The Environmental Data Initiative (EDI)

Make metadata with the EML assembly line

EDI is funded by the NSF DEB

5 Phases of Publishing Ecological Data

1. Assemble data and metadata
2. Format and QC data tables
3. Create EML metadata
4. Submit your data package (data and metadata) to repository
5. Cite your data package

5 Phases of Publishing Ecological Data

1. Assemble data and metadata
2. **Format and QC data tables**
3. **Create EML metadata**
4. Submit your data package (data and metadata) to repository
5. Cite your data package

www.environmentaldatainitiative.org

Quality metadata is VERY important!

- For understanding:
  - What the data are
  - When the data were collected
  - How the data were collected
  - Who collected the data
  - How the data have been modified
- For assessing fitness for use
- Critical to:
  - Open and reproducible science
  - Synthesis science
  - Extending the life and value of data

Structured metadata is VERY important!

- If not structured then:
  - Can’t search for the data in repositories or aggregators
  - Can’t search content at the dataset level
  - The speed of science and knowledge formation is slowed
  - Potential information loss
  - Can’t convert to other metadata standards

But, making quality EML is challenging!

- Creating quality EML requires:
    - Detailed info about: methods, data entities, personnel, keywords, etc.
    - Detailed technical info about data entities
    - Understanding of the EML schema (where content can go)
    - Understanding EML best practices (where content should go)
    - Construction of EML (how to create it)

But, making quality EML is challenging!

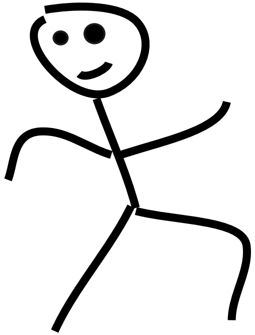
- Creating quality EML requires:
  - Detailed info about: methods, data entities, personnel, keywords, etc.
  - Detailed technical info about data entities
  - Understanding of the EML schema (where content can go)
  - Understanding EML best practices (where content should go)
  - Construction of EML (how to create it)

The EMLassemblyline R code package

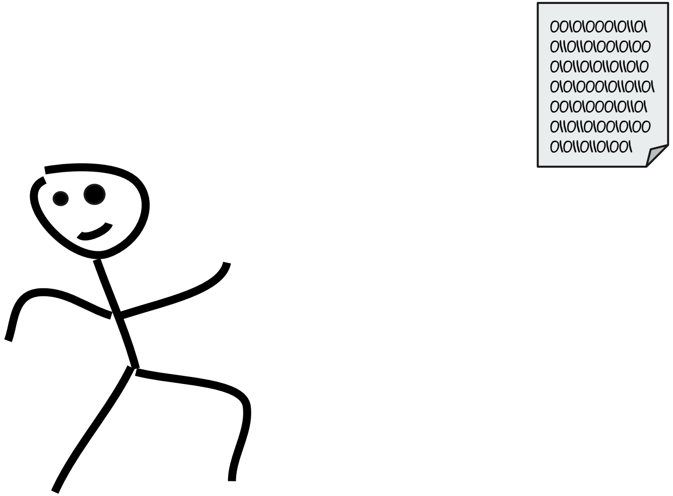
- User:
  - Supplies core information in a familiar way (tables, text files)
- EMLassemblyline:
  - Provides meaningful error messages and guidance
  - Extracts technical info about data entities
  - Builds and validates the EML
  - Embodies EML best practices
  - Simplifies versioning
  - Facilitates automation

Step 1:

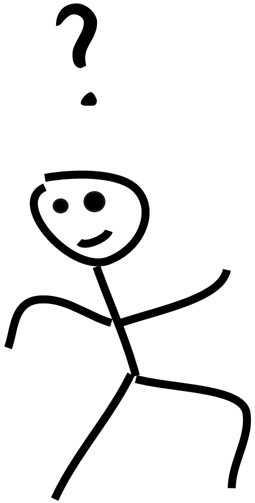
00101000101101
011011010010100
0101101011011010
0101000101101101
00101000101101
011011010010100
0101101101001



