Attribution alone

DOI 10.30687/978-88-6969-557-5/026

405

seguire un percorso evolutivo lineare che parta dai primi studi della metà del secolo scorso e giunga ai tempi più recenti. Proprio come un fiume carsico, il flusso di conoscenze e di pratiche si è più volte separato in diversi rivoli, alcuni dei quali destinati a scomparire del tutto, altri riemersi a sorpresa dal sottosuolo per tornare a convergere verso il flusso principale, o a gonfiarsi isolatamente, oppure a disperdersi di nuovo. Fra i fattori di questa frammentazione ci sono il mancato riconoscimento accademico delle Digital Humanities come disciplina a sé stante, l'ondivago interesse della comunità scientifica nel corso dei decenni e la disomogenea distribuzione di fondi alla ricerca.¹

In questo contesto, l'attività di Paolo Mastandrea è stata determinante per costruire una cerniera fra la comunità dei filologi classici italiani riconosciuti dall'accademia e la comunità di chi si occupa a vario titolo di Digital Humanities.

2 Le Digital Humanities: molte comunità

In Italia, prima del 2011, più che una comunità di umanisti digitali si poteva identificare un nucleo di autorevoli studiosi a guida di gruppi di lavoro distribuiti sul territorio nazionale. Restringendo lo sguardo al campo delle discipline filologico-letterarie e in un ideale percorso geografico, pur senza voler fare un elenco esaustivo, si possono menzionare almeno Tito Orlandi e Giuseppe Gigliozzi (Roma), Andrea Bozzi (Pisa), Dino Buzzetti (Bologna), Roberto Busa (Milano) e Paolo Mastandrea (Venezia).

Ciascuno di questi studiosi aveva una diversa visione del rapporto fra informatica e discipline umanistiche in relazione a molteplici obiettivi di ricerca. Per quanto le generalizzazioni siano sempre insoddisfacenti, si possono tuttavia individuare alcune tendenze: dalla spiccata propensione all'astrazione di Tito Orlandi (2010) al mirabile equilibrio fra teoria e pratica di Gigliozzi (1997), dall'impegno a creare un ponte tra filologia e linguistica computazionale di Bozzi (2019) all'approccio ermeneutico al trattamento dei testi digitali introdotto da Buzzetti (Buzzetti, McGann 2006). Busa ha esteso le pratiche dei filologi grazie alla metodologia della linguistica computazionale (Busa 1990). Mastandrea, saldamente radicato alle discipline filologiche tradizionali per gli aspetti teorici, ha dimostrato grande apertura verso l'informatica per gli aspetti strumentali (Mastandrea, Tessarolo 2011).

¹ Anche se il lavoro è frutto di una riflessione collettiva (introduzione e conclusioni), le sezioni 2, 4, 5 e 6 sono di responsabilità di Federico Boschetti, le sezioni 3 e 8 sono state scritte da Linda Spinazzè, mentre le sezioni 7 e 9 sono da ascrivere ad Angelo Mario Del Grosso.

I primi percorsi universitari (triennali, magistrali, dottorali) di Informatica Umanistica e di Digital Humanities sono nati proprio nelle città dove questi studiosi operavano, creando il terreno fertile per estendere la formazione a tutto il territorio nazionale, come si può evincere da una mappa aggiornata dell'offerta formativa.²

Il 25 marzo 2011 si costituiva l'*Associazione di Informatica Umanistica e Cultura Digitale* (AIUCD) nel tentativo di dare identità a una comunità tanto vivace e produttiva sia sul piano metodologico che pratico, quanto eterogenea per ambiti disciplinari di provenienza. L'associazione, crescendo negli anni, ha acquisito autorevolezza anche all'estero, grazie sia al coordinamento con la European Association of Digital Humanities (EADH) e l'Alliance of Digital Humanities Organizations (ADHO), sia al potere attrattivo dei temi affrontati nei convegni annuali: è stato proprio il convegno nazionale AIUCD 2016 a Venezia ad aver inaugurato la forte presenza di interventi stranieri.³

L'università e le associazioni sono i luoghi privilegiati per il dibattito scientifico, ma per le Digital Humanities non vanno trascurati gli stretti legami tra aspetti scientifici e aspetti tecnologici, tra il 'cosa' e il 'come' della ricerca umanistica con metodi e strumenti digitali. Vale la pena dunque di prendere in considerazione almeno qualche esempio appartenente a tre tipologie di organismi che mirano a diffondere buone pratiche in questo campo: gli osservatori (indipendentemente dal loro *status* formale), i consorzi e le infrastrutture di ricerca.

Limitando nuovamente il campo delle Digital Humanities all'ambito più ristretto della filologia digitale e del Literary Computing, spiccano due osservatori fra loro complementari: il *Catalog of Digital Scholarly Editions* curato da Patrick Sahle⁴ e il *Catalogue of Digital Editions* curato da Greta Franzini.⁵ Queste importanti iniziative di catalogazione adottano il metodo induttivo per la classificazione delle risorse digitali, dal basso verso l'alto, cioè dalle edizioni digitali effettivamente prodotte nel corso degli anni verso alcune tipologie stabilite in base alle funzionalità effettivamente osservate (Michelone 2021).

