

Atti del IX Convegno Annuale
dell'Associazione per l'Informatica
Umanistica e la Cultura Digitale
(AIUCD)

LA SVOLTA INEVITABILE:
SFIDE E PROSPETTIVE PER
L'INFORMATICA UMANISTICA

15 – 17 gennaio 2020
Milano
Università Cattolica del Sacro Cuore

A CURA DI:
Cristina Marras
Marco Passarotti
Greta Franzini
Eleonora Litta

ISBN: 978-88-942535-4-2

ASSOCIAZIONE per
l'INFORMATICA UMANISTICA
e la CULTURA DIGITALE



Copyright © 2020

Associazione per l'Informatica Umanistica e la Cultura Digitale

Copyright of each individual chapter is maintained by the authors.

This work is licensed under a Creative Commons Attribution Share-Alike 4.0 International license (CC-BY-SA 4.0). This license allows you to share, copy, distribute and transmit the text; to adapt the text and to make commercial use of the text providing attribution is made to the authors (but not in any way that suggests that they endorse you or your use of the work). Attribution should include the following information:

Cristina Marras, Marco Passarotti, Greta Franzini, Eleonora Litta (a cura di),
Atti del IX Convegno Annuale AIUCD. La svolta inevitabile:
sfide e prospettive per l'Informatica Umanistica.

Available online as a supplement of Umanistica Digitale: <https://umanisticadigitale.unibo.it>

All links were visited on 29th December 2019, unless otherwise indicated.

Every effort has been made to identify and contact copyright holders and any omission or error will be corrected if notified to the editors.

Prefazione

La nona edizione del convegno annuale dell'Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD 2020; Milano, 15-17 gennaio 2020) ha come tema “La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica”, con lo specifico obiettivo di fornire un'occasione per riflettere sulle conseguenze della crescente diffusione dell'approccio computazionale al trattamento dei dati connessi all'ambito umanistico. Questo volume raccoglie gli articoli i cui contenuti sono stati presentati al convegno. A diversa stregua, essi affrontano il tema proposto da un punto di vista ora più teorico-metodologico, ora più empirico-pratico, presentando i risultati di lavori e progetti (conclusi o in corso) che considerino centrale il trattamento computazionale dei dati.

Dunque, la svolta inevitabile qui a tema va intesa innanzitutto come metodologica e, più nello specifico, computazionale. Ad essa la ricerca umanistica contemporanea assiste, con diversi gradi di accoglienza, critica addirittura rifiuto. La computabilità del dato empirico (anche) in area umanistica è, infatti, il tratto distintivo e il vero valore aggiunto che le innovazioni tecnologiche degli ultimi decenni hanno comportato in questo ambito. Nonostante negli anni il settore delle cosiddette Digital Humanities si sia voluto caratterizzare, anche a partire dalla propria denominazione, insistendo maggiormente sull'aspetto digitale che non su quello computazionale, i tempi sembrano ormai maturi perché il termine Computational Humanities, o il troppo precocemente accantonato Humanities Computing, (ri)prenda il posto oggi ancora occupato da Digital Humanities.¹ Digitale è, infatti, il formato dei dati con cui attualmente si ha in gran parte a che fare nel nostro settore: ma è computazionale l'uso che di questi dati si fa ed è un fatto che gran parte dei lavori prodotti nell'area delle Digital Humanities consista nel “fare conti” sui dati.²

Come tanti suoi predecessori, anche il formato digitale passerà; mentre il metodo, e la svolta che esso comporta, resterà, perché solidamente ancorato all'evidenza empirica del dato che è il punto di partenza e, quindi, il centro di analisi di molta ricerca umanistica. Per questa ragione, la svolta computazionale nelle scienze umanistiche è innanzitutto metodologica: a cambiare radicalmente non è tanto il formato dei dati, ma il modo con cui ad essi ci si approccia e l'uso che di essi si fa.

Non va negato un certo scetticismo reazionario che, ora esplicito, ora sottaciuto, parte del mondo della ricerca umanistica nutre nei confronti dei metodi e degli strumenti che la svolta computazionale ha messo a disposizione di noi ricercatori, che viviamo l'attuale scorcio di storia della scienza. Negli anni, tale scetticismo ha alimentato una irragionevole distinzione, e conseguente separazione, tra umanisti “tradizionali” e umanisti “digitali”, quasi che si debbano identificare due aree al fine di evitare che gli uni infastidiscano troppo gli altri con le proprie ricerche, trascurando che esse trattano i medesimi oggetti e hanno quale fine comune la produzione di nuova conoscenza.

Siffatta separazione è dovuta a errori imputabili all'una e all'altra parte. Da un lato, certi umanisti “digitali” tendono a produrre ricerca che rischia di scadere nella superficialità, assumendo che l'alta quantità dei dati trattati possa compensarne l'eventuale bassa qualità e dimenticando, così, che le ricerche di area umanistica molto raramente lavorano su Big Data e non possono (anzi, non vogliono) accontentarsi di tendenze percentuali fondate su dati imprecisi. Dall'altro lato, i “tradizionali” sono spesso afflitti da un conservatorismo protezionista incompatibile con la natura stessa del lavoro di ricerca, che è in sé progressivo e in costante evoluzione. Ne consegue un dialogo interrotto tra le due parti: i “digitali” sono considerati dei tecnici (inteso in senso riduttivo) che brutalizzano il delicato dato umanistico, mentre i “tradizionali” vengono derubricati a dinosauri incartapecoriti che ormai non hanno più niente di nuovo da dire.