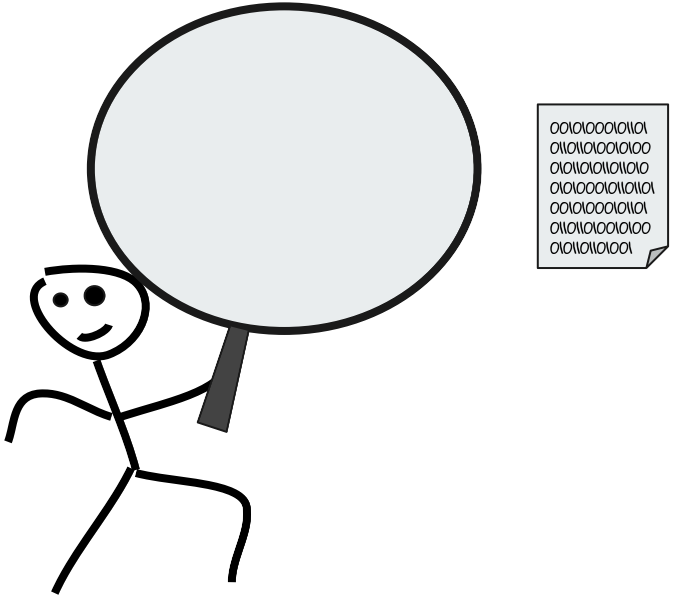
Step 1: Identify data type(s)



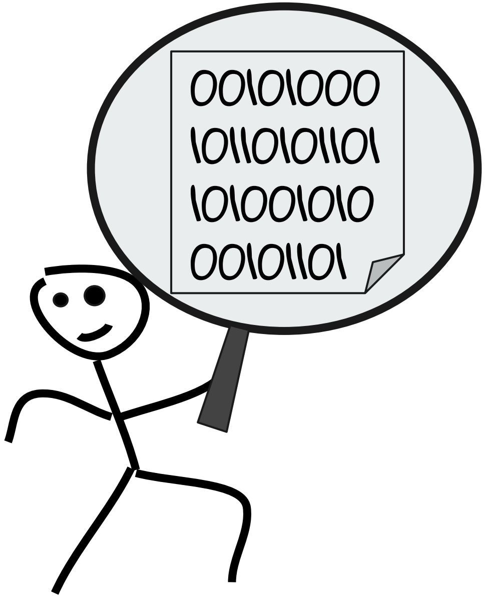
Step 1: Identify data type(s)



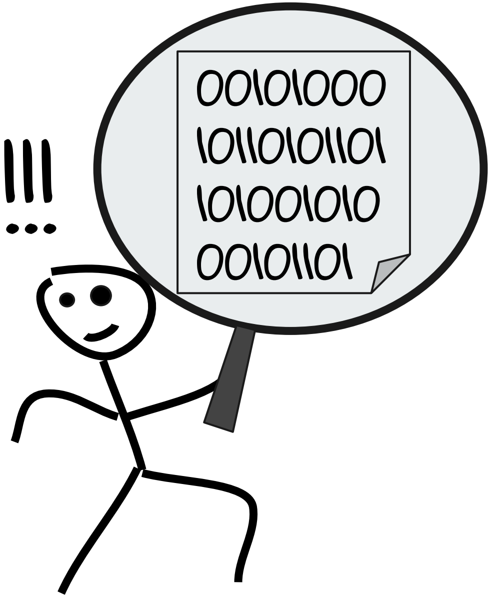
Step 1: Identify data type(s)



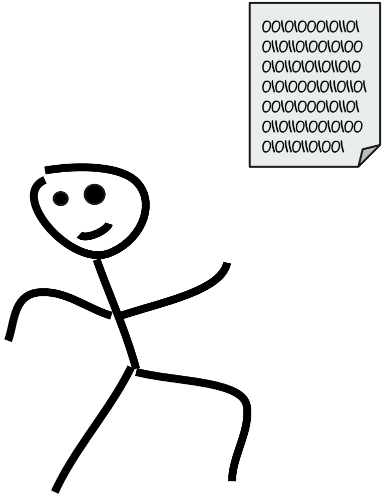
Step 1: Identify data type(s)



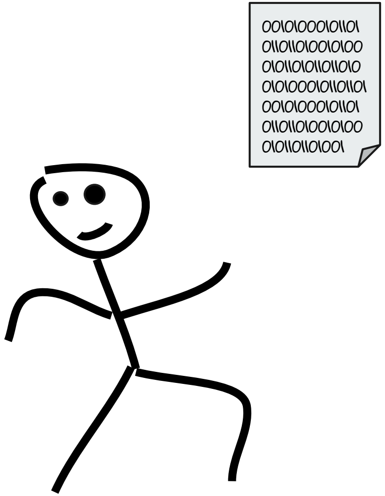
Step 1: Identify data type(s)