I consorzi, come il W3C per il web e il TEI Consortium per le edizioni digitali, suggeriscono linee guida che, in base all'autorevolezza del consorzio stesso, possono diventare standard *de iure* o *de facto*.

² Si veda <http://www.aiucd.it/didattica>.

³ Si veda il book of abstract: <http://amsacta.unibo.it/5559/1/aiucd2016-boa.pdf>

⁴ Bergische Universität Wuppertal: <https://www.digitale-edition.de/exist/apps/editions-browser/about.html>. In questo repertorio *Musisque Deoque* è compreso all'interno della categoria «collection of texts».

⁵ Ospitato dal *Centre of Digital Humanities* dell'University College di Londra, in questo osservatorio il corpus *Musisque Deoque* non viene citato: <https://dig-ed-cat.acdh.oew.ac.at>.

Aderire alle linee guida della TEI, anche se non garantisce la piena interoperabilità semantica, assicura almeno l'interscambio⁶ e favorisce l'uso e il riuso dei dati. Per questo nel corso degli anni non solo si sono moltiplicate le risorse digitali native in formato XML-TEI, ma si è proceduto anche alla conversione di materiali presenti in vasti archivi digitali da formati proprietari a XML-TEI.

Infine, le grandi infrastrutture di ricerca, fra cui principalmente CLARIN-ERIC⁷ per le risorse linguistiche e testuali, non solo definiscono buone pratiche (l'inglese adotta il superlativo: *best practices*) e linee guida, ma forniscono alla comunità scientifica anche gli strumenti per la preservazione a lungo termine degli archivi digitali, per la ricerca federata dei contenuti, per l'esecuzione di analisi linguistiche sulle banche dati testuali e così via.

3 La filologia classica: una disciplina storica da valorizzare nel contesto europeo

Il testo, fenomeno di per sé complesso e sfuggente a una definizione condivisa, è l'oggetto di studio della filologia. La filologia, in particolare la filologia classica, è disciplina storica e interpretativa che si esprime grazie alla realizzazione dell'edizione critica. Non esiste una definizione univoca del modello formale di un'edizione: tipologia ed età del testo, combinate alle scuole nazionali e a filoni di ricerca concorrenti, determinano caratteristiche essenziali anche molto variabili da un'edizione critica all'altra (Trovato 2016, 391). Allo stesso modo forse in modo ancora più acceso, resta il dibattito tra filologi che intendono l'edizione scientifica digitale semplicemente come una diversa forma di presentazione del testo e quelli che ritengono che tutto il processo di creazione di un'edizione critica in formato elettronico debba essere ripensato e riconosciuto come un nuovo processo/metodo scientifico (Mancinelli, Pierazzo 2020).

Le metodologie per il trattamento automatico dei testi, proprie della linguistica computazionale, insieme alle pratiche per la rappresentazione digitale dell'edizione scientifica,⁸ hanno contribuito a delineare gli attuali confini della filologia digitale, che per inserirsi nel panorama europeo dell'Open Science⁹ ha il pressante bisogno di su-

⁶ Si veda in proposito Bauman 2011.

⁷ Il portale di CLARIN è accessibile all'indirizzo <https://www.clarin.eu> mentre il nodo nazionale, CLARIN-IT, è accessibile all'indirizzo <https://www.clarin-it.it>

⁸ Esula dagli obiettivi di questo contributo ripercorrere la storia del dibattito sulla filologia digitale; rimandiamo per una recentissima panoramica a Michelone 2021.

⁹ Le parole del commissario europeo Carlos Moedas nel 2015 (https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_15_5243) hanno dato l'avvio all'odierna politica di Open Science (European Commission 2016) che permea la nuova de-

perare divisioni e frammentazioni per formalizzare modelli e condividere pratiche a livello internazionale. Da quando il nuovo assetto accademico europeo di *Horizon 2020* ha iniziato a spingere al confronto e alla collaborazione multidisciplinare le varie scuole nazionali si è sentita l'esigenza di creare il *Lexicon of Scholarly Edition*.¹⁰ Infatti, il nuovo *Horizon Europe* rende obbligatoria la diffusione e condivisione dei dati al centro della missione della ricerca in modo trasversale per tutte le discipline:¹¹ nel prossimo futuro, dunque, non si parlerà più semplicemente di questioni relative all'accesso aperto dei prodotti della ricerca in termini di 'pubblicazione' o di 'prodotto finito' (ad esempio l'edizione critica digitale), ma si dovrà garantire preservazione di e accesso a tutti i dati emersi/usati dalla ricerca. Nella comunità scientifica europea delle Digital Humanities è allora oramai più che necessaria la creazione di una vera e propria infrastruttura condivisa per gli studi filologici, classici e non.

