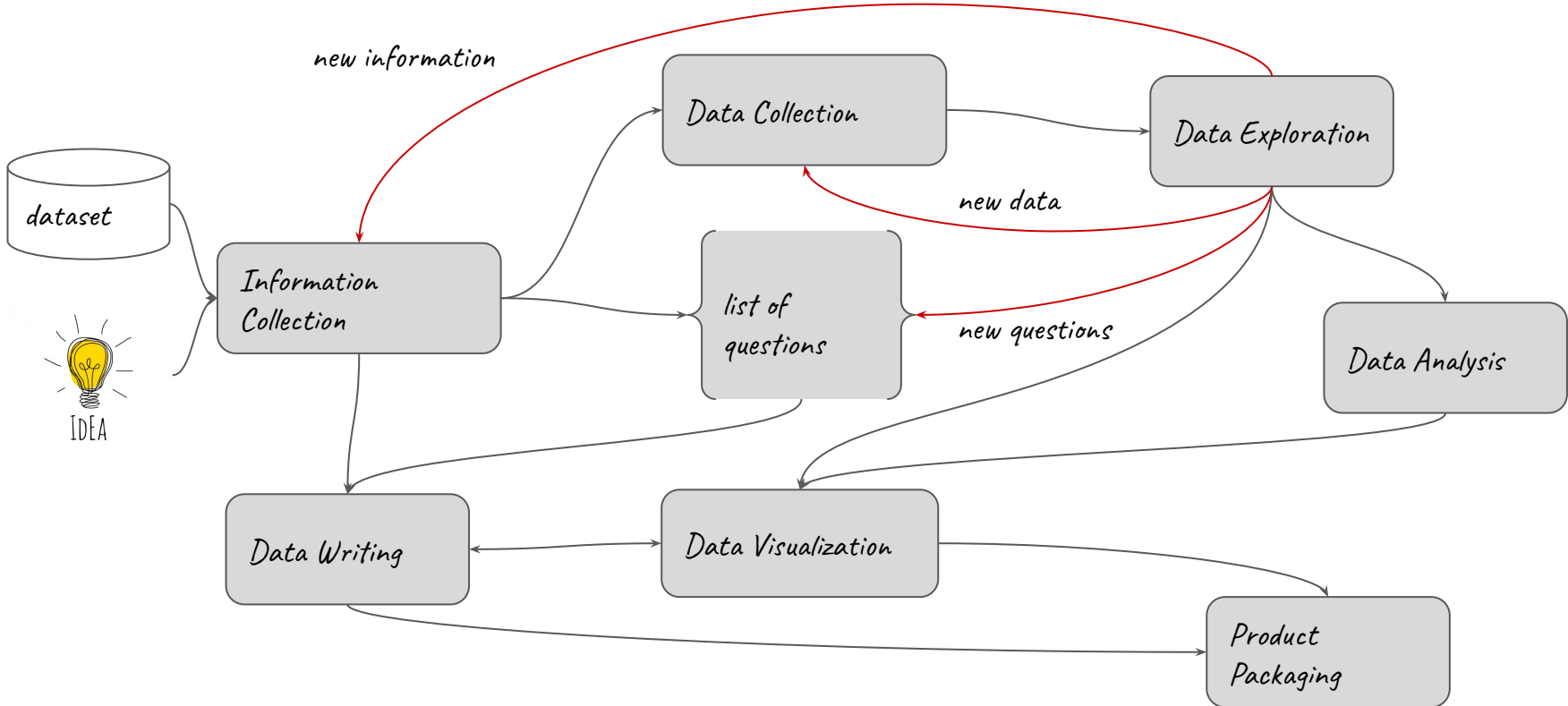


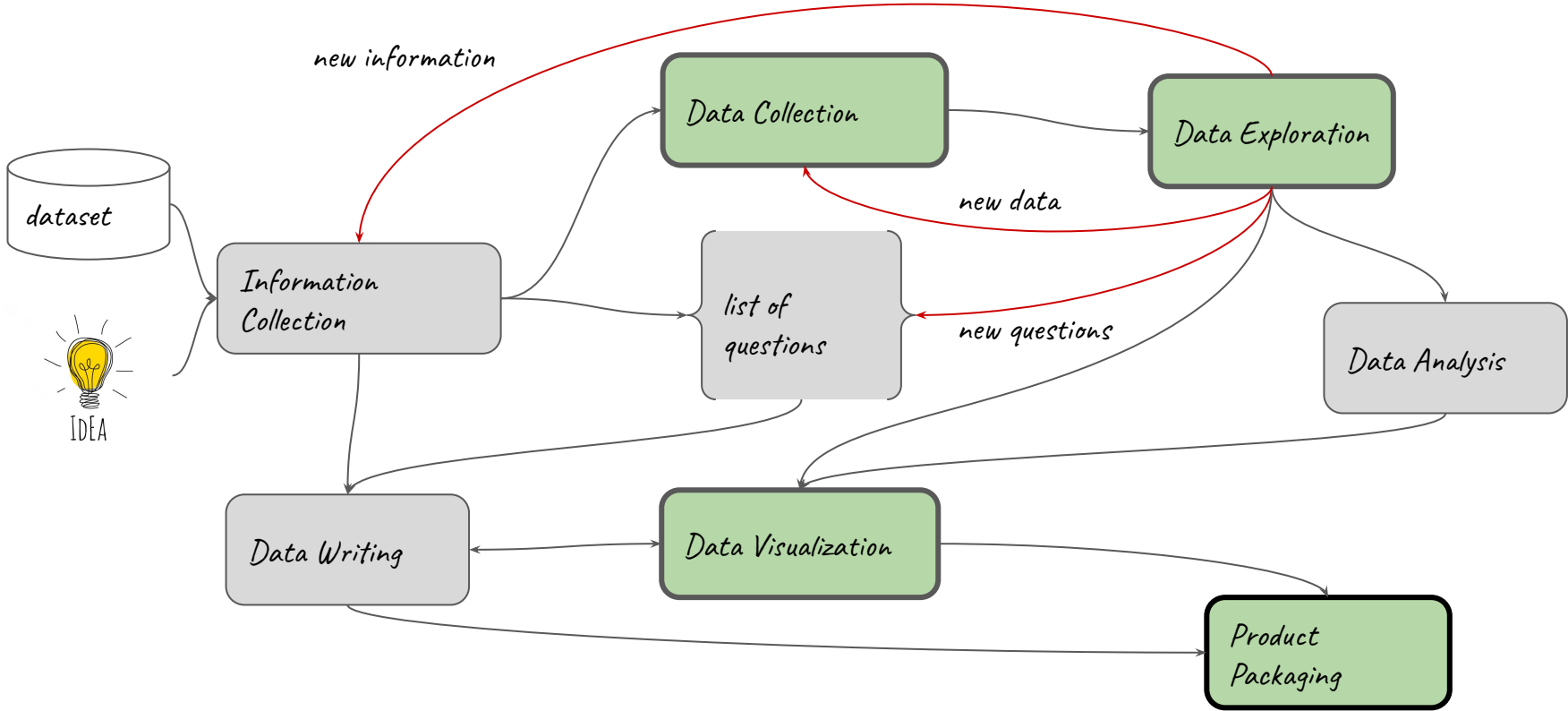
Gli Strumenti del Data Journalist

Angelica Lo Duca
angelica.loduca@iit.cnr.it

Data Journalism Workflow



Data Journalism Workflow - Software



Strumenti

- Data Collection
 - Python + Jupyter Notebook + librerie specifiche
- Data Cleaning
 - Open Refine
- Data Exploration/Data Visualization
 - Tableau
- Product Packaging
 - HTML/CSS/JSS, ...