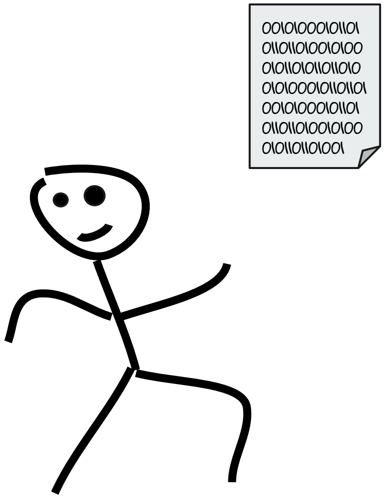
Step 1: Identify data type(s)



Step 2:


00101000101101
011011010010100
0101101011011010
0101000101101101
00101000101101
011011010010100
0101101101001

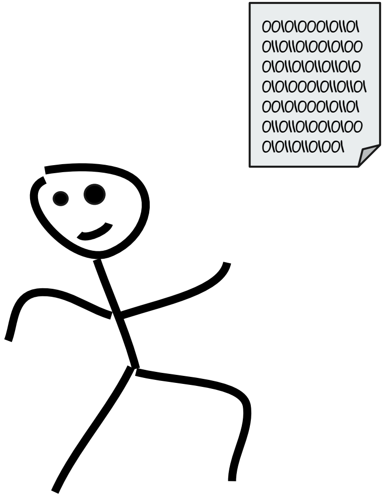
Step 2: Consult instructions



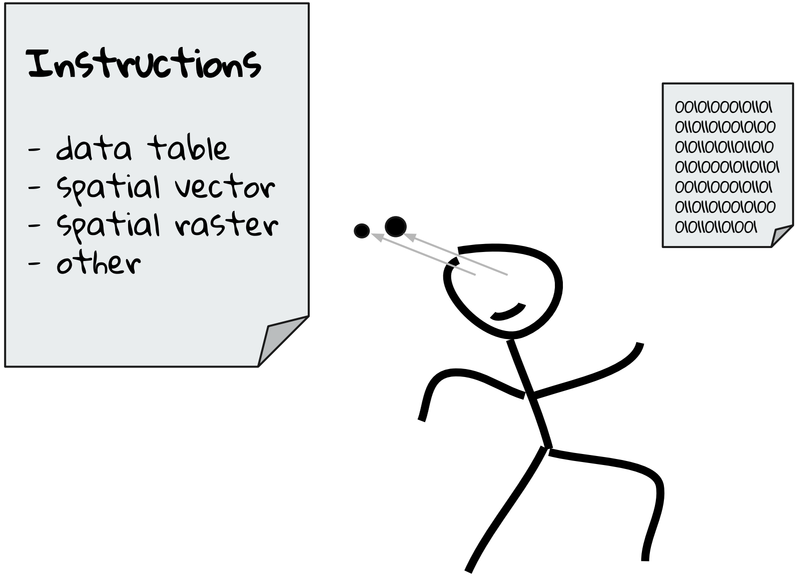
Step 2: Consult instructions

Instructions

- data table
- spatial vector
- spatial raster
- other



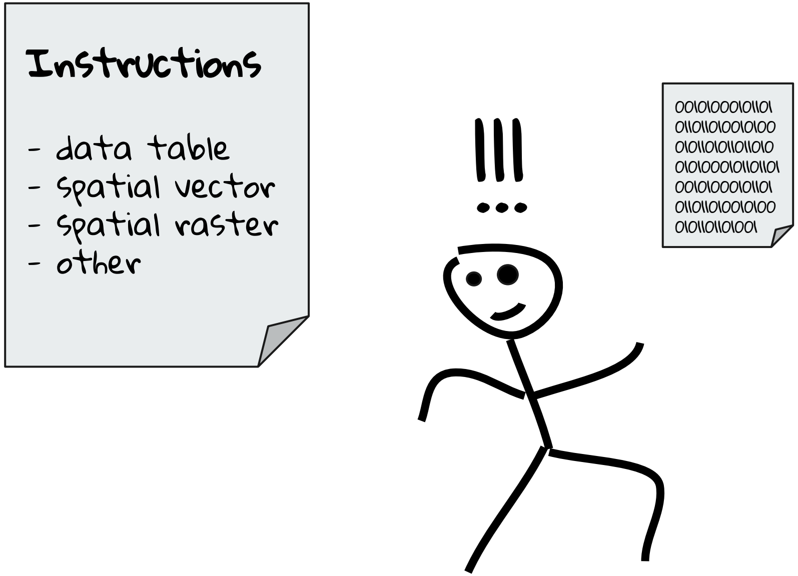
Step 2: Consult instructions