Ma la svolta computazionale non è né “digitale”, né “tradizionale”. Semplicemente, essa è inevitabile. Chi ne fa cattivo uso, come certo mondo “digitale”, non sa valorizzarne la forza della portata; chi la rifiuta a priori, si pone fuori dalla realtà e, volutamente ignorando il nuovo, ferisce la ragione stessa del far ricerca.

¹ Una valida sintesi della questione relativa alla denominazione del settore, con una buona bibliografia a supporto, è riportata in un articolo di Leah Henrikson pubblicato su 3:AM Magazine (24 Ottobre 2019) e disponibile presso <https://www.3amagazine.com/3am/humanities-computing-digital-humanities-and-computational-humanities-whats-in-a-name/>

² Da, Nan Z. “The computational case against computational literary studies.” *Critical Inquiry* 45.3 (2019): 601-639.

Ma resta che la svolta è inevitabile: non si comprende perché sul tavolo dell'umanista del 2020 non possano trovarsi al contempo un'edizione critica cartacea e i risultati di un analizzatore morfologico automatico proiettati sullo schermo di un computer. Entrambi sono strumenti che diversamente trattano il comune oggetto d'interesse di tanta ricerca, ovvero i dati.

Ma di una svolta non solo metodologica questa edizione 2020 del convegno AIUCD vuole trattare e farsi carico, aspirando anzi a mettere in atto anche una piccola, ma sostanziale svolta organizzativa. Per la prima volta, la call for papers di un convegno dell'Associazione, ha richiesto l'invio non di abstract, ma di articoli completi della lunghezza di un massimo di 4 pagine (bibliografia esclusa). Di concerto con il Comitato Direttivo dell'Associazione, abbiamo deciso di orientarci in tal senso per due ragioni principali. Primo, crediamo che, giunto alla propria nona edizione, il convegno annuale della AIUCD sia ormai sufficientemente maturo per passare a una fase il cui obiettivo sia quello di accogliere nel programma del convegno proposte che nel formato dell'articolo completo consentissero ai revisori una valutazione piena e più accurata. Ciò si lega anche alla seconda ragione. Il nostro settore come è noto è molto veloce: i dati (e i risultati su di essi basati) tendono a cambiare nel giro di poco tempo. Ricevere articoli completi ci ha consentito di mettere i contenuti del presente volume nelle mani dei partecipanti (e più in generale della comunità tutta) il primo giorno del convegno, fornendo così una realistica fotografia dello stato dei lavori al gennaio 2020.

Tutti gli articoli selezionati per essere presentati al convegno hanno cittadinanza in questo volume. Anche questa è una svolta: diversamente dall'uso fino ad oggi adottato, gli articoli pubblicati non sono più il risultato di una selezione a posteriori rispetto al convegno, ma tutti quelli effettivamente apparsi nel programma di AIUCD 2020. In tal senso, una certa esclusività promossa a livello di selezione scientifica si fa inclusività in termini di pubblicazione e, dunque di visibilità dei lavori presentati. Ogni proposta è stata valutata da tre revisori; si è dovuto ricorrere a una quarta valutazione solo nel caso di due proposte su cui i tre revisori avevano espresso opinioni che rendevano difficile prendere una decisione in merito alla loro accettazione, o meno. Al proposito delle differenze tra i revisori, abbiamo constatato divergenze piuttosto frequenti e, in alcuni casi, nette tra coloro che provengono dall'area linguistico-computazionale e quanti, invece, sono a vario titolo legati ai diversi settori dell' "umanistica digitale". Mentre i linguisti computazionali sono tradizionalmente usi a valutare articoli completi e tendono a richiedere che i contenuti di essi descrivano motivazioni, metodi e risultati (preferibilmente replicabili) di lavori di ricerca in corso, o completati, i revisori di area umanistico-digitale sono disposti a valutare positivamente anche idee e proposte che ancora non si siano incarnate in una reale applicazione ai dati. La constatazione di tale diversità è il risultato della composizione volutamente inter- e trans-disciplinare del comitato dei revisori, a rappresentare la natura trasversale di AIUCD e, di riflesso, del suo convegno annuale. Nel prendere le decisioni in merito alle proposte, abbiamo cercato un equilibrio tra gli atteggiamenti delle due parti, favoriti dall'aver a disposizione un livello di dettaglio sul lavoro descritto. La richiesta di articoli completi ha avuto un impatto non molto rilevante sul numero delle proposte inviate, che sono state 71, di cui 67 sottoposte al processo di revisione, mentre 4 sono state escluse perché non confacenti ai criteri richiesti dalla call for papers (tra cui anonimato e originalità). Alla precedente edizione del convegno AIUCD (Udine, 23-25 gennaio 2019) erano state inviate 82 proposte, di cui 75 sottoposte a revisione. Conseguenze più sostanziali si sono, invece, riscontrate sulla percentuale delle proposte accettate e rifiutate. Delle 67 proposte valutate, 45 sono state accettate per apparire nel programma del convegno e, quindi, in questo volume, mentre 22 sono state rifiutate, risultando così in una percentuale di accettazione pari al 67.16%. All'edizione udinese, la percentuale si era attestata intorno all'84%. La contrazione del numero di proposte accettate è strettamente connessa alla richiesta di articoli completi invece che di abstract.

Il programma del convegno ha incluso due sessioni poster. Dei 45 contributi accettati, 21 sono stati giudicati adatti alla presentazione in modalità poster. Rispetto alle consuetudini del settore, che tende a relegare le proposte meno interessanti o più problematiche nelle sessioni poster, abbiamo deciso di assegnare la modalità di comunicazione in forma di poster non secondo la qualità, ma piuttosto in base alla tipologia della proposta. Dunque, tendenzialmente le proposte che presentano lavori che hanno portato a risultati pratici (come strumenti, risorse, o interfacce) sono state giudicate più adatte a una presentazione in formato poster, mentre le discussioni teoriche, disciplinari, o metodologiche hanno occupato le sessioni di comunicazioni orali. Resta che non sussiste differenza alcuna in termini di selezione qualitativa tra un articolo i cui contenuti sono stati

presentati al convegno in forma orale, o in forma di poster, come dimostra l'aver riservato il medesimo numero di pagine a tutti gli articoli presenti in questo volume.