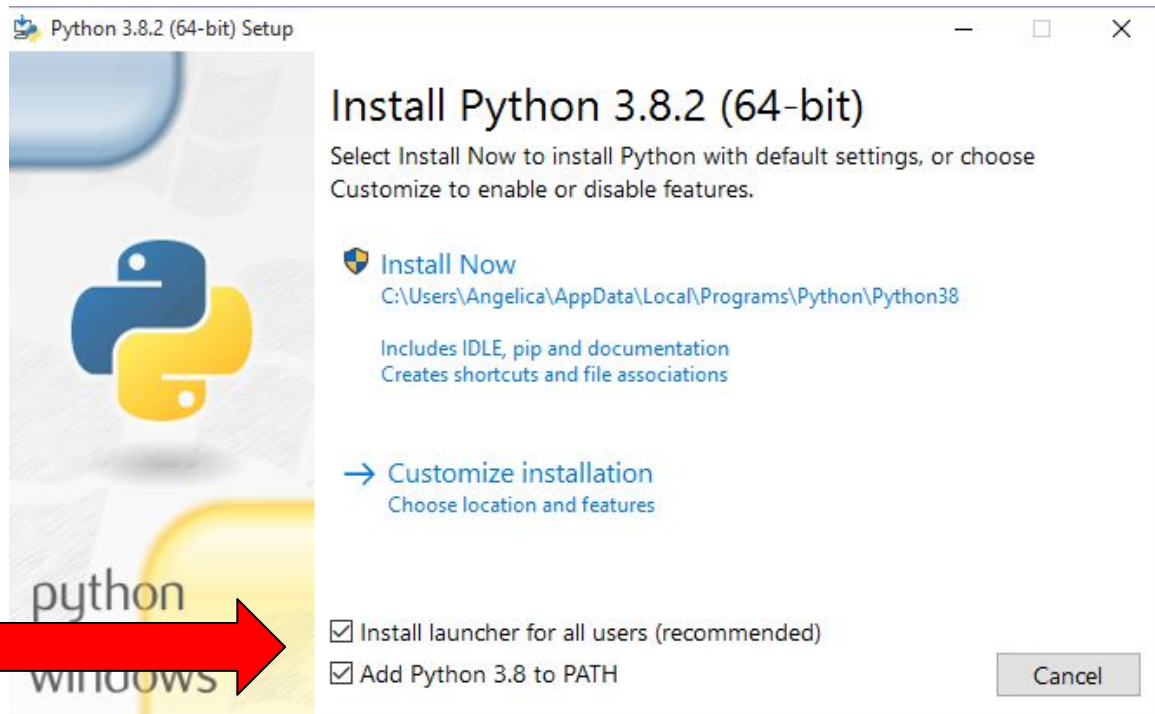
Data Collection

Python 3 - Windows

- <https://www.python.org/downloads/> (pagina generale)

Version	Operating System	Description	MD5 Sum	File Size	GPG
Gzipped source tarball	Source release		f9f3768f757e34b342dbc06b41cbc844	24007411	SIG
XZ compressed source tarball	Source release		e9d6ebc92183a177b8e8a58cad5b8d67	17869888	SIG
macOS 64-bit installer	Mac OS X	for OS X 10.9 and later	f12203128b5c639dc08e5a43a2812cc7	30023420	SIG
Windows help file	Windows		7506675dccb9a1569b54e600ae66c9fb	8507261	SIG
Windows x86-64 embeddable zip file	Windows	for AMD64/EM64T/x64	1a98565285491c0ea65450e78afe6f8d	8017771	SIG
Windows x86-64 executable installer	Windows	for AMD64/EM64T/x64	b5df1cbb2bc152cd70c3da9151cb510b	27586384	SIG
Windows x86-64 web-based installer	Windows	for AMD64/EM64T/x64	2586cdad1a363d1a8abb5fc102b2d418	1363760	SIG
Windows x86 embeddable zip file	Windows		1b1f0f0c5ee8601f160cfad5b560e3a7	7147713	SIG
Windows x86 executable installer	Windows		6f0ba59c7dbeba7bb0ee21682fe39748	26481424	SIG
Windows x86 web-based installer	Windows		04d97979534f4bd33752c183fc4ce680	1325416	SIG

Installazione Python 3 su Windows



Selezionare entrambe le caselle con il segno di spunta.