Instructions

- data table
- spatial vector
- spatial raster
- other

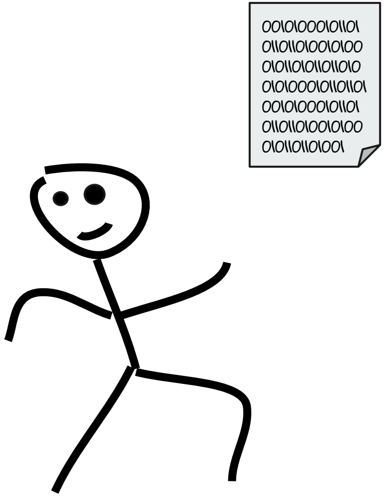
Step 2: Consult instructions



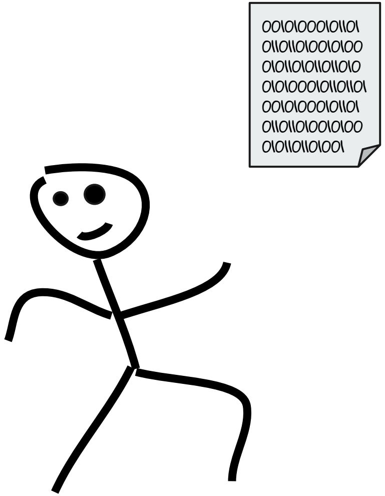
Instructions

- data table
- spatial vector
- spatial raster
- other

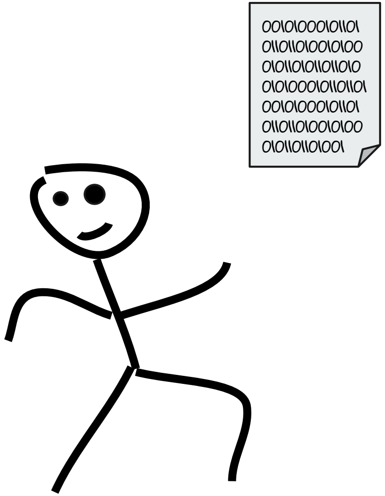
Step 2: Consult instructions



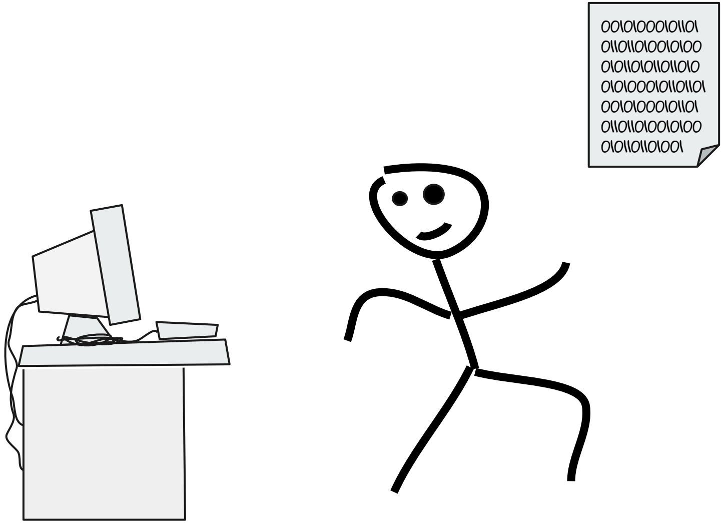
Step 3:


00101000101101
011011010010100
0101101011011010
0101000101101101
00101000101101
011011010010100
0101101101001