4 L'importanza dei dati FAIR

Se negli scorsi decenni i dati venivano ricercati e letti per essere rielaborati solo da parte di esseri umani, con l'avvento dei *Linked Open Data* (Bauer, Kaltenböck 2012), prima di giungere al risultato finale interpretabile dallo studioso, sono possibili molti passaggi intermedi gestiti interamente da macchine. Ugualmente, se ai prodromi di imprese di costituzione di biblioteche digitali l'attenzione era incentrata quasi esclusivamente sulla qualità del prodotto finito (in particolare sulla conformità dei documenti digitalizzati agli originali cartacei), oggi la prospettiva si è allargata alla valorizzazione del processo di creazione, preservazione e riuso della risorsa digitale (tracciabilità degli interventi, adozione di *policies* per la conservazione

cade di finanziamenti alla ricerca *Horizon Europe* (https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en).

10 <https://lexiconse.uantwerpen.be/>. Inaugurato da un'idea dell'*European Society of Textual Scholarship*, il progetto ha inteso raccogliere in un unico lexicon le definizioni più autorevoli di concetti che, tradotti da una lingua all'altra – pur nello stesso ambito disciplinare –, possono portare con sé significati intrinseci anche molto differenti. Il sito, in un'ottica europea di Open Science, nel corso del 2020 è stato reingegnerizzato in Jekyll/GitHub e depositato in *Zenodo*, un archivio digitale con specifiche garanzie di preservazione dei dati e opzioni di accessibilità.

11 «The beneficiaries must disseminate their results as soon as feasible, in a publicly available format, subject to any restrictions due to the protection of intellectual property, security rules or legitimate interests. [...] The beneficiaries must ensure open access to peer-reviewed scientific publications relating to their results. [...] The beneficiaries must manage the digital research data generated in the action ('data') responsibly, in line with the FAIR principles and by taking all of the following actions» (HE MGA, v.1.0, art. 17, 108 s.).

a lungo termine e adeguamento agli standard che favoriscono l'interoperabilità). Inoltre, il disaccoppiamento dei dati dagli strumenti di interrogazione offre nuove opportunità d'uso, spesso non previste nei progetti di ricerca iniziali. Si pensi ad esempio all'impiego delle banche dati testuali per studi di linguistica storica o di antropologia del mondo antico, dove nel primo caso i dati potrebbero essere sottoposti a strumenti di analisi morfologica come LEMLAT¹² e nel secondo potrebbero essere gestiti da strumenti di annotazione manuale come *Euporia*.¹³ Per questi scenari complessi, è necessario che i dati testuali e i metadati che li descrivono siano facilmente reperibili, accessibili, interoperabili e riusabili. Questi quattro concetti, che in inglese sono resi con *Findable, Accessible, Interoperable, Reusable*, sono raccolti nell'acronimo FAIR e vengono ben spiegati nell'articolo seminale di Wilkinson et al. (2016).

Perché una risorsa digitale sia reperibile, è necessario che sia associata ad un identificatore univoco e permanente e che i suoi metadati siano completi e indicizzati da motori di ricerca autorevoli. Perché sia accessibile, bisogna che il protocollo di comunicazione usato per trovarla sia standard e sia aperto, e che i metadati siano messi a disposizione anche quando i dati non lo sono (si pensi ad esempio ai riferimenti bibliografici offerti dalle biblioteche digitali anche in assenza del relativo *full text*). Perché sia interoperabile, la risorsa digitale deve usare formati standard (ad esempio XML-TEI), deve impiegare vocabolari condivisi e i (meta)dati devono includere riferimenti qualificati ad altri (meta)dati. Infine, una risorsa, per essere riusabile, deve essere soggetta a un'adeguata licenza d'uso, deve dichiarare in modo dettagliato la propria provenienza e deve seguire gli standard rilevanti all'interno di una specifica comunità scientifica (ad esempio EpiDoc, per la comunità degli epigrafisti digitali).¹⁴

5 Ripetibilità, replicabilità e riproducibilità

I principi FAIR creano le condizioni per l'esercizio della scienza aperta, che consiste nel «condurre la ricerca scientifica in modo completamente trasparente e rendere i risultati disponibili a tutti» (Watson 2015). Così, trasparenza significa fornire al proprio revisore (o ad altre controparti) gli strumenti per poter criticare e falsificare (in senso popperiano) le proprie affermazioni su evidenze oggettive, ripercorrendo tutti i passi della propria ricerca. Non più dunque i so-

¹² <http://www.ilc.cnr.it/lemLat>.

¹³ <https://github.com/CoPhi/euporia>.

¹⁴ <https://epidoc.stoa.org>.

li dati aggregati (le conclusioni dell'indagine), ma l'intero archivio: tutte le risorse.

I risultati sono tanto più solidi quanto più sono ripetibili, replicabili e riproducibili. Nell'accezione di Plesser (2018), un risultato è *ripetibile* se rimane invariato quando viene eseguito dal medesimo gruppo di ricerca nelle medesime condizioni. È *replicabile* quando viene eseguito da gruppi di ricerca differenti sui medesimi dati ed è *riproducibile* quando viene ottenuto da gruppi di ricerca indipendenti senza scambi reciproci.

