Abilità informatiche

A.A. 2023/2024

02b - Gestione dei dati

Sebastian Barzaghi
sebastian.barzaghi2@unibo.it
https://orcid.org/0000-0002-0799-1527

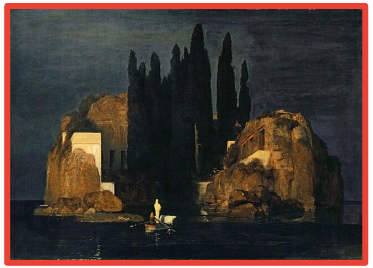
Riassunto della lezione precedente (02a)

Cos'è un dato?

Qualsiasi cosa che può essere quantificata, qualificata o interpretata in qualche modo e usata come evidenza

Ciò che è considerabile "dato" dipende da chi lo usa, come, e per quale scopo






University of York, "What is data?" URL: https://subjectguides.york.ac.uk/data/lore.

Cos'è un dataset?

Una collezione di dati (spesso in un formato leggibile da un computer)

Riflette le circostanze che hanno portato alla sua creazione e gestione (comunità, individui, organizzazioni, ambienti, strumenti, limiti, bias, responsabilità…)

	id	title	alt	type
1	1	Edipo risolve l'enigma della Sfinge	Oedipus and the Sphinx	Pittura
2	2	Eracle di Mantinea	statuette	Scultura
3	3	Eracle cattura il cinghiale di Erimanto	amphora	Pittura vascolare
4	4	Eracle brandisce la clava contro Caco	Hercule tuant Cacus	Pittura
5	5	Psiche riceve il primo bacio da Amore	PsychÃ© et l'Amour, dit aussi PsychÃ© recevant le premier baiser de l'Amour	Pittura
6	6	Diana cacciatrice	Statuette: Diane	Scultura
7	7	Eracle giunge all'Olimpo tra gli Dei	olpÃ©	Pittura vascolare
8	8	Odisseo massacra i pretendenti di Penelope	CratÃ´re des prÃ©tendants	Pittura vascolare
9	9	Dioniso, accompagnato dal suo corteo, incontra Arianna	Sarcophage; couvercle de sarcophage	Scultura
10	10	Venere di Milo	VÃ©nus de Milo	Scultura
11	11	Teseo uccide il Minotauro	Skyphos Rayet	Pittura vascolare
12	12	Gigantomachia	Amphore de Milo	Pittura vascolare
13	13	Orfeo incanta gli animali	OrphÃ©e charmant les animaux	Disegno
14	14	Clio, Euterpe e Talia	Clio, Euterpe et Thalie	Pittura
15	15	Danzatrici di Ruvo	Danzatrici di Ruvo	Pittura murale
16	16	Nike di Samotracia	Victoire de Samothrace	Scultura
17	17	Ercole Farnese	Ercole Farnese	Scultura
18	18	Atena MattÃ©i	AthÃ©na MattÃ©i	Scultura
19	19	Polifemo e Galatea si baciano		Pittura murale
20	20	Neottolemo trasporta il corpo di Astianatte		Scultura
21	21	Ajax and Cassandra		Pittura
22	22	Diana appoggiata a un cervo	Diane appuyÃ©e sur un cerf	Scultura
23	23	Eracle iniziato ai Misteri Eleusini		Scultura
24	24	Afrodite Callipige	Afrodite Callipige	Scultura

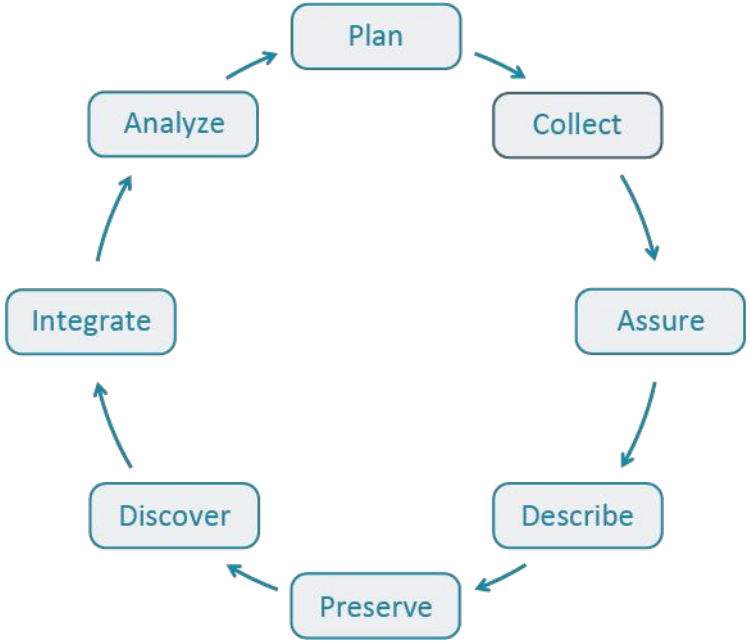
Parte del dataset *Mythologiae*. Fonte: https://mythologiae.unibo.it/