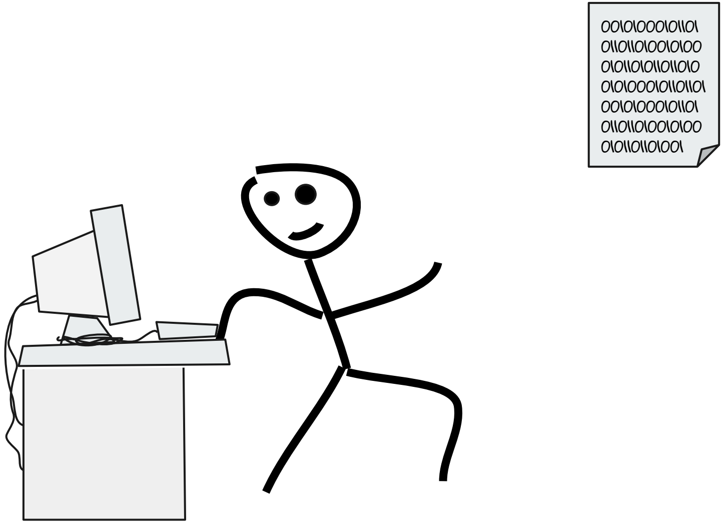
Step 3: Import templates



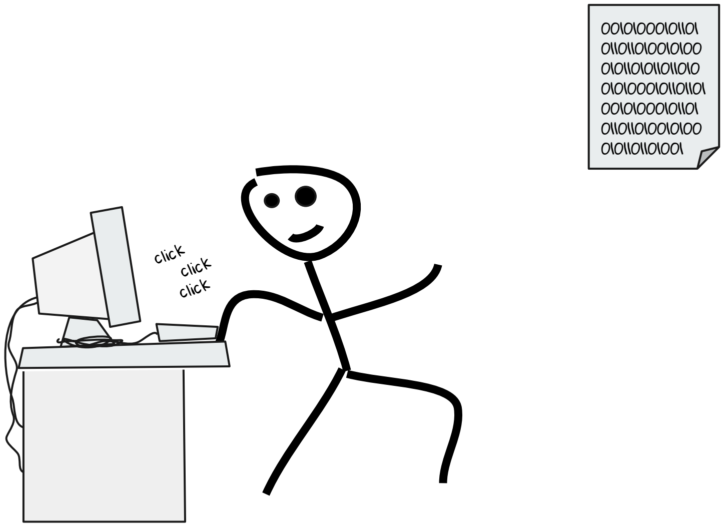
Step 3: Import templates



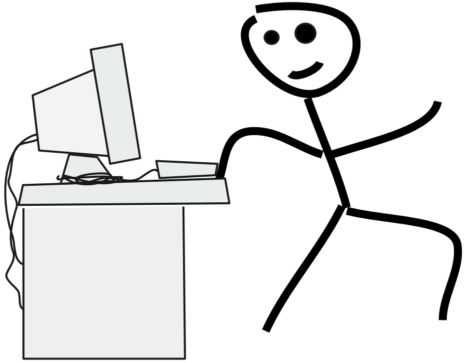
Step 3: Import templates