Declinando questi concetti nell'ambito delle biblioteche digitali di latino, per garantire la ripetibilità dei risultati, ad esempio di un conteggio degli esametri a noi pervenuti prodotti in una certa epoca, è necessario preservare a lungo termine la versione precisa del corpus su cui si è condotto l'esperimento di conteggio. Per garantire la replicabilità, è necessario mettere a disposizione di tutti e non solo del proprio gruppo di ricerca la particolare versione su cui è stato effettuato il conteggio. Infine, per garantire la riproducibilità del risultato, si devono fornire in modo esaustivo i metadati del proprio esperimento di conteggio, in modo che altri, per altra via, possano arrivare a risultati confrontabili (ad esempio su testi digitali ottenuti dal *Perseus Project* oppure dalla collezione di *Musisque Deoque*).

6 Evitare il rischio di una nuova *vulgata*

La trasmissione dei testi nell'era digitale si espone sia agli errori umani tradizionali, sia a nuovi errori dovuti alle peculiarità del mezzo digitale. In questa fase delicata della «Quarta Rivoluzione» (Roncaglia 2010), come sottolinea Mastandrea in Celentano et al. (2004), il rischio di una nuova *vulgata* è molto elevato, se non si prendono le dovute precauzioni.

A dimostrazione del fenomeno, si è scelto qui un caso di studio basato sulla collezione di testi di *Poeti d'Italia in lingua latina*. Sottoposto al sistema di scansione metrica automatica elaborato per il progetto *Pedecerto* da Luigi Tessarolo e Luca Mondin, il verso 105 dell'edizione digitale di Aldegati, *eleg.* 20:

Dulcis Hylas inter ludit et ille choros.

atteso come esametro, è stato correttamente segnalato come ametrico. Fatta la verifica sull'edizione a stampa Bottari (1980), è risultato evidente un *saut du même au même* da parte di uno degli scriventi, curatore dell'edizione digitale delle elegie di Aldegati, in quanto il distico

Dulcis Hylas inter nymphas colludit amantes
Foelices inter ludit et ille choros.

era stato collassato in un unico verso aberrante. Il ripristino del testo metricamente corretto ha comportato anche il ripristino della numerazione dei versi. Ma ciò che interessa qui è notare come l'errore si sia propagato in rete in almeno tre casi: l'errore di collassamento dei versi 105-6 compare su un blog personale, dove l'autore tuttavia cita correttamente la sua fonte digitale.¹⁵ Inoltre, l'errore compare nella biblioteca digitale del *Perseus Project*, in quanto la collezione di *Poeti d'Italia in lingua latina* vi è confluita nel 2011, pur senza una chiara menzione della sua provenienza.¹⁶ Tramite il *Perseus Project*, la collezione di *Poeti d'Italia* (con l'errore qui presentato) è stata poi acquisita da *Corpus Corporum*;¹⁷ anche in questo caso non vi è un link diretto ai responsabili del progetto *Poeti d'Italia in lingua latina*.

7 La galassia *Musisque Deoque*

Con la definizione 'galassia *Musisque Deoque*' indichiamo tutte quelle risorse che dal 2005 in avanti sono state rese disponibili in rete, gratuitamente, grazie all'impegno scientifico del gruppo di ricerca guidato da Paolo Mastandrea e al supporto tecnologico di Luigi Tesarolo. Oltre ai 343.709 versi latini di *Musisque Deoque* (642 opere per 245 diversi autori), rientrano in questa galassia i testi del progetto *Poeti d'Italia in lingua latina* (839 opere su 315 autori per un totale di 486.404 versi) e quelli del progetto *Hellenica*, una piccola raccolta di testi poetici in lingua greca (10 autori per 116.316 versi). Si aggiunge a questi corpora testuali lo strumento *Pedecerto*, che consente un'interpretazione metrico-verbale dei componimenti in esametri e pentametri.

15 G. Dall'Orto, <http://www.giovannidallorto.com/lavori/aldegati.html>. Il testo da: Marci Aldegathi Mantuani elegia [...] come pubblicato online sul sito *Poeti d'Italia in lingua latina*. Viene aggiunta la clausola «l'autore ringrazia [...] chi gli segnalerà eventuali errori» e viene lasciato un recapito di posta elettronica.

16 <https://www.perseus.tufts.edu/hopper/text?doc=Perseus%3atext%3a2011.01.0077>. Tuttavia, sul sito del progetto, l'indicazione nel TEI header non facilita una retroazione diretta dei lettori verso i responsabili scientifici di *Musisque Deoque* e di *Poeti d'Italia in lingua latina*, ma solo verso il Perseus Project.

17 <http://mlat.uzh.ch/?c=15&w=AldMar.Eleg>.