Verifica della correttezza dell'installazione

- Chiudere (se aperto) e Aprire un terminale
 - cercare nella barra di ricerca la parola “terminale”
 - selezionare prompt dei comandi
- Digitare **python --version**
 - come output dovrebbe dare: Python 3.12.2
- In alternativa digitare py (per utenti Windows)

Python 3 - Mac OS

- <https://www.python.org/downloads/> (pagina generale)

Version	Operating System	Description	MD5 Sum	File Size	GPG
Gzipped source tarball	Source release		f9f3768f757e34b342dbc06b41cbc844	24007411	SIG
XZ compressed source tarball	Source release		e9d6ebc92183a177b8e8a58cad5b8d67	17869888	SIG
macOS 64-bit installer	macOS	Python 3.9 and later	f12203128b5c639dc08e5a43a2812cc7	30023420	SIG
Windows help file	Windows		7506675dccb9a1569b54e600ae66c9fb	8507261	SIG
Windows x86-64 embeddable zip file	Windows	for AMD64/EM64T/x64	1a98565285491c0ea65450e78afe6f8d	8017771	SIG
Windows x86-64 executable installer	Windows	for AMD64/EM64T/x64	b5df1cbb2bc152cd70c3da9151cb510b	27586384	SIG
Windows x86-64 web-based installer	Windows	for AMD64/EM64T/x64	2586cdad1a363d1a8abb5fc102b2d418	1363760	SIG
Windows x86 embeddable zip file	Windows		1b1f0f0c5ee8601f160cfad5b560e3a7	7147713	SIG
Windows x86 executable installer	Windows		6f0ba59c7dbeba7bb0ee21682fe39748	26481424	SIG
Windows x86 web-based installer	Windows		04d97979534f4bd33752c183fc4ce680	1325416	SIG

Verifica della correttezza dell'informazione

- Aprire un terminale
- Mac OS di default ha installato python 2.7
 - digitando **python --version** dovrebbe dare come output Python 2.7.10
- Per usare python 3 si deve usare il comando python3
 - digitando **python3 --version** dovrebbe dare come output Python 3.12.2

Python 3 - Ubuntu

- In un terminale scrivere i seguenti comandi:
 - `sudo apt update`
 - `sudo apt install software-properties-common`
 - `sudo add-apt-repository ppa:deadsnakes/ppa`
 - `sudo apt update`
 - `sudo apt install python3.12 python3.12-venv python3.12-dev`
 - `python3.12 -V`
- **Installare pip3**
 - `sudo apt install python3-pip`

Jupyter Notebooks

- <https://jupyter.org/install> (Windows)
 - pip install jupyterlab <invio>
 - pip install notebook <invio>
- <https://jupyter.org/install> (Mac/Linux)
 - pip3 install jupyterlab <invio>
 - pip3 install notebook <invio>

Esecuzione di Jupyter

Da terminale digitare:

- Windows/Linux
 - `jupyter notebook`
- Mac
 - `jupyter-notebook`
 - `python3 -m notebook`

Data Collection

	Librerie
Web API Flickr	<code>pip install flickrapi</code> <code>pip3 install flickrapi (Linux/Mac)</code>
Estrazione da PDF	<code>pip install tabula-py</code> <code>pip3 install tabula-py (Linux/Mac)</code> tabula - https://tabula.technology/
Web Scraping	selenium + chrome driver

Selenium + Chrome Driver per python

- Da terminale digitale
 - `pip install selenium`
 - `pip3 install selenium` (Mac/Linux)
 - <https://selenium-python.readthedocs.io/installation.html>
- Web Driver per python
 - verificare la versione del browser

Chrome:	https://sites.google.com/a/chromium.org/chromedriver/downloads
Edge:	https://developer.microsoft.com/en-us/microsoft-edge/tools/webdriver/
Firefox:	https://github.com/mozilla/geckodriver/releases
Safari:	https://webkit.org/blog/6900/webdriver-support-in-safari-10/

Versione di Chrome

- Per conoscere la versione di Chrome, nella barra degli indirizzi digitare
 - `chrome://settings/help`



Google Chrome

Aggiornamento quasi terminato. Riavvia Google Chrome per completarlo. Le finestre di navigazione in incognito non verranno riaperte.