S. Ciston, "A CRITICAL FIELD GUIDE FOR WORKING WITH MACHINE LEARNING DATASETS," K. Crawford and M. Ananny, Eds., Knowing Machines project, Feb. 2023.

Gestione dei dati

Processo **critico** di creazione, raccolta, organizzazione, descrizione, archiviazione, e condivisione dei dati

Obiettivo: produrre dataset autodescritti, sostenibili e utilizzabili



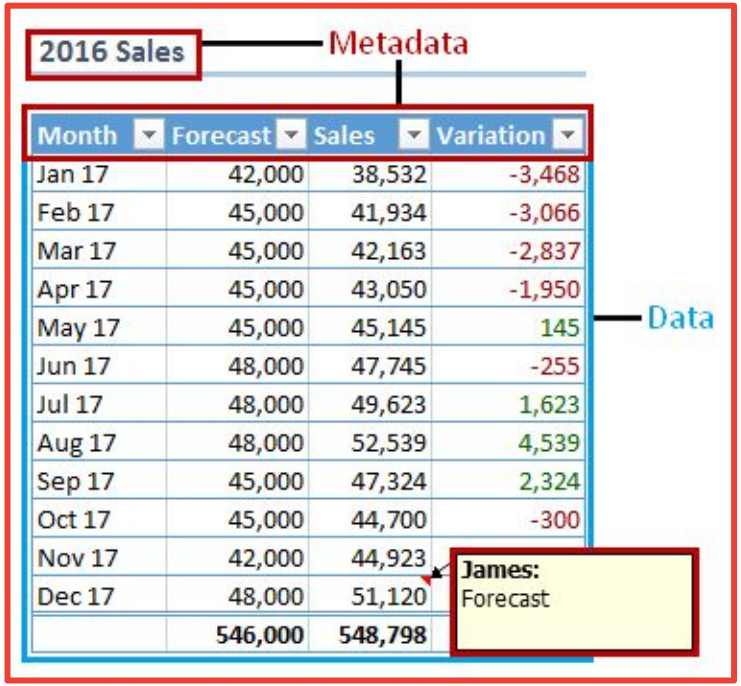
Una possibile rappresentazione del processo di gestione dei dati. Fonte: https://dataoneorg.github.io/Education/bestpractices/

Strasser, C., Cook, R., Michener, W., & Budden, A. (2012). Primer on Data Management: What you always wanted to know.

Cos'è un metadato?

Dato per descrivere o rappresentare una caratteristica di un altro dato

- Tipo di dato
  - Descrittivi
  - Strutturali
  - Amministrativi
- Multiforme
- Associativo

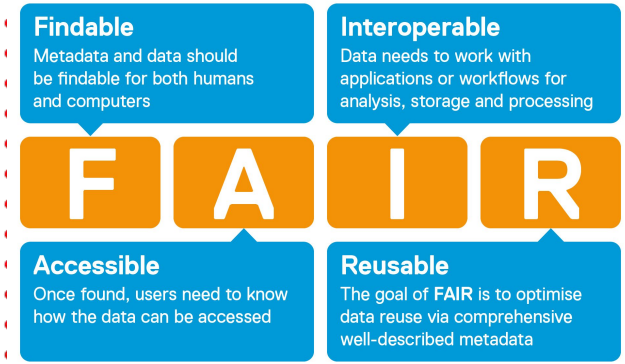


Fonte: https://dataedo.com/kb/data-glossary/what-is-metadata

Australian Research Data Commons (ARDC). (2020). ARDC Metadata Guide. Zenodo. https://doi.org/10.5281/zenodo.6459832

Perché usare i metadati?

- Trovabilità
- Accessibilità
- Interoperabilità
- Riusabilità



Fonte: https://scibite.com/solutions/enterprise-fair-data-mdm/

Remember the Mars Climate Orbiter incident from 1999?