8 Per una breve storia dei progetti della galassia *Musisque Deoque*

Il progetto di Ricerca Nazionale *Musisque Deoque* è stato un'opportunità di dialogo e confronto straordinario per la comunità disciplinare di latinisti-filologi italiani. Grazie alle cordate accademiche nazionali susseguitesesi dal 2005 fino al 2011, la capofila unità veneziana ha di volta in volta riunito professori e giovani ricercatori da svariate università (Trieste, Parma, Padova, Perugia, Roma, Cagliari, Napoli, Calabria, CNR di Pisa)¹⁸ in uno spirito di aggregazione affatto nuovo per le pratiche solipsistiche del filologo latinista tradizionale, forzando ricercatori di formazione squisitamente umanistica al lavoro di squadra e a modalità rigorose di condivisione. Il comune obiettivo di costituire una collezione completa di tutta la poesia latina, integrata e aggiornata da apparati critici ed esegetici elettronici, è stato perseguito grazie alla collaborazione di latinisti di diverse generazioni sotto la guida di Luigi Tessarolo, che fin dagli anni Novanta aveva collaborato a imprese pionieristiche di editoria digitale della Zanichelli, come le prime raccolte di testi digitali in CD-ROM (*Aurea Latinitatis Bibliotheca*, poi *Poesis 1 e 2* quindi *Poetria Nova*). Proprio *Poesis 2* (1999), raccolta di testi digitalizzati già corretti e revisionati negli errori di scansione, è stata la base di partenza su cui è stata costruita la nuova collezione *Musisque Deoque*. Nel corso del primo progetto, tra il 2005/2006 si affrontarono le prime sfide tecniche: è stato necessario – e non è stato compito facile – stabilire le fasi di lavoro e l'infrastruttura di tutto il sistema distribuito. Nel 2005 il concetto di lavoro in *cloud* non era capillarmente diffuso e il contesto accademico italiano era decisamente impreparato a gestire una rete informatica distribuita. Così, il modello informatico di riferimento naturale puntava alla costruzione di un sistema condiviso che raccogliesse e aggiornasse i dati in tempo reale in un unico server. Di fatto si optò per una gestione decisamente separata lato operatore-lato utente, adottando un modello già testato nell'esperienza del precedente progetto *Poeti d'Italia in lingua latina*.¹⁹ Il primo biennio di lavoro fu dedicato alla definizione del protocollo della struttura dati e al design del programma di back-end per l'inserimento/controllo dei testi di base e l'aggiunta di varianti e/o commenti.

18 Si veda l'elenco dei responsabili scientifici e dei principali collaboratori alla pagina <http://mizar.unive.it/mqdq/public/index/collaboratori>.

19 Il progetto finanziato da due PRIN consecutivi (1991 e 2001) aveva curato la creazione di un archivio digitale di testi poetici latini dell'umanesimo e rinascimento, provvedendo alla scansione e correzione di molti versi difficilmente reperibili perché raccolti in testi peregrini o addirittura inediti e trascritti da tesi di laurea (Mastandrea, Pastore Stocchi 2004).

Numerosi e impegnativi furono gli incontri per stabilire una linea comune sulla seconda fase, ovvero lo spoglio variantistico da operarsi nel nuovo apparato elettronico che doveva necessariamente evitare «di riportare la massa dei dati della tradizione manoscritta» ma acquisire solo «varianti 'significative' funzionali alla conoscenza della circolazione dei testi e ai modi della loro ricezione presso altri successivi scrittori» (Mastandrea 2011, 3). La pubblicazione degli atti dei convegni connessi ai vari progetti nazionali²⁰ rendono conto solo in parte della vastità di problemi filologici messi sul piatto nel cercare di raggiungere un compromesso relativo a come doveva intendersi il nuovo apparato elettronico ricercabile in *Musisque Deoque*. Nel 2010, al termine dei due progetti nazionali *Musisque Deoque* (2005 e 2007), il portale – grazie al lavoro di circa 50 collaboratori – contava la revisione e l'aggiunta di 'note d'apparato' per una somma di circa 105.000 versi ovvero circa il 30% del corpus totale.

Nel 2011, con il progetto *Memorata Poetis* ci si rivolse alla sperimentazione di alcune idee di ampliamento e approfondimento dell'infrastruttura tecnologica di *Musisque Deoque*. L'intento era quello di costruire un motore di ricerca che rispondesse a quesiti che si spostassero dalla ricerca lessicale alla ricerca tematica su corpora multilinguistici, fluttuando dalle letterature antiche alle moderne, dalle occidentali alle (medio)orientali.²¹ L'ambizioso progetto si è rivelato un laboratorio dove non vi era più una sola comunità disciplinare – quella dei latinisti – ma un vero drappello di ricercatori legati a filosofie, approcci e metodi afferenti a vari ambiti disciplinari (grecisti, italianisti, arabisti, anglisti...).

Nel 2013 con *Pedecerto*²² la risorsa *Musisque Deoque* è stata integrata con uno strumento per la scansione dei versi dattilici latini che oltre a produrre un catalogo ineccepibile delle casistiche su tutto il corpus poetico latino fino alla tarda antichità²³ offre anche una risorsa per la scansione automatica dei versi dattilici.

20 Mastandrea, Zurli 2009; Mastandrea, Spinazzè 2011.

21 <http://www.memoratapoetis.it/public/>; proprio per la sua ampiezza linguistica e disciplinare, il progetto non rientra nella ristretta galassia *Musisque Deoque*.

22 Il progetto è stato messo a punto dall'Università di Udine nell'ambito del FIRB *Traditio patrum* guidato da Emanuela Colombi come *principal investigator*. Luca Mondin (Ca' Foscari) ha curato gli aspetti scientifici, Luigi Tessoro quelli informatici; la consulenza metricologica è stata fornita da Andrea Bacianini.