I contenuti dei testi qui raccolti in ordine alfabetico testimoniano la varietà dei temi che usualmente sono trattati nei convegni della AIUCD. Essi spaziano da riflessioni generali sul settore di ricerca alla realizzazione di nuove risorse linguistiche e strumenti di analisi dei dati, da lavori di filologia ed editoria digitale a temi connessi alla digitalizzazione delle fonti in ambito bibliotecario. Oltre alla presentazione dei contenuti degli articoli di questo volume, il programma del convegno ha previsto tre relazioni su invito (una per ciascuno dei tre giorni della sua durata), che sono state rispettivamente presentate da Roberto Navigli (Sapienza, Università di Roma), Julianne Nyhan (University College London) e Steven Jones (University of South Florida).

Il contributo di Roberto Navigli, intitolato *Every time I hire a linguist my performance goes up (or: the quest for multilingual lexical knowledge in a deep (learning) world)*, è un esempio di ricerca che dice della ineludibilità del legame e, auspicabilmente, della collaborazione tra mondo scientifico e mondo umanistico e, nello specifico, tra la comunità che si riconosce nella AIUCD e quella della linguistica computazionale. Gli interventi di Julianne Nyhan (*Where does the history of the Digital Humanities fit in the longer history of the Humanities? Reflections on the historiography of the 'old' in the work of Fr Roberto Busa S.J.*) e Steven Jones (*Digging into CAAL: Father Roberto Busa's Center and the Prehistory of the Digital Humanities*) si posizionano nell'alveo della storia della disciplina, particolarmente riferendo in merito ai loro studi sulle attività di padre Roberto Busa. La figura di Busa è strettamente legata all'Università Cattolica del Sacro Cuore di Milano, dove a partire dalla fine degli anni settanta il gesuita tenne un corso di Linguistica Computazionale e Matematica e fondò un gruppo di ricerca che, nel 2009, fu trasformato in un Centro di Ricerca; quel CIRCSE che con l'AIUCD ha organizzato il convegno annuale dell'associazione di cui questo volume raccoglie gli Atti. Nel 2010, un anno prima di lasciarci, padre Busa volle donare alla Biblioteca della Cattolica il proprio archivio personale. Una ricchissima documentazione del lavoro di Busa e della sua diffusione, oltre che delle sue relazioni personali e professionali (ricostruibili attraverso il vasto epistolario), l'Archivio Busa è attualmente in fase di catalogazione e digitalizzazione da parte della Biblioteca d'Ateneo. Una selezione di materiale tratto dall'Archivio è stata resa direttamente accessibile ai partecipanti dell'edizione milanese del convegno AIUCD in una piccola mostra allestita nell'atrio dell'aula dei lavori congressuali. Le teche della mostra raccolgono fogli di lavoro, lettere, schede perforate, nastri e articoli di quotidiani che trattano del lavoro di padre Busa: una forma di ringraziamento che l'Università Cattolica, il CIRCSE e la comunità scientifica tutta vuole riservare a uno dei pionieri dell'analisi linguistica automatica.

I nostri ringraziamenti vanno innanzitutto alla Presidente di AIUCD Francesca Tomasi e a Fabio Ciotti, che in quel ruolo l'ha preceduta, per aver scelto Milano quale sede dell'edizione 2020 del convegno. Da loro è venuto il primo, fondamentale, sostegno alla "svolta organizzativa" di cui abbiamo voluto farci portatori. Ringraziamo altresì il Consiglio Direttivo dell'Associazione, il Comitato di Programma e tutti i revisori, che hanno lavorato alacremente per metterci nelle condizioni di definire il miglior programma possibile. La sede milanese dell'Università Cattolica del Sacro Cuore ci ha supportato a livello amministrativo e logistico; teniamo particolarmente a ringraziare l'Ufficio Formazione Permanente, nello specifico di Elisa Ballerini, la Biblioteca d'Ateneo, e specificatamente Paolo Senna, che ci ha messo a disposizione i materiali dell'Archivio Busa, l'Ufficio Eventi e la Direzione di Sede, che hanno fornito gli spazi per il convegno. Grazie soprattutto a chi ha inviato proposte, ai relatori e ai partecipanti tutti, perché sono loro i protagonisti essenziali dell'evento.

La nostra speranza è che il lavoro fatto sia utile ancora prima che apprezzato. E che i suoi risultati si mantengano nelle edizioni a venire, con l'obiettivo di migliorare sempre, guardando avanti; perché saper vedere le svolte e affrontarle è la ragione stessa della ricerca.