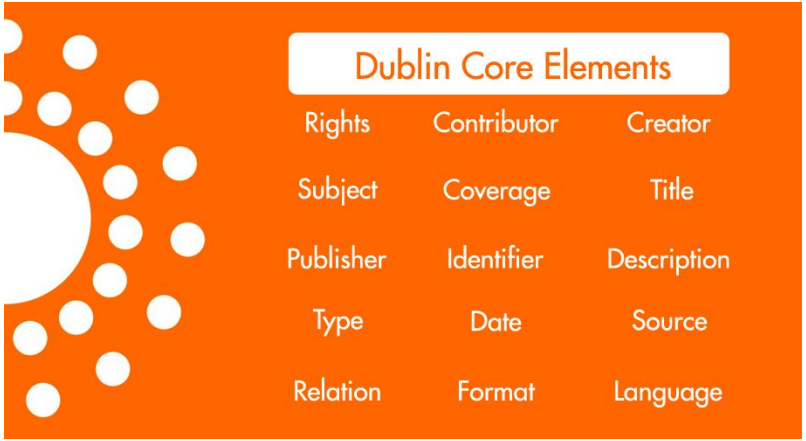
Fonte: https://www.simscale.com/blog/nasa-mars-climate-orbiter-metric/

Australian Research Data Commons (ARDC). (2020). ARDC Metadata Guide. Zenodo. https://doi.org/10.5281/zenodo.6459832

Schemi di metadati

Una struttura concettuale che specifica quali asserzioni (metadati) utilizzare e secondo quali regole

- Insieme di *elementi*
- Definizione di *elementi*
- Relazioni tra *elementi*
- Regole



Fonte: https://historygonedigital.wordpress.com/2017/10/02/dublin-core-metadata-element-set/

Australian Research Data Commons (ARDC). (2020). ARDC Metadata Guide. Zenodo. https://doi.org/10.5281/zenodo.6459832

Schemi di codifica

Insieme di regole che specificano sintassi o lessico utilizzati nelle asserzioni (metadati) di uno schema di metadati

SdC sintattica → ISO 8601 Date="2003-07" ISO 639 Language="en"

Vocabolari controllati → DCMI Type Vocabulary Type="Text" http://purl.org/dc/dcmitype/Text

Authority file → VIAF **Verdi, Giuseppe, 1813-1901.** VIAF ID: 22329110 (Personal) Permalink: http://viaf.org/viaf/22329110

Australian Research Data Commons (ARDC). (2020). ARDC Metadata Guide. Zenodo. https://doi.org/10.5281/zenodo.6459832;

2.3 Open Science

Definizione
Open Science e
scienze umanistiche

Cosa intendiamo con “scienza”?

Da *scientia* (conoscenza) → Conoscenza acquisita attraverso lo studio o la pratica; padronanza di una disciplina o area particolare

Il **processo organizzato** attraverso il quale l'umanità cerca di scoprire e comprendere i **fenomeni naturali, sociali e culturali**, convalidando le scoperte attraverso la **condivisione dei dati** e la revisione tra pari, al fine di **utilizzare** tale conoscenza a proprio vantaggio

Möller, L. (2023). UNESCO Recommendation on Open Science. In *66. Helmholtz Open Science online seminar*. https://unesdoc.unesco.org/ark:/48223/pf0000379949

Cos’è la Scienza Aperta?

Un cambiamento di paradigma nella scienza

Termine ombrello per diverse pratiche volte a rendere la ricerca scientifica più diffusa, accessibile e trasparente

Melanie Imming, & Jon Tennant. (2018). Sticker open science: just science done right (ENG). Zenodo. https://doi.org/10.5281/zenodo.128557

Cos’è la Scienza Aperta?

Combinazione di vari movimenti e pratiche volti a rendere la conoscenza scientifica disponibile, accessibile e riutilizzabile per tutti

Aumentare le collaborazioni scientifiche e la condivisione delle informazioni a vantaggio della scienza e della società

Comprende tutte le discipline scientifiche e gli aspetti delle pratiche accademiche, comprese le scienze di base e applicate, le scienze naturali e sociali e le discipline umanistiche

Ulrike Wuttke. (2018, novembre 19). Introduction to Humanities Research Data Management. Zenodo. https://doi.org/10.5281/zenodo.149125

Ok, ma di *cosa* si tratta?