23 Si veda lo stato dell'arte nel sito dedicato: <http://www.pedecerto.eu/public/pagine/arte>.

9 La Galassia *Musisque Deoque* nella direzione FAIR

Dal 2006 il sito *Musisque Deoque* (con *Poeti d'Italia in lingua latina*) è un riferimento per gli studiosi: a livello accademico se ne apprezza le potenzialità di supporto all'intertestualità, ma anche i risvolti tecnici. Il motore di ricerca è uno strumento raffinato che permette di ottenere, con elevata qualità ed efficienza, corrispondenze di parole, sillabe e cadenze metriche. Ma è anche nelle ricadute didattiche che la galassia *Musisque Deoque* può ultimamente - a maggior ragione in questi tempi di didattica a distanza - essere rivalutata da un pubblico allargato di docenti di latino delle scuole superiori.²⁴

L'adeguamento delle biblioteche digitali di *Musisque Deoque* e di *Poeti d'Italia in lingua latina* ai principi FAIR è stato condotto per fasi. Il primo intervento operato in questa direzione, datato 2011, è stato la conversione dei documenti della collezione di *Poeti d'Italia in lingua latina*, originariamente marcati in XML, con schema proprietario in XML-TEI P5. Questo ha favorito il disaccoppiamento dei dati dal motore di ricerca e la conseguente inclusione della biblioteca digitale in quella del *Perseus Project*, che ha ampia visibilità internazionale.

Grazie alla creazione di una unità di ricerca dell'Istituto di Linguistica Computazionale 'A. Zampolli' (CNR-ILC) presso il Venice Centre for Digital and Public Humanities (VeDPH) nel 2019, è stata data una forte spinta all'ingresso dei progetti digitali sui testi latini sviluppati a Venezia in CLARIN-IT, il nodo italiano dell'infrastruttura di ricerca per le risorse linguistiche e testuali con sede al CNR-ILC. CLARIN-IT promuove in Italia i principi della scienza aperta in ambito umanistico e l'adeguamento delle banche dati ai principi FAIR (de Jong et al. 2018).

Si è proceduto dunque alla conversione in XML-TEI anche degli apparati critici di *Musisque Deoque* (con il supporto di Daniele Fusi, ricercatore al VeDPH e Angelo Mario Del Grosso, ricercatore al CNR-ILC), all'adozione della licenza aperta Creative Commons Attribution Share-Alike 4.0 International (CC-BY-SA 4.0), all'arricchimento del *teiHeader* di ciascun documento delle collezioni, all'indicazione delle responsabilità dei singoli curatori della digitalizzazione e della versione del documento.

24 Nel biennio 2016-17 Paola Tomè (2015) all'interno della sua borsa Marie Curie curò un'ampia iniziativa di disseminazione degli strumenti digitali applicati alla didattica coinvolgendo in un vero e proprio ambizioso progetto (Greek Studies in XVth Century Europe) tre diversi licei classici del Veneto una biblioteca storica come la Marciana di Venezia e un variegato team accademico che comprendeva grecisti, linguisti computazionali, storici del libro. Vedi anche Manca 2011, Balbo 2013 e Monella 2020. Il progetto, con denominazione 'Wiki-MQDQ', di cui è responsabile Martina Venuti e che vede la stretta collaborazione con CNR-ILC, è stato finanziato dal VeDPH con fondi di premialità: <https://cophilab.ilc.cnr.it/cadmus/mqdq>.

Attualmente è in corso a Venezia, nell'ambito di un progetto sostenuto dal VeDPH (referente Martina Venuti), un processo di trasferimento tecnologico tra Luigi Tassarolo, Daniele Fusi e Angelo Mario Del Grosso per una ristrutturazione architettuale che assicuri il perdurare delle funzionalità e l'estensione nel medio-lungo termine dell'intero sistema *Musisque Deoque*.

Di pari passo con l'aggiornamento di dati e metadati dei corpora testuali della galassia MQDQ in un'ottica FAIR, Daniele Fusi ha ottimizzato per *Musisque Deoque* il suo software nativo *Cadmus*, sviluppando *cadmus-mqdq*²⁵ ovvero l'applicazione online di editing dei testi/apparati che ha come scopo primario il supporto nell'ampliamento del corpus variantistico di *Musisque Deoque*.

Il corpus testuale digitale corredato di dati d'apparato di *Musisque Deoque*²⁶ è invece la fonte per l'analisi, l'elaborazione e il popolamento del *database* del sistema che risiede nella parte architettuale del motore di ricerca *Musisque Deoque* curato da Angelo Mario Del Grosso. Ad oggi il DB è costituito da 44 tabelle suddivise in tre macro aree: la prima sessione relativa allo spoglio del testo e all'indicizzazione; un secondo blocco relativo ai dati d'apparato e alle varianti; infine una terza parte che abbraccia i dati relativi alle fonti, agli autori, alle opere, e in generale ai metadati delle risorse codificate. La nuova architettura disegnata per il sistema *Musisque Deoque* prevede l'uso di GitHub per lo stoccaggio dei dati e il continuo aggiornamento dei contenuti. I diversi servizi che implementano le singole fasi del processo sono incardinati in specifiche applicazioni parte di una architettura più ampia di servizi che prevede l'uso della tecnologia Docker.²⁷

La creazione delle schede dei metadati e il deposito integrale presso CLARIN-IT delle due banche dati²⁸ per *Poeti d'Italia in lingua latina*, ha completato il processo per quanto riguarda l'adeguamento ai principi FAIR, con la consulenza scientifica di Monica Monachini, responsabile di CLARIN-IT, e Riccardo Del Gratta, ricercatore del CNR-ILC, che già aveva curato insieme a Federico Boschetti l'estrazione dei metadati dagli archivi di ALIM per la fruizione tramite il *Virtual Language Observatory* di CLARIN (Boschetti et al. 2021).