Cristina Marras
Marco Passarotti
Greta Franzini
Eleonora Litta

Chair e Comitati

General Chair

- Cristina Marras

Chair del comitato scientifico e di programma

- Marco Passarotti

Comitato scientifico e di programma

- Maristella Agosti
- Stefano Allegrezza
- Federica Bressan
- Cristiano Chesi
- Fabio Ciraci
- Greta Franzini
- Angelo Mario Del Grosso
- Eleonora Litta
- Pietro Maria Liuzzo
- Federico Meschini
- Johanna Monti
- Federico Nanni
- Marianna Nicolosi
- Dario Rodighiero
- Marco Rospocher
- Chiara Zuanni

Comitato Organizzatore

- Greta Franzini
- Eleonora Litta

Indice dei Contenuti

| | |
|--|-----------|
| EcoDigit-Ecosistema Digitale per la fruizione e la valorizzazione dei beni e delle attività culturali del Lazio | 1 |
| Luigi Asprino, Antonio Budano, Marco Canciani, Luisa Carbone, Miguel Ceriani, Ludovica Marinucci, Massimo Mecella, Federico Meschini, Marialuisa Mongelli, Andrea Giovanni Nuzzolese, Valentina Presutti, Marco Puccini, Mauro Saccone | |
| Encoding the Critical Apparatus by Domain Specific Languages: The Case of the Hebrew Book of Qohelet | 7 |
| Luigi Bambaci, Federico Boschetti | |
| 600 maestri raccontano la loro vita professionale in video: un progetto di (fully searchable) open data | 14 |
| Gianfranco Bandini, Andrea Mangiatordi | |
| Ripensare i dati come risorse digitali: un processo difficile? | 19 |
| Nicola Barbuti | |
| Verso il riconoscimento delle Digital Humanities come Area Scientifica: il catalogo online condiviso delle pubblicazioni dell'AIUCD | 24 |
| Nicola Barbuti, Maurizio Lana, Vittore Casarosa | |
| Il trattamento automatico del linguaggio applicato all'italiano volgare. La redazione di un <i>formario</i> tratto dalle prime dieci <i>Lettere</i> di Alessandra M. Strozzi | 28 |
| Ottavia Bersano, Nadezda Okinina | |
| Annotazione semantica e visualizzazione di un corpus di corrispondenze di guerra | 34 |
| Beatrice Dal Bo, Francesca Frontini, Giancarlo Luxardo | |
| The Use of Parallel Corpora for a Contrastive (Russian-Italian) Description of Resource Markers: New Instruments Compared to Traditional Lexicography | 39 |
| Anna Bonola, Valentina Nosedà | |
| PhiloEditor: Simplified HTML Markup for Interpretative Pathways over Literary Collections | 47 |
| Claudia Bonsi, Angelo Di Iorio, Paola Italia, Francesca Tomasi, Fabio Vitali, Ersilia Russo | |
| An Empirical Study of Versioning in Digital Scholarly Editions | 55 |
| Martina Bürgermeister | |

| | |
|--|------------|
| ELA: fasi del progetto, bilanci e prospettive Emmanuela Carbé, Nicola Giannelli | 61 |
| Digitized and Digitalized Humanities: Words and Identity Claire Clivaz | 67 |
| La geolinguistica digitale e le sfide lessicografiche nell'era delle digital humanities: l'esempio di VerbaAlpina Beatrice Colcuc | 74 |
| Una proposta di ontologia basata su RDA per il patrimonio culturale di Vincenzo Bellini Salvatore Cristofaro, Daria Spampinato | 82 |
| Biblioteche di conservazione e libera fruizione dei manoscritti digitalizzati: la Veneranda Biblioteca Ambrosiana e la svolta inevitabile grazie a IIF Fabio Cusimano | 89 |
| Repertori terminologici plurilingui fra normatività e uso nella comunicazione digitale istituzionale e professionale Klara Dankova, Silvia Calvi | 98 |
| The Digital Lexicon Translaticium Latinum: Theoretical and Methodological Issues Chiara Fedriani, Irene De Felice, William Michael Short | 106 |
| Selling Autograph Manuscripts in 19th c. Paris: Digitising the Revue des Autographes Simon Gabay, Lucie Rondeau du Noyer, Mohamed Khemakhem | 113 |
| Enriching a Multilingual Terminology Exploiting Parallel Texts: an Experiment on the Italian Translation of the Babylonian Talmud Angelo Mario Del Grosso, Emiliano Giovannetti, Simone Marchi | 119 |
| Towards a Lexical Standard for the Representation of Etymological Data Fahad Khan, Jack Bowers | 125 |
| Workflows, Digital Data Management and Curation in the RETOPEA Project Ilenia Eleonor Laudito | 130 |

| | |
|---|------------|
| Il confronto con Wikipedia come occasione di valorizzazione professionale: il case study di Biblioteca digitale BEIC Lisa Longhi | 136 |
| Making a Digital Edition: The Petrarchive Project Isabella Magni | 142 |
| Extending the DSE: LOD Support and TEI/IIIF Integration in EVT Paolo Monella, Roberto Rosselli Del Turco | 148 |
| Mapping as a Contemporary Instrument for Orientation in Conferences Chloe Ye-Eun Moon, Dario Rodighiero | 156 |
| Argumentation Mapping for the History of Philosophical and Scientific Ideas: The TheSu Annotation Scheme and its Application to Plutarch's Aquane an ignis Daniele Morrone | 163 |
| Leitwort Detection, Quantification and Discernment Racheli Moskowitz, Moriyah Schick, Joshua Waxman | 171 |
| From Copies to an Original: The Contribution of Statistical Methods Amanda Murphy, Raffaella Zardoni, Felicita Mornata | 178 |
| FORMAL. Mapping Fountains over Time and Place. Mappare il movimento delle fontane monumentali nel tempo e nello spazio attraverso la geovisualizzazione Pamela Palomba, Emanuele Garzia, Roberto Montanari | 185 |
| Paul is Dead? Differences and Similarities before and after Paul McCartney's Supposed Death. Stylometric Analysis of Transcribed Interviews Antonio Pascucci, Raffaele Manna, Vincenzo Masucci, Johanna Monti | 191 |
| Digital Projects for Music Research and Education from the Center for Music Research and Documentation (CIDoM), Associated Unit of the Spanish National Research Council Juan José Pastor Comín, Francisco Manuel López Gómez | 198 |
| Prospects for Computational Hermeneutics Michael Piotrowski, Markus Neuwirth | 204 |
| EModSar: A Corpus of Early Modern Sardinian Texts Nicoletta Puddu, Luigi Talamo | 210 |
| Shared Emotions in Reading Pirandello. An Experiment with Sentiment Analysis Simone Reborà | 216 |
| DH as an Ideal Educational Environment: The Ethnographic Museum of La Spezia Letizia Ricci, Francesco Melighetti, Federico Boschetti, Angelo Mario Del Grosso, Enrica Salvatori | 222 |