Creazione e condivisione di Open Data liberamente disponibili per l'accesso e il (ri)utilizzo

Ideale: Dati senza restrizioni da copyright, brevetti o altri meccanismi di controllo → risultati trasparenti

Realità: *As open as possible, as closed as necessary*

Ulrike Wuttke. (2018, novembre 19). Introduction to Humanities Research Data Management. Zenodo. https://doi.org/10.5281/zenodo.149125

Perché la Scienza Aperta?

Argomento scientifico: maggiore riproducibilità dei risultati della ricerca; maggiore trasparenza dei metodi di ricerca e valutazione

Argomento sociologico: più impatto sociale della ricerca e coinvolgimento dei cittadini

Argomento utilitaristico: più citazioni; più attenzione mediatica; più possibilità di collaborazioni e opportunità di lavoro e finanziamento; più risparmio di tempo e denaro; requisiti dei finanziatori a livello nazionale e internazionale

Ulrike Wuttke. (2018, novembre 19). Introduction to Humanities Research Data Management. Zenodo. https://doi.org/10.5281/zenodo.149125

Scienza Aperta e discipline umanistiche

Computer, Web, Big Data, AI → nuovi metodi di ricerca e oggetti di studio

Lentezza generalizzata (e a volte ostilità):

- valori e metodi percepiti come contrastanti
- mancanza di finanziamenti
- problemi con copyright e licenze
- natura ibrida dei dati umanistici
- mancanza di (consapevolezza su) documentazione e standardizzazione
- mancanza di (cultura su) interoperabilità, riutilizzo e condivisione

Edmond, J. , Garnett, V. , (2018) "*Manage, Improve and Open Up Your Research Data*", PARTHENOS. https://training.parthenos-project.eu/sample-page/manage-improve-and-open-up-your-research-and-data/ (Retrieved on 21/04/2024).

Scienza Aperta e discipline umanistiche

Forti incentivi:

- nuove prospettive di ricerca
- eliminazione della duplicazione
- integrazione delle conoscenze
- pubblicazione e diffusione della ricerca
- aumento dell'impatto e della visibilità
- allineamento alle politiche dell'UE (e quindi finanziamenti)

Edmond, J. , Garnett, V. , (2018) "*Manage, Improve and Open Up Your Research Data*", PARTHENOS. https://training.parthenos-project.eu/sample-page/manage-improve-and-open-up-your-research-and-data/ (Retrieved on 21/04/2024).

Scienza Aperta per umanisti: perché?

Paura di danneggiare la propria carriera?

Piccoli cambiamenti nel flusso di lavoro, come l'adozione di ORCID e l'uso delle opzioni di deposito open access, possono fare una grande differenza

→ un migliore accesso ai materiali rari

→ una maggiore rappresentazione di posizioni marginali

→ una maggiore credibilità della ricerca umanistica

Edmond, J., & Tóth-Czifra, E. (2018). Open data for humanists, a pragmatic guide. https://shs.hal.science/halshs-02115443

Dati umanistici

I ricercatori delle discipline umanistiche e del patrimonio culturale hanno dati?
→ Sì, molti, ma di solito non usano il termine “dati”

Dati prodotti e utilizzati nei processi scientifici come digitalizzazione, studio di fonti, esperimenti, misurazioni, interviste e indagini

Esempi di dati umanistici: fonti primarie (testi, immagini, video), fonti secondarie, fonti digitalizzate, fonti digitali, strumenti digitali (software), annotazioni, ecc.

Gualandi, B., Pareschi, L., & Peroni, S. (2023). What do we mean by “data”? A proposed classification of data types in the arts and humanities. *Journal of Documentation*, *79*(7), 51-71.

In che modo i dati umanistici sono diversi?

Contesti di ricerca specifici e interdisciplinari

Basati sull'arricchimento continuo (strati di interpretazione)

Problemi legati alla proprietà dei dati

Gran parte di essi non è digitale

I sistemi semiotici dei dati umanistici possono essere molto personali e individuali

Trinity College, Dublin. (2017, July 10). *Manage, Improve and Open up your Research and Data – Parthenos training*. https://training.parthenos-project.eu/sample-page/manage-improve-and-open-up-your-research-and-data/

Principi FAIR, di nuovo

Migliorare e aumentare la condivisione dei dati di ricerca è vantaggioso

Quali regole di base dovrebbero essere date su come le persone condividono, quando e dove?