L'interrogazione dei dati non solo tramite interfaccia grafica per l'interazione uomo-macchina ma anche attraverso Web API per l'inte-

²⁵ <https://cophilab.ilc.cnr.it/cadmus/mqdq>.

²⁶ Ad oggi vi sono 153.896 versi interessati dall'apparato su un totale di 353.096 versi.

²⁷ Si tratta di una complessa tecnologia sviluppata negli ultimi anni in grado di semplificare notevolmente la distribuzione delle applicazioni software in modo da rendere efficiente la condivisione e l'installazione in ambienti collaborativi.

²⁸ Consultabili rispettivamente al link <https://hdl.handle.net/20.500.11752/OPEN-555> per *Musisque Deoque* e <https://hdl.handle.net/20.500.11752/OPEN-556>

razione macchina-macchina favorirà inoltre l'ingresso di questi strumenti nell'ecosistema dei Linked Open Data, delle catene di analisi messe a disposizione da CLARIN tramite ad esempio *WebLicht*,²⁹ e della Federated Content Search,³⁰ che permette la ricerca di occorrenze all'interno di una o più risorse linguistiche e testuali messe a disposizione di tutti i centri nazionali CLARIN.

10 Conclusione: un bilancio e una prospettiva futura

I progetti della galassia *Musisque Deoque* sono nati per rispondere a domande di ricerca aperte e rilevanti nell'ambito degli studi filologici e letterari, come i meccanismi dell'intertestualità, la genesi delle varianti significative, le modalità di circolazione dei testi in prospettiva sincronica e diacronica. Per questo motivo le risorse digitali e gli strumenti di interrogazione che sono stati sviluppati nel corso dei decenni avevano come principale pubblico di riferimento la comunità scientifica degli studiosi di letteratura e dei filologi.

Quando si è scelto di scalare dalla dimensione nazionale (in termini di visibilità e finanziamenti) alla dimensione internazionale, i principali interlocutori appartenevano alla comunità delle Digital Humanities. Per questo è stato necessario avviare un processo di standardizzazione dei formati, di definizione di licenze aperte e, infine, di ricorso all'infrastruttura di ricerca CLARIN-IT. La galassia *Musisque Deoque* nel corso del tempo si è evoluta in un archivio digitale in linea con i principi FAIR di reperibilità, accessibilità, interoperabilità e riusabilità. La prospettiva futura non può che andare in questa direzione.

²⁹ <https://weblicht.sfs.uni-tuebingen.de>.

³⁰ <https://www.clarin.eu/content/federated-content-search-clarin-fcs>.

Bibliografia

- Balbo, A. (2013). «La nuvola greca e latina: rete, cloud computing e antichità classica nel XXI secolo». Pagnotta, F. (a cura di), *L'età di Internet. Umanità, cultura, educazione*. Milano, 53-68.
- Bauer, F.; Kaltenböck, M. (2012). *Linked Open Data: The Essentials: A Quick Start Guide for Decision Makers*. Wien.
- Bauman, S. (2011). «Interchange vs. Interoperability». *Balisage: The Markup Conference 2011* (Montréal, 2-5 August 2011). Montréal. <https://doi.org/10.4242/BalisageVol7.Bauman01>.
- Boschetti, F. et al. (2021). «“Tea for Two”: The Archive of the Italian Latinity of the Middle Ages Meets the CLARIN Infrastructure». Navarretta, C.; Eskevich, M. (eds), *Selected Papers from the CLARIN Annual Conference 2020*. Linköping Electronic Conference Proceedings 180, 37-46 <https://doi.org/10.3384/ecp1805>.
- Bottari, G. (1980). *Marcantonio Aldegati, poeta latino del Quattrocento*. Palermo.
- Bozzi, A. (2019). *Percorsi di linguistica e di filologia computazionali*. A cura di M.S. Corradini Bozzi e G. Ferrari. Pisa.
- Busa, R. (1990). «Informatics and New Philology». *Computers and the Humanities*, 24(5/6), 339-43.
- Buzzetti, D.; McGann, J. (2006). «Electronic Textual Editing: Critical Editing in a Digital Horizon». Burnard, L. et al. (eds), *Electronic Textual Editing*. <https://tei-c.org/Vault/ETE/Preview/mcgann.html>.
- Celentano, A.; Cortesi, A.; Mastandrea, P. (2004). «Informatica umanistica: una disciplina di confine». *Mondo Digitale*, 4, 44-55.
- De Jong, F. et al. (2018). «CLARIN: Towards FAIR and Responsible Data Science Using Language Resources». Calzolari, N. et al. (eds), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (7-12 May 2018). Miyazaki (Japan): European Language Resources Association. <https://www.aclweb.org/anthology/L18-1515>.
- Europäische Kommission (2016). *Open Innovation, Open Science, Open to the World: A Vision for Europe*. Luxembourg.
- Gigliozzi, G. (1997). *Il testo e il computer: manuale di informatica per gli studi letterari*. Milano.
- Maas, P. (1952). *Critica del Testo*. Translated by N. Martinello. Firenze.
- Manca, M. (2011). «Come usare (e non usare) i computer nella didattica dell'antico». Perrelli, R.; Mastandrea, P. (a cura di), *Latinum est, et legitur: metodi e temi dello studio dei testi latini*. Amsterdam.
- Mancinelli, T.; Pierazzo, E. (2020). *Che cos'è un'edizione scientifica digitale*. Roma.
- Mastandrea, P. (a cura di) (1999). *Poesis 2: CD-ROM dei testi della poesia latina*. Bologna.
- Mastandrea, P. (2009). «Gli archivi elettronici di *Musisque Deoque*. Ricerca intertestuale e cernita fra varianti antiche (con qualche ripensamento sulla tradizione indiretta dei poeti latini)». Mastandrea, P.; Zurli, L. (a cura di), *Poesia latina. Nuova e-filologia. Opportunità per l'editore e per l'interprete*. Roma, 41-72.
- Mastandrea, P.; Pastore Stocchi, M. (2004). «Poeti d'Italia in lingua latina. Un archivio elettronico da Dante al primo trentennio del XVI secolo». Bernardi Perini, G. (a cura di), *Il Latino nell'età dell'Umanesimo*. Firenze, 35-50.