| | |
|--|------------|
| A Digital Review of Critical Editions: A Case Study on Sophocles, Ajax 1-332 Camilla Rossini | 227 |
| Strategie e metodi per il recupero di dizionari storici Eva Sassolini, Marco Biffi | 235 |
| Encoding Byzantine Seals: SigiDoc Alessio Sopracasa, Martina Filosa | 240 |
| Preliminary Results on Mapping Digital Humanities Research Gianmarco Spinaci, Giovanni Colavizza, Silvio Peroni | 246 |
| Epistolario De Gasperi: National Edition of De Gasperi's Letters in Digital Format Sara Tonelli, Rachele Sprugnoli, Giovanni Moretti, Stefano Malfatti, Marco Odorizzi | 253 |
| Visualizing Romanesco; or, Old Data, New Insights Gianluca Valenti | 260 |
| What is a Last Letter? A Linguistics/Preventive Analysis of Prisoner Letters from the Two World Wars Giovanni Pietro Vitali | 265 |
| L'organizzazione e la descrizione di un fondo nativo digitale: PAD e l'Archivio Franco Buffoni Paul Gabriele Weston, Primo Baldini, Laura Pusterla | 273 |

Ripensare i dati come risorse digitali: un processo difficile?

Nicola Barbuti

Università degli Studi di Bari Aldo Moro
nicola.barbuti@uniba.it

Abstract

English. The Art. 2 of the *EU Council Conclusions of 21 May 2014 on cultural heritage as a strategic resource for a sustainable Europe (2014 / C 183/08)* recognizes the existence of the new digital cultural heritage (*born digital* and *digitized*). Starting from this assumption, we need to rethink digital and digitization as social and cultural expressions of the contemporary age. Digital resources shall record and represent both digitization processes and themselves in their life cycle, they are no longer mere gateway to improve the access to reality. So, we have to define clear and homogeneous criteria to validate and certify them as a memory and sources of knowledge for future generations. In this regard, the present paper outlines a first proposal for identification of digital cultural resources, based on the expansion of the *R: Reusable* of the *FAIR Principles for the management of scientific metadata* in *R⁵: Reusable, Readable, Relevant, Reliable and Resilient*.

Italiano. L'art. 2 delle *Conclusioni del Consiglio UE del 21 maggio 2014 sul patrimonio culturale come risorsa strategica per un'Europa sostenibile (2014 / C 183/08)* riconosce l'esistenza del nuovo patrimonio culturale digitale (*born digital* e *digitalizzato*). Partendo da questo assunto, si rende indispensabile ripensare il digitale e la digitalizzazione come espressioni sociali e culturali dell'età contemporanea. È necessario, perciò, riconsiderare le risorse digitali come registrazioni e rappresentazioni di processi, e non più come semplici mediatori atti a migliorare l'accesso alla realtà, definendo criteri chiari e omogenei per convalidarli e certificarli come memoria e fonti di conoscenza per le generazioni future. A riguardo, nel presente lavoro si delinea una prima proposta, basata sull'ampliamento della *R: Re-usable* dei *FAIR Principles* per la gestione dei metadati scientifici in *R⁵: Reusable, Readable, Relevant, Reliable and Resilient*.

1 Introduzione

L'art. 2 delle *Conclusioni del Consiglio dell'UE del 21 maggio 2014 sul patrimonio culturale come risorsa strategica per un'Europa sostenibile (2014 / C 183/08)* recita¹: *Il patrimonio culturale è costituito dalle risorse ereditate dal passato in tutte le sue forme e aspetti – tangibile, intangibile e digitale (nato digitale e digitalizzato), inclusi monumenti, siti, paesaggi, abilità, pratiche, conoscenze ed espressioni della creatività umana, nonché raccolte conservate e gestite da enti pubblici e privati come musei, biblioteche e archivi. Ha origine dall'interazione tra persone e luoghi nel tempo ed è in continua evoluzione. Queste risorse sono di grande valore per la società dal punto di vista culturale, ambientale, sociale ed economico e quindi la loro gestione sostenibile costituisce una scelta strategica per il 21° secolo.*

Partendo da questo presupposto, dobbiamo necessariamente cambiare il nostro approccio al digitale e alla digitalizzazione iniziando a considerarli rappresentazioni qualificanti l'età contemporanea e la *digital transformation* che la connota. Ciò implica l'urgenza di individuare e classificare tra le risorse digitali prodotte fino a oggi e in produzione, siano esse singoli oggetti, o complesse *digital libraries*, o sistemi 3D, quelle che possono essere identificate come il nuovo *Digital Cultural Heritage* (DCH), distinguendole da quelle prodotte per mera semplificazione di processi gestionali o per la fruizione estemporanea e immediata di scadenti rappresentazioni relative a entità tangibili e intangibili.

¹<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52014XG0614%2808%29>

Da diversi anni, la digitalizzazione e la qualità e preservazione delle risorse digitali sono riconosciute tra le principali emergenze da affrontare in tutto il mondo. Nel 2012 l'UNESCO ha tenuto la sua conferenza a Vancouver con il significativo titolo *The Memory of the World in the Digital Age: Digitization and Preservation* (Duranti and Shaffer [ed. by], 2012), nel cui ambito è stata redatta ed emanata la *Vancouver Declaration on Digitisation and Preservation*², con l'IFLA e l'International Council of Archives (ICA) tra i principali responsabili.