→ Sviluppo dei principi FAIR (*Findable, Accessible, Interoperable, Reusable*)

Sviluppati da FORCE 11 (un'organizzazione pan-disciplinare), questi principi forniscono una comprensione di base del valore che la condivisione dei dati può offrire e i requisiti di base per farlo

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Principi FAIR, di nuovo: *Findability*

F1. *ai (meta)dati è assegnato un identificatore univoco globale e persistente* ~ come se ogni oggetto avesse un numero di telefono unico che puoi chiamare quando ne hai bisogno

F2. *i dati sono descritti con metadati ricchi* ~ come se ogni oggetto avesse una scheda con tutte le informazioni su di esso

F3. *i (meta)dati sono registrati o indicizzati in una risorsa ricercabile* ~ come i libri elencati in un catalogo

F4. *i metadati specificano l'identificatore dei dati* ~ come se ogni oggetto avesse il suo nome per distinguersi dagli altri

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Principi FAIR, di nuovo: *Accessibility*

A1. *i (meta)dati sono recuperabili dal loro identificatore utilizzando un protocollo di comunicazione standardizzato* ~ come se ci fosse un modo predefinito e comune per chiedere e ricevere i dati, come quando si usa un linguaggio o un metodo di comunicazione che tutti possono capire e usare

A1.1. *il protocollo è aperto, gratuito e universalmente implementabile* ~ come avere un libro che puoi leggere gratuitamente ovunque e da qualsiasi dispositivo

A1.2. *il protocollo consente una procedura di autenticazione e autorizzazione, se necessario* ~ mostrare una tessera d'identità per entrare in un posto

A2. *i metadati sono accessibili, anche quando i dati non sono più disponibili* ~ avere le istruzioni su come fare qualcosa anche se l'oggetto di cui hai bisogno non è più disponibile; le istruzioni ti permettono ancora di sapere cosa fare

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Principi FAIR, di nuovo: *Interoperability*

I1. *i (meta)dati utilizzano un linguaggio formale, accessibile, condiviso e ampiamente applicabile per la rappresentazione della conoscenza* ~ come se tutti parlassero la stessa lingua per comunicare tra loro

I2. *i (meta)dati utilizzano vocabolari che seguono i principi FAIR* ~ come se tutti usassero le stesse parole per descrivere le stesse cose

I3. *i (meta)dati includono riferimenti qualificati ad altri (meta)dati* ~ come se ci fossero collegamenti tra diverse pagine di un libro per aiutarti a trovare tutte le informazioni correlate

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Principi FAIR, di nuovo: *Reusability*

R1. *i meta(dati) hanno una pluralità di attributi accurati e rilevanti* ~ come se ci fossero molte informazioni su un oggetto in modo da poterlo comprendere appieno

R1.1. *i (meta)dati sono rilasciati con una licenza d'uso dei dati chiara e accessibile* ~ come se ci fosse una nota che spiega esattamente cosa puoi fare con un oggetto

R1.2. *i (meta)dati sono associati alla loro provenienza* ~ come se ci fosse una storia che spiega da dove viene un oggetto e chi l'ha creato

R1.3. *i (meta)dati soddisfano gli standard della comunità rilevanti per il dominio* ~ come se ci fossero regole condivise su come fare le cose in modo che tutti possano fidarsi dei risultati

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Gestione dei dati, di nuovo

La gestione dei dati ha l'obiettivo di rendere il processo di ricerca più efficiente

Titolo: "L'altro Rinascimento: Ulisse Aldrovandi e le meraviglie del mondo. Un'analisi della collezione"

Obiettivo: Analizzare e arricchire una collezione di oggetti culturali appartenuta ad Ulisse Aldrovandi (1522 - 1605) per comprendere meglio lo sviluppo della cultura scientifica in quel periodo

Pianificare

Capire i *chi, cosa, quando, dove, come, quanto* del progetto

Data Management Plan (DMP) → Documento che contiene informazioni su come gestire, organizzare, documentare tutti i dati che in qualche maniera fanno parte del progetto durante il loro ciclo di vita

Può intimidire e sembrare eccessivo, ma serve a pensare sistematicamente attraverso il processo di ricerca da una "prospettiva dei dati"

Esempio: Gualandi, B., & Peroni, S. (2024). Data Management Plan: second version (1.0). Zenodo. https://doi.org/10.5281/zenodo.10727879

Strumenti utili: *Argos, DMPTool, DMPOnline*

https://argos.openaire.eu/home

https://dmptool.org/

https://dmponline.dcc.ac.uk/

Raccogliere

- Quali dati raccoglierai o creerai?
- Che tipo, formato e volume di dati?
- Ci sono dati esistenti?
- Come saranno raccolti o creati i dati?
- Quali standard e metodologie utilizzerai?
- Come strutturerai e nominerai le tue cartelle e file?