- Mastandrea, P.; Spinazzè, L. (a cura di) (2011). *Nuovi archivi e mezzi d'analisi per i testi poetici latini = Un consuntivo dei lavori del programma 'Musisque Deoque'* (Venezia, 21-23 giugno 2010). Amsterdam.
- Mastandrea, P.; Tassarolo, L. (a cura di) (2010). *Poetria Nova. A CD-ROM of Latin Medieval Poetry (650-1250 A.D.) with a Gateway to Classical and Late Antiquity Texts*. Firenze.
- Mastandrea, P.; Tassarolo, L. (2011). «Introduzione». Mastandrea, P.; Spinazzè, L. (a cura di), *Nuovi archivi e mezzi d'analisi per i testi poetici latini. Un consuntivo dei lavori del programma 'Musisque Deoque'* (Venezia, 21-23 giugno 2010). Amsterdam, 1-11.
- Mastandrea, P.; Tassarolo, L. (2014). «Da *Musisque Deoque* a *Memorata Poetis*. Le vie della ricerca intertestuale». Tomasi, F.; Agosti, M. (a cura di), *Collaborative Research Practices and Shared Infrastructures for Humanities Computing*. Padova.
- Mastandrea, P.; Zurli, L. (a cura di) (2009). *Poesia latina. Nuova e-filologia. Opportunità per l'editore e per l'interprete*. Roma.
- Michelone, F. (2021). «The Critical Edition Between Digital and Print: Methodological Considerations». *Umanistica Digitale*, 10, 25-48. <https://doi.org/10.6092/issn.2532-8816/12626>.
- Monella, P. (2020). *Metodi digitali per l'insegnamento classico e umanistico*. Milano.
- Orlandi, T. (2010). *Informatica testuale: teoria e prassi*. Roma.
- Pasquali, G. (1952²). *Storia della tradizione e critica del testo*. 2. ed. con nuova prefazione e aggiunta di tre appendici. Firenze.
- Perrelli, R.; Mastandrea, P. (a cura di) (2011). *Latinum est, et legitur: metodi e temi dello studio dei testi latini*. Amsterdam.
- Plessner, H.E. (2018). «Reproducibility vs. Replicability: A Brief History of a Confused Terminology». *Frontiers in Neuroinformatics*, 11, 76. <https://doi.org/10.3389/fninf.2017.00076>.
- Roncaglia, G. (2010). *La quarta rivoluzione: sei lezioni sul futuro del libro*. Roma.
- Tomè, P. (2015). «“Greek Studies in XVth Century Europe”. Progetto di disseminazione della ricerca Marie Curie con alcuni licei classici del Veneto». Spinazzè, L. (a cura di), *Filologia digitale dalla ricerca alla didattica*. Trento, 89-94.
- Trovato, P. (2016). «Come pubblicare i testi di pellegrinaggio. Edizioni storiche vs edizioni letterarie o semplicemente buone edizioni?». *Nuova Rivista Storica*, 100(2), 391-420. <https://doi.org/10.1400/242805>.
- Watson, M. (2015). «When Will 'Open Science' Become Simply 'Science'?». *Genome Biology*, 16(1), 101. <https://doi.org/10.1186/s13059-015-0669-2>.
- Wilkinson, M.D. (2016). «The FAIR Guiding Principles for Scientific Data Management and Stewardship». *Scientific Data*, 3. <https://doi.org/10.1038/sdata.2016.18>.