Da allora, la situazione non sembra essere molto cambiata, nonostante gli sforzi intrapresi per accelerare l'elaborazione di soluzioni a criticità di una complessità che, forse, non hanno precedenti storici.

Gli attuali approcci ai processi di creazione delle risorse digitali, infatti, sembrano non recepire l'evoluzione che negli ultimi anni ha riguardato la digitalizzazione, ancora oggi associata semplicemente alla riproduzione fotografica, mentre, invece, è diventata un processo complesso guidato da regole definite e condivise. Anche l'importanza della qualità dei dati digitali è del tutto sottovalutata nel relegarne la funzione a meri strumenti di mediazione per la fruizione del reale in forma virtuale, sebbene da più parti si riconosca che questi dovrebbero rispondere a requisiti di intellegibilità, affidabilità, pertinenza, persistenza, e registrare le trasformazioni delle funzioni legate al loro riutilizzo nel tempo. L'interpretazione strumentale, infatti, ancora orienta e condiziona negativamente soprattutto la strutturazione degli schemi di metadati con cui indicizzare gli oggetti digitali prodotti e la composizione delle descrizioni loro associate, formulate per essere meri codici funzionali esclusivamente alla ricerca e al recupero dei dati in rete.

Proprio i metadati e i contenuti descrittivi, invece, dovrebbero essere oggetto di particolare attenzione, in quanto sono la sola possibilità di registrare e rappresentare in modo intellegibile i processi di digitalizzazione, creazione e trasformazione che caratterizzano il ciclo di vita dei dati, e di conservare così le informazioni necessarie a conoscerli e a qualificarli come risorse digitali con funzioni anche culturali.

Il tema della funzione essenziale dei metadati nel management e nella fruizione dei dati digitali è il focus dei *FAIR Guiding Principles for Scientific Data Management and Stewardship*³, le linee guida per il management dei dati scientifici pubblicate nel 2016, da qualche tempo uno dei temi di maggior interesse nell'ambito del più ampio dibattito sulle possibilità di applicare le metodologie del *data science* alla creazione e gestione dei *data humanities*⁴.

A riguardo, nella CIDOC Conference 2018 si è tenuto un workshop sull'effettiva efficacia dei *FAIR Principles* rispetto agli scenari che oggi la digitalizzazione propone, e soprattutto rispetto a quelli che già si preannunciano imminenti⁵.

Il presente lavoro sintetizza alcune riflessioni maturate da quel proficuo confronto, relative alla necessità di provvedere a un ampliamento del requisito *R*: *Reusable* nei requisiti *R*⁵: *Reusable, Readable, Relevant, Reliable and Resilient*, finalizzato a facilitare l'applicabilità dei *FAIR Principles* ai *data humanities* e, conseguentemente, l'identificazione e certificazione come DCH dei dati rispondenti a tali requisiti nell'informe magma digitale in cui oggi fluttuiamo.

2 Verso un ampliamento dei *FAIR* da *R* a *R*⁵

L'assunto di partenza per avviare la riflessione è che i dati digitali non possono più essere creati finalizzandoli alla mera funzione di strumenti di mediazione per una fruizione della realtà alternativa a quella fisica: è necessario ripensarli quali risorse digitali che si qualificano come *record*, entità dinamiche e diacroniche che registrano e conservano nelle descrizioni i processi di digitalizzazione che li hanno creati e quelli che hanno caratterizzato il loro successivo ciclo di vita.

² http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/unesco_abc_vancouver_declaration_en.pdf

³ <https://www.go-fair.org/fair-principles/>

⁴ <https://www.rd-alliance.org/open-consultation-fair-data-humanities-until-15th-july-2019>; <https://www.gofair.org/implementation-networks/overview/co-operas/>; <https://operas.hypotheses.org/>

⁵ <http://www.cidoc2018.com/sites/default/files/CIDOC2018-BookOfAbstracts-Final-v-1-2.pdf>

I metadati descrittivi diventano perciò fondamentali e inscindibili dagli oggetti digitali, in quanto sono proprio l'accuratezza e la qualità delle descrizioni a qualificarli come *record* e, quindi, a renderli risorse digitali pensate e strutturate per essere fruite diacronicamente dagli utenti del futuro, che devono comprendere cosa il dato rappresenta alla pari degli utenti contemporanei.

L'adozione dei *FAIR Principles* lascia aperte alcune questioni. Innanzitutto, non siamo del tutto persuasi che Ricercabilità (*Findable*), Accessibilità (*Accessible*, che assolutamente non è identificabile con *Open*) e Interoperabilità (*Interoperable*) siano requisiti idonei a qualificare i dati come *record* e risorse digitali, conferendogli funzioni nuove e più evolute da quelle strumentali attualmente riconosciute.

Un dato che sia ricercabile, accessibile e interoperabile con altri non fornisce alcuna garanzia di qualità, sufficienza e affidabilità dei contenuti informativi che contiene. Inoltre, la ricercabilità e, di conseguenza, l'accessibilità e interoperabilità che sono a essa vincolate hanno senso nella misura in cui un dato sia oggetto di interesse da parte dei fruitori. E l'interesse per un dato è legato strettamente non alla sua mera funzione di chiave d'accesso a un oggetto digitale semplice o complesso, ma all'essere risorsa informativa e cognitiva per la quantità e qualità dei contenuti descrittivi che mette a disposizione dell'utente già in fase di lettura dei suoi metadati.