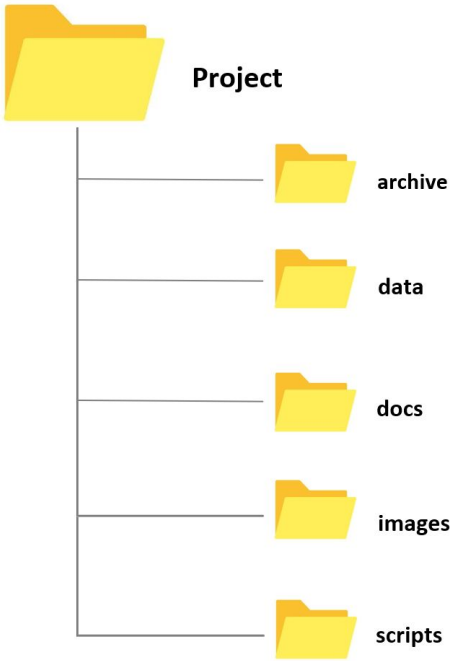
Lo studente...

Visita la mostra per fotografare e raccogliere immagini degli oggetti presenti. Utilizza Zotero per annotare informazioni bibliografiche su ogni opera. Organizza le immagini in cartelle separate per sala.

Australian Research Data Commons (ARDC). (2020). ARDC Metadata Guide. Zenodo. https://doi.org/10.5281/zenodo.6459832

Indicazioni utili: *Struttura*

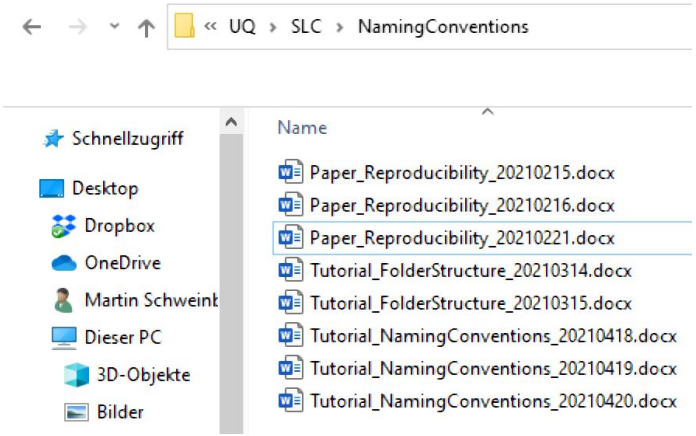
- Raggruppa i file con proprietà comuni in cartelle
- Usa nomi di cartelle significativi e chiari
- Mantieni un numero gestibile di sottocartelle
- Struttura le cartelle in modo gerarchico
- Separa il lavoro corrente da quello completato
- Conserva i dati grezzi separatamente da quelli elaborati



Schweinberger, Martin. 2022. *Data Management, Version Control, and Reproducibility*. Brisbane: The University of Queensland. url: https://slcladal.github.io/repro.html (Version 2022.10.10).

Indicazioni utili: *Nomenclatura*

- Mantieni nomi brevi ma significativi (se usi abbreviazioni, tieni traccia in un file README)
- Includi date nel formato YYYY-MM-DD
- Evita spazi, punti e caratteri speciali come / \ : * ? " < > |
- Includi informazioni rilevanti nei nomi dei file, come l'ID e il numero di versione



Schweinberger, Martin. 2022. *Data Management, Version Control, and Reproducibility*. Brisbane: The University of Queensland. url: https://slcladal.github.io/repro.html (Version 2022.10.10).