Versione 80.0.3987.116 (Build ufficiale) (a 64 bit)

Riavvia

Installazione Web Driver

- Dopo aver scaricato il Web driver, estrarre il file e posizionarlo in una cartella a piacere del proprio file system (ad esempio una cartella sul desktop)
- Impostare la variabile di ambiente \$PATH in modo da contenere il percorso alla cartella prescelta
 - Windows - <https://youtu.be/mHtlBq5cP2Y>
 - Mac OS - modificare il file .bash_profile o .profile
 - aprire un terminale e dalla home directory scrivere ls -a per identificare il proprio file di profilo
 - nella propria home digitare il comando **vi .profile**
 - in fondo al file scrivere
 - digitare il tasto i (insert)
 - export PATH="`<percorso al web driver>`:\$PATH"
 - premere ESC :wq per salvare il file
 - chiudere il terminale e riavviare il computer
 - aprire il terminale e digitare echo \$PATH (dovrebbe mostrare anche il percorso alla

Data Cleaning

Data Cleaning

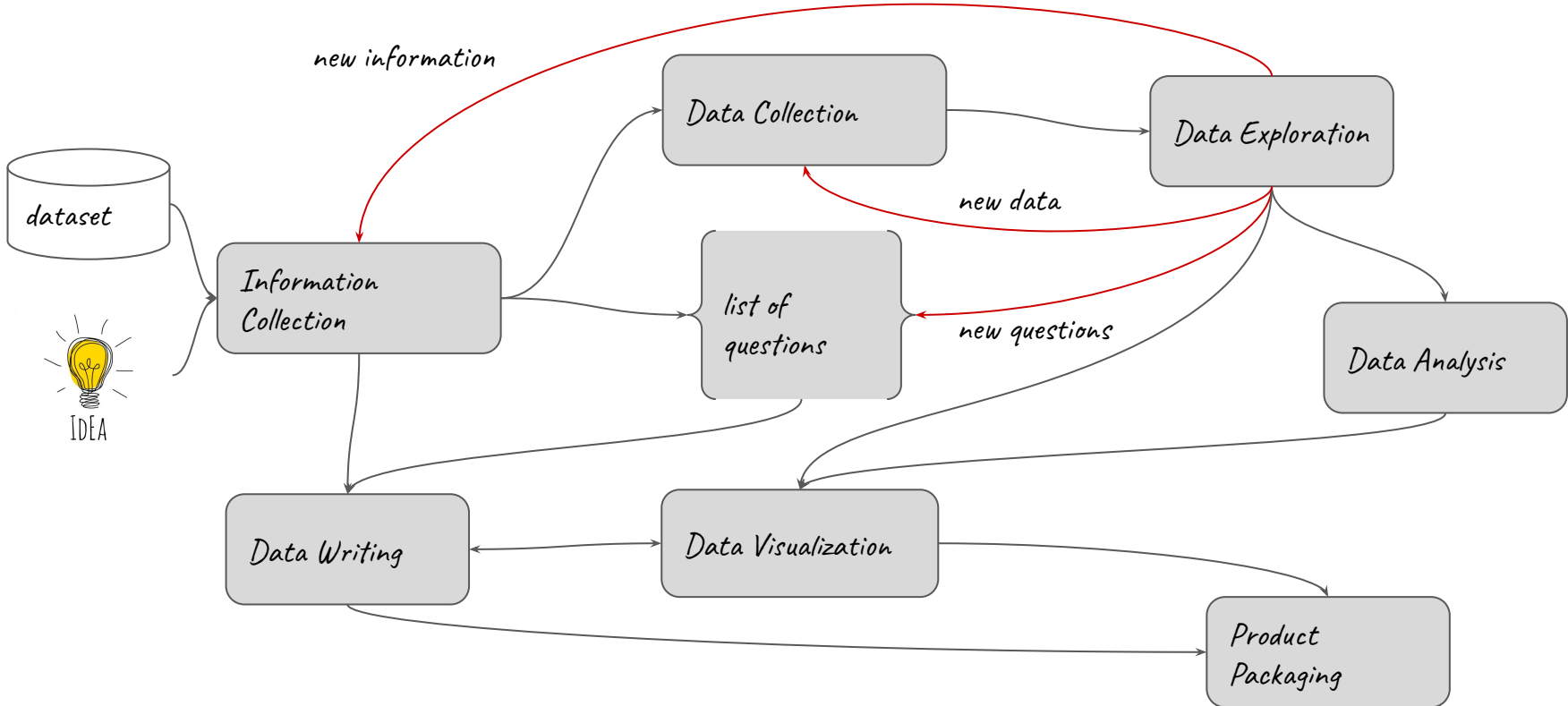
- Open Refine
 - <https://openrefine.org/download.html>