Siamo perciò del parere che il requisito che conferisce significato e senso ai primi tre, e dal quale questi dipendono indissolubilmente, sia la Riutilizzabilità (*Reusable*). L'utilizzo e, soprattutto il riutilizzo dei dati, infatti, sono secondo noi i fattori che ne garantiscono la sostenibilità nel tempo e, quindi, la sopravvivenza, in quanto requisiti caratterizzati da dinamismo e diacronia che, quasi sempre, implicano trasformazioni nelle funzioni delle entità che ne sono oggetto: per avere un'idea utilizzando un paradigma analogico, si pensi al Colosseo e al suo ciclo di vita.

Le registrazioni descrittive dei metadati sono perciò fondamentali per garantire qualità e persistenza delle risorse digitali, qualora siano improntate a equilibrate soluzioni quantitative/qualitative e rispondano a ulteriori requisiti che, secondo noi, sono altrettanto essenziali quanto la riutilizzabilità.

Anche la *Reusability*, infatti, di per sé non costituisce una garanzia di qualità del dato e del suo valore quale risorsa informativa e cognitiva. Anzi: proprio le variabili cui una risorsa è soggetta perché riutilizzabile possono essere fonte di distorsione e difformità dei contenuti, il cui valore informativo e cognitivo può perciò non essere più certificabile come affidabile.

La R di Reusable andrebbe perciò, secondo noi, ampliata in R5 con i seguenti requisiti:

- *Readability*: da intendersi non nell'accezione semantica di leggibilità, ma in quella concettuale di *intelligibilità* della risorsa digitale per tutte i possibili target di utenti interessati a fruirne; è requisito fondamentale per conferire ai metadati la funzione informativa e cognitiva necessaria a qualificarli come risorsa culturale, e si basa sull'equilibrato rapporto quantitativo/qualitativo dei contenuti descrittivi e sull'accuratezza formale, stilistica e linguistica dei contenuti;
- *Relevance*: la *pertinenza* è legata alla coerenza sia tra i diversi elementi che compongono il tracciato dei metadati, sia degli stessi elementi con i rispettivi contenuti descrittivi, ed è generata dalla relazione tra il dato e gli utenti interessati ai contenuti informativi e cognitivi in esso registrati; è, perciò, strettamente legata all'uso, al riutilizzo e alle possibili trasformazioni di funzione registrate nelle descrizioni; è requisito indispensabile affinché la risorsa, di solito creata con funzioni e scopi non necessariamente culturali, sia *persistente* e possa essere identificabile e riconoscibile nella sua struttura formale e descrittiva anche se varia nel tempo le proprie funzioni, evolvendosi in fonte di conoscenza sui processi che registra e, quindi, in record culturale digitale;
- *Reliability*: l'*affidabilità* è la certificazione e validazione della qualità della risorsa digitale rilevabili dalle registrazioni delle sue descrizioni durante tutto il suo ciclo di vita, in relazione a tutte le possibili trasformazioni ed evoluzioni funzionali cui può essere stata soggetta; è, dunque, strettamente connessa alla capacità dell'entità digitale di registrare e preservare gli elementi qualificanti la qualità informativa e cognitiva dei suoi contenuti descrittivi, anche nell'evoluzione delle funzioni e nelle variazioni di forme e funzioni nel tempo;
- *Resilience*: come l'intelligibilità, anche la *resilienza* applicata ai dati e, soprattutto, ai metadati è requisito fondamentale per conferire alle risorse digitali la nuova dimensione culturale;

chiosando la definizione comunemente in uso in ambito informatico⁶, essa va intesa come *la capacità di una risorsa digitale di adattarsi alle condizioni di utilizzo e riutilizzo, di resistere all'usura, di essere duttile nelle trasformazioni e nell'evoluzione delle sue funzioni, al fine di garantire la disponibilità del proprio potenziale cognitivo e informativo nello spazio e nel tempo*; è, quindi, indispensabile per garantire la sostenibilità e il riutilizzo delle risorse digitali nel medio-lungo termine, provvedendo a preservare sia le informazioni utili a conoscere i processi della loro creazione, sia quelle sulla loro funzione originale, sia, infine, le registrazioni delle trasformazioni ed evoluzioni funzionali che ne hanno caratterizzato il ciclo di vita.

3 Conclusioni

Tirando le conclusioni su quanto sopra sinteticamente delineato, è nostra opinione che l'adozione dei requisiti FAIR con la R ampliata in R^5 sia prerequisito indispensabile nel processo di creazione dei dati digitali e, soprattutto, dei metadati che li descrivono, in quanto gli conferirebbero le funzioni di potenziale DCH, rendendoli sostenibili, permanenti, affidabili e, nel contempo, storicizzandoli come fonti di conoscenza dei processi e delle complessità che caratterizzano la rapidissima evoluzione della *digital transformation*.

Non il dato in sé, infatti, ma l'interesse degli utenti presenti e futuri per la fruizione del dato in quanto risorsa informativa e cognitiva deve diventare il prerequisito su cui fondare l'intero processo di creazione, pubblicazione e preservazione di risorse digitali. L'applicazione dei requisiti R^5 , dunque, deve diventare oggetto di attenzione fin dalla fase di analisi e progettazione dei processi sia di digitalizzazione che di creazione di qualsiasi schema di metadati con cui descrivere e gestire gli oggetti digitali in produzione.

Solo così si potrà dare un serio inizio, nel medio termine, all'individuazione di quanto possa essere identificato come DCH nella massa di dati che oggi sovrabbonda nel web e, nel contempo, si potranno definire linee guida omogenee e condivise che presiedano alla creazione di nuove risorse avendo chiaro fin dal principio se gli si voglia conferire il potenziale valore di entità culturali.