Indicazioni utili: *Zotero, Zoterobib, Europeana, re3data*

zotero 

https://www.zotero.org/ +
https://chromewebstore.google.com/detail/zotero-connector/ekhagklcjbdpajgpjgmbionohlpdbjgc?hl=it

zoterobib

https://zbib.org/

europeana

https://www.europeana.eu/it

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

https://www.re3data.org/

Descrivere

- Quali documentazione e metadati accompagneranno i dati?
- Come catturerai/creerai questa documentazione e questi metadati?
- Quali standard utilizzerai e perché?
- Evita l'ambiguità
- Utilizza schemi di metadati, schemi di codifica sintattica, vocabolari controllati, authority file

Lo studente...

Crea un README che include informazioni dettagliate sul progetto e una tabella CSV contenente i metadati (nomi delle colonne della tabella) e i dati (i valori nelle celle della tabella) per ogni oggetto della collezione (righe della tabella)

Indicazioni utili: *README*

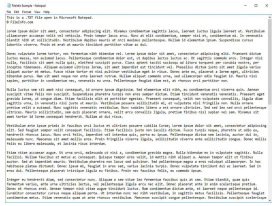
Un documento di testo semplice (in formato .txt o .md), progettato per essere letto prima del resto della documentazione

Dovrebbe includere i metodi di raccolta dati, la struttura dei file, l'elaborazione dei dati, dettagli sull'attrezzatura utilizzata, informazioni sui partecipanti, accordi legali ed etici, formati dei file, glossario delle colonne, gestione dei dati mancanti, metodologia e software utilizzato

Breve guida: https://www.makeareadme.com/

Maisam M. Dadkan, & Garrett Speed. (2023, febbraio 21). How to Write a README. Zenodo. https://doi.org/10.5281/zenodo.7660046

Strumento utile: *txt, CSV, Markdown*



https://www.ionos.it/digitalguide/server/configurazione/file-txt/

https://www.aifa.gov.it/documents/20142/1102225/200520_Guida%20CSV.pdf

https://experienceleague.adobe.com/it/docs/contributor/contributor-guide/writing-essentials/markdown

Sarah Simpkin, "Getting Started with Markdown," *Programming Historian* 4 (2015), https://doi.org/10.46430/phen0046.

Archiviazione e backup

- Come verranno archiviati e eseguiti i backup dei dati durante la ricerca?
- Come verranno recuperati i dati in caso di incidente?
- Quali sono i rischi per la sicurezza dei dati e come verranno gestiti?
- Come garantirai che i collaboratori possano accedere ai tuoi dati in modo sicuro?

Lo studente...

Conserva le immagini degli oggetti e la rispettiva documentazione sul suo computer personale, su un hard disk esterno e su Google Drive

Indicazioni utili: *Regola del 3-2-1*

Conservare almeno tre (3) copie dei tuoi dati e archiviare copie di backup su due (2) diversi supporti di archiviazione, di cui uno (1) situato in un luogo esterno

- 1 sul portatile
- 1 su chiavetta
- 1 su repository online (es. OneDrive, GitHub)

Schweinberger, Martin. 2022. *Data Management, Version Control, and Reproducibility*. Brisbane: The University of Queensland. url: https://slcladal.github.io/repro.html (Version 2022.10.10).

Pubblicazione

- Usa Identificatori Persistenti (ad esempio DOI e ORCID): riferimento duraturo e univoco a un oggetto digitale (articolo di giornale, dataset, campione scientifico, opera d'arte, tesi di dottorato, pubblicazione o persona)
- Come gestirai le questioni di copyright e diritti di proprietà intellettuale (DPI)?

Lo studente...

Una volta completata la raccolta e l'analisi dei dati, pubblica il materiale raccolto su Zenodo. Ottiene un DOI per il materiale per garantirne la citabilità. Assegna ai metadati una licenza CC0 per assicurarne il totale riuso, e la licenza CC BY-NC-SA 4.0 per le immagini.

Strumenti utili: *ORCID, Zenodo, Figshare, arXiv*

https://orcid.org/

https://zenodo.org/

https://figshare.com/

https://arxiv.org/

Australian Research Data Commons (ARDC). (2020). ARDC Metadata Guide. Zenodo. https://doi.org/10.5281/zenodo.6459832

Strumenti utili: *CCChooser*

https://chooser-beta.creativecommons.org/

Engelfriet, A. (2009). Choosing an open source license. *IEEE software*, *27*(1), 48-49. https://doi.org/10.1109/MS.2010.5

Abilità informatiche

A.A. 2023/2024

02b - Fine

Sebastian Barzaghi
sebastian.barzaghi2@unibo.it
https://orcid.org/0000-0002-0799-1527