Data Exploration/Visualization

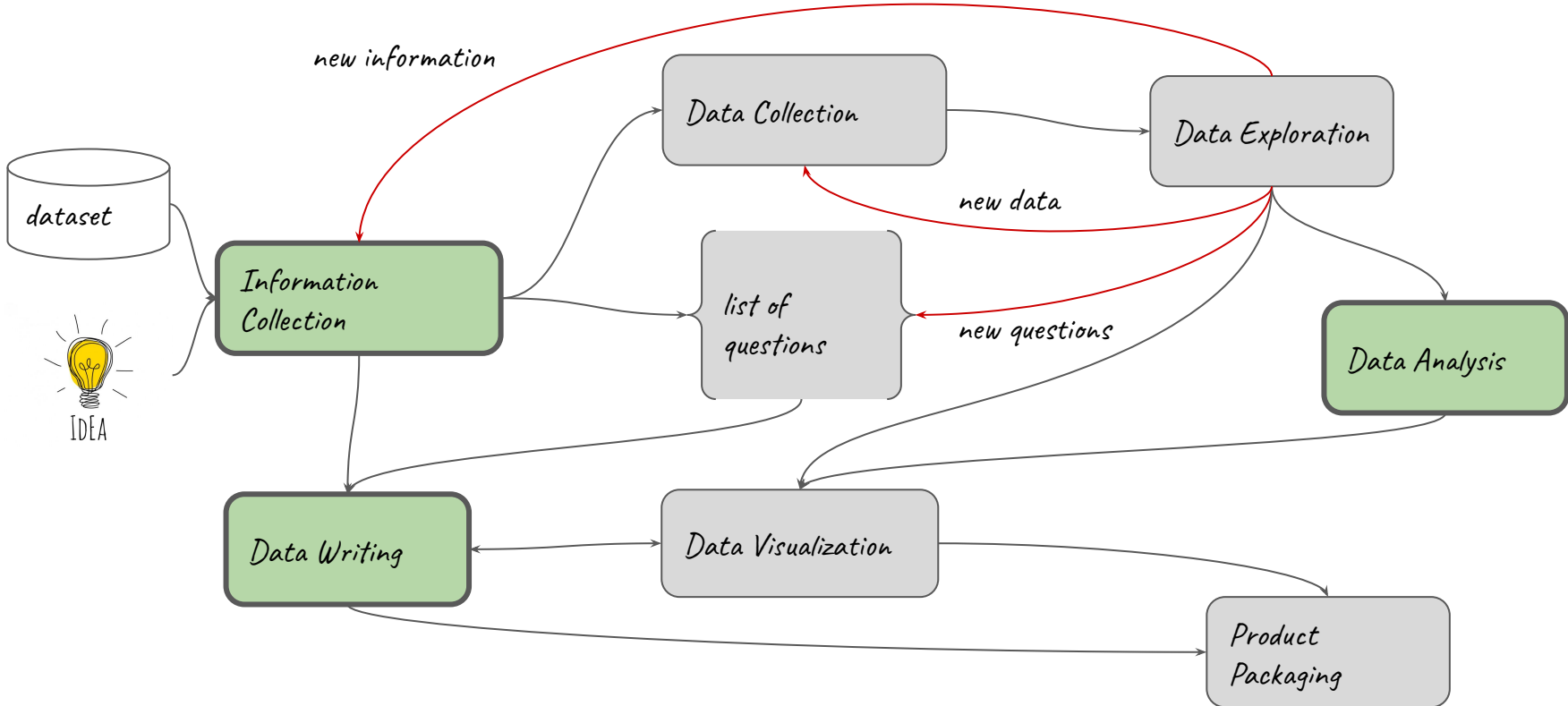
Data Exploration

- Tableau
 - <https://public.tableau.com/en-us/s/download>
 - <https://www.tableau.com/academic/students>