In questo modo, nel giro di pochi anni le *Conclusioni EU* del 2014 potranno finalmente essere sostanziate con un nuovo DCH ufficialmente riconosciuto. Diversamente, continueremo a considerare digitalizzazione e digitale solo come un modo diverso e accattivante di fruire il tangibile, perdendo di vista quanto invece tutto ciò sia già oggi l'*humus* identitario che, pur a livelli diversi, identifica l'era digitale contemporanea.

Bibliografia

<https://www.go-fair.org/fair-principles/>

<https://www.go-fair.org/implementation-networks/overview/co-operas/>

<https://operas.hypotheses.org/>

http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/unesco_abc_vancouver_declaration_en.pdf

<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52014XG0614%2808%29>

<http://www.interpares.org/>

Agenzia per l'Italia Digitale (AgID), Presidenza del Consiglio dei Ministri, *Linee guida sulla conservazione dei documenti informatici*, Versione 1.0 – dicembre 2015, pp. 45 ss. http://www.agid.gov.it/sites/default/files/linee_guida/la_conservazione_dei_documenti_informatici_rev_def.pdf

⁶ <https://it.wikipedia.org/wiki/Resilienza>

- Lila Bailey. 2015. *Digital Orphans: The Massive Cultural Black Hole On Our Horizon*, Techdirt, Oct 13th 2015 <<https://www.techdirt.com/articles/20151009/17031332490/digitalorphans-massive-cultural-blackhole-our-horizon.shtml>>.
- Nicola Barbuti. 2016. *Le nuove entità culturali digitali tra Intangible Cultural Heritage e Patrimonio Culturale Immateriale*, in *The Creative Network – Conferenza GARR*, Firenze, 30 novembre-02 dicembre 2016 <<https://www.eventi.garr.it/it/conf16/home/materiali-conferenza-2016/paper/19-conf2016-paper-barbuti/file>>
- Nicola Barbuti. 2017. *Dalla Digital Culture al Digital Cultural Heritage: l'evoluzione impossibile?*, in *AIUCD 2017 Conference – Book of Abstract. Il telescopio inverso: big data e distant reading nelle discipline umanistiche*, p. 14-17, AIUCD <<http://aiucd2017.aiucd.it/wp-content/uploads/2017/01/book-of-abstract-AIUCD-2017.pdf>>
- Nicola Barbuti, and Ludovica Marinucci. 2018. *Dal Digital Cultural Heritage alla Digital Culture. Evoluzioni nelle Digital Humanities*, DH 2018 <<https://dh2018.adho.org/dal-digital-cultural-heritage-alla-digital-culture-evoluzioni-nelle-digital-humanities/>>
- Enrico Daga, and Leif Isaksen. 2016. *Proceedings of the 1st Workshop on Humanities in the Semantic Web*, co-located with 13th ESWC Conference 2016 (ESWC 2016), Anissaras, Greece, May 29th, 2016 <<http://ceur-ws.org/Vol-1608/paper-05.pdf>>.
- Luciana Duranti, and Elizabeth Shaeffer [ed. by]. 2012. *The Memory of the World in the Digital Age: Digitization and Preservation. An international conference on permanent access to digital documentary heritage*, UNESCO Conference Proceedings, 26-28 September 2012, Vancouver <http://ciscra.org/docs/UNESCO_MOW2012_Proceedings_FINAL_ENG_Compressed.pdf> .
- Vincenzo Gambetta. 2009. *La conservazione della memoria digitale*, [Rubano] : Siav.
- Pallab Ghosh. 2016. *Google's Vint Cerf warns of 'digital Dark Age'*, BBC News, Science & Environment, 13 February 2016 <<http://www.bbc.com/news/science-environment-31450389>>
- Maria Guercio. 2008. *Gli archivi come depositi di memorie digitali*, «Digitalia», Anno III, n. 2, Rom : ICCU, pp. 37-53.
- Maria Guercio. 2013. *Conservare il digitale. Principi, metodi e procedure per la conservazione a lungo termine di documenti digitali*, Roma-Bari : Laterza.
- Joint Steering Committee for Development of RDA. 2015. *Resource Description and Access (RDA)* <http://www.iccu.sbn.it/opencms/export/sites/iccu/documenti/2015/RDA_Traduzione_ICCU_5_Novembre_REV.pdf> .
- Wouter Kool, Brian Lavoie, and Titia van der Werf. 2014. *Preservation Health Check: Monitoring Threats to Digital Repository Content*, Dublin (Ohio) : OCLC Research <<http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-preservation-health-check-2014.pdf>> .
- Brian Lavoie, and Richard Gartner. 2013. *Preservation Metadata (2nd edition), DPC Technology Watch Report*, 03 May 2013, DPC Technology Watch Series <<http://www.dpconline.org/docman/technology-watch-reports/894-dpctw13-03/file>> .
- Library of Congress. *PREMIS – Preservation Metadata: Implementation Strategies*, v. 3.0 <<http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>> .
- Gilberto Marzano. 2011. *Conservare il digitale. Metodi, norme, tecnologie*, Milano : Editrice Bibliografica.
- Mellon Foundation and Digital Preservation Coalition Sponsor Formation of Task Force for Email Archives, 1 November 2016 <<https://mellon.org/resources/news/articles/mellon-foundation-and-digital-preservation-coalition-sponsor-formation-task-force-email-archives/>> .
- OCLC. *PREMIS (PREservation Metadata: Implementation Strategies) Working Group*, 2005 <<http://www.oclc.org/research/projects/pmwg/>>.
- Sustainable Economics for a Digital Planet: Ensuring Long-term Access to Digital Information*, Final Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access (F. Berman and B. Lavoie, co-chairs), La Jolla, February 2010 <http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf>.