Data Journalism Workflow



Data Journalism Workflow - Altri Strumenti



Fase	Strumento
Information Collection	Web
Data Analysis	Librerie Python per l'analisi dei dati (non le vedremo nel corso)
Data Writing	Editor + Strumenti di assistenza alla scrittura



Information Collection

- **Data.gov:** Data.gov è un repository di dataset che sono stati forniti da diverse agenzie governative negli Stati Uniti. Questi dataset coprono una vasta gamma di argomenti, tra cui salute, istruzione, trasporti e altro ancora.
- **Google Dataset Search:** Google Dataset Search è un motore di ricerca appositamente progettato per aiutare i ricercatori a trovare dataset. Indicizza dataset da varie fonti, tra cui editori accademici, organizzazioni governative e repository di dati.
- **Open Data Network:** Open Data Network è una raccolta di portali di dati aperti provenienti da tutto il mondo. Puoi utilizzarlo per trovare dataset da una varietà di fonti, tra cui agenzie governative, organizzazioni no-profit e istituti accademici.



Information Collection (cont.)

- **Amazon Web Services (AWS) Public Datasets:** AWS offre una serie di dataset pubblici a cui puoi accedere gratuitamente. Questi dataset coprono una vasta gamma di argomenti, tra cui genomica, astronomia e altro ancora.
- **Data.world:** Data.world è una piattaforma che ospita un gran numero di dataset provenienti da una varietà di fonti. Fornisce anche strumenti per l'esplorazione e la collaborazione sui dati.
- **Reddit Datasets:** Il subreddit r/datasets è una comunità in cui le persone condividono dataset che hanno trovato o creato. Puoi sfogliare il subreddit o cercare argomenti specifici per trovare dataset che potrebbero essere utili per il tuo progetto.



dove posso trovare un dataset sui cambiamenti climatici?



Ci sono diverse fonti da cui puoi trovare dataset sui cambiamenti climatici. Ecco alcuni suggerimenti:



1. NASA's Global Climate Change: la NASA ha una vasta raccolta di dati sul clima, compresi i dati sul clima storico e quelli provenienti dai satelliti. Puoi trovare i dati sul loro sito web dedicato al cambiamento climatico: <https://climate.nasa.gov/vital-signs/>
2. NOAA Climate.gov: la National Oceanic and Atmospheric Administration degli Stati Uniti ha una vasta raccolta di dati sul clima, tra cui i dati sulle temperature, le precipitazioni e altri fattori climatici. Puoi trovare questi dati sul loro sito web dedicato al clima: <https://www.climate.gov/maps-data/datasets>
3. Berkeley Earth: Berkeley Earth è un'organizzazione indipendente che raccoglie dati sul clima da tutto il mondo. Puoi trovare i loro dati sul loro sito web: <http://berkeleyearth.org/data/>
4. European Climate Data and Information: l'ECMWF (European Centre for Medium-Range Weather Forecasts) raccoglie e distribuisce dati sul clima in Europa. Puoi trovare i loro dati sul loro sito web: <https://www.ecmwf.int/en/forecasts/datasets>
5. World Bank Climate Change Knowledge Portal: la Banca Mondiale offre una vasta raccolta di dati e informazioni sul clima provenienti da tutto il mondo. Puoi trovare i loro dati sul loro portale: <https://climateknowledgeportal.worldbank.org/>

Questi sono solo alcuni dei molti posti dove puoi trovare dataset sul cambiamento

...ChatGPT...

Data Writing

- Editor di Testo (Word, Open Office, Google Doc, Notepad...)
- Strumenti di assistenza:
 - Grammarly
 - Quillbot
 - ...
- Grammarly e Quillbot forniscono:
 - la versione gratuita e a pagamento
 - un plugin per Chrome