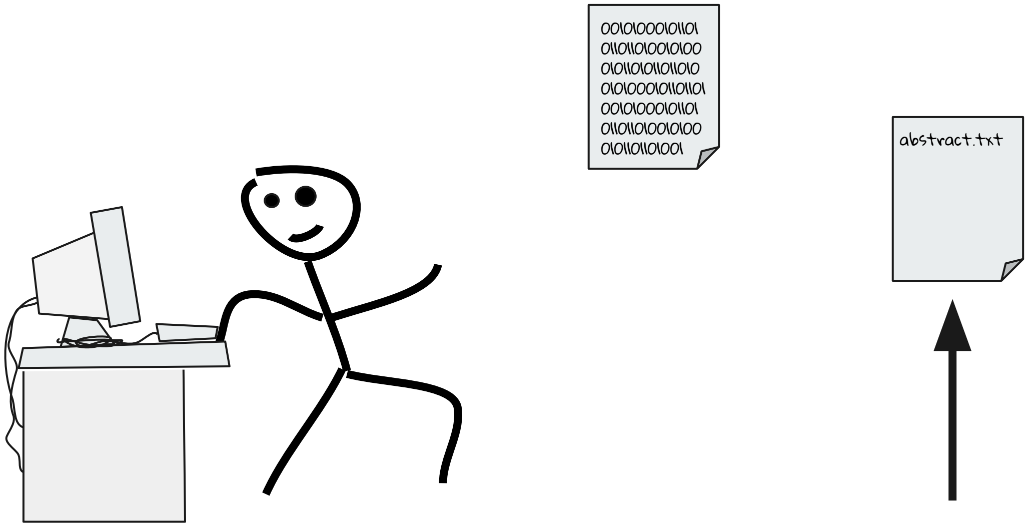
Step 3: Import templates



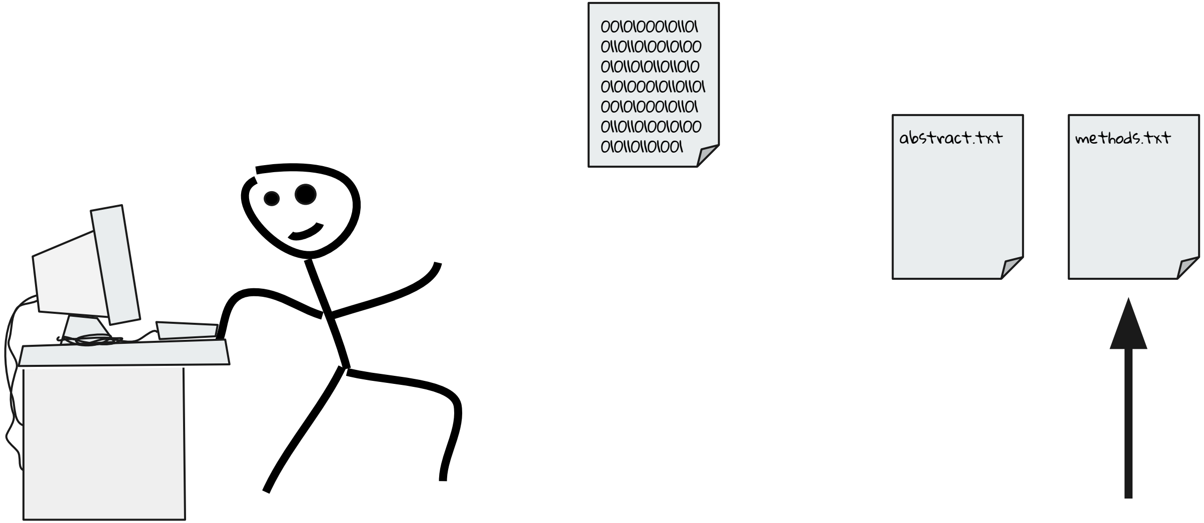
Step 3: Import templates



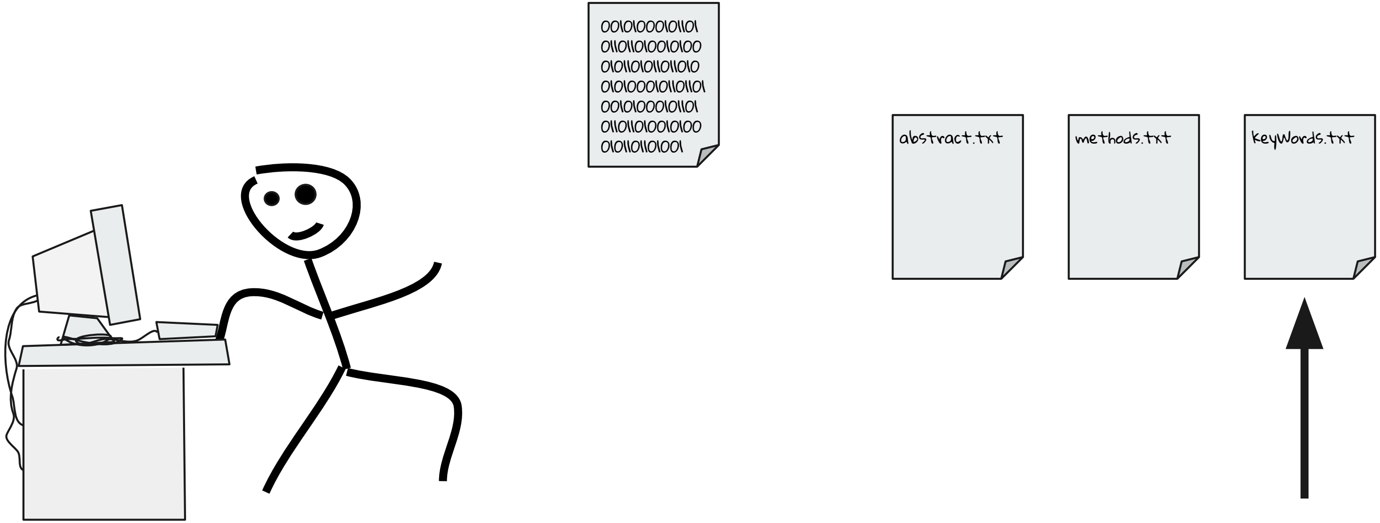
Step 3: Import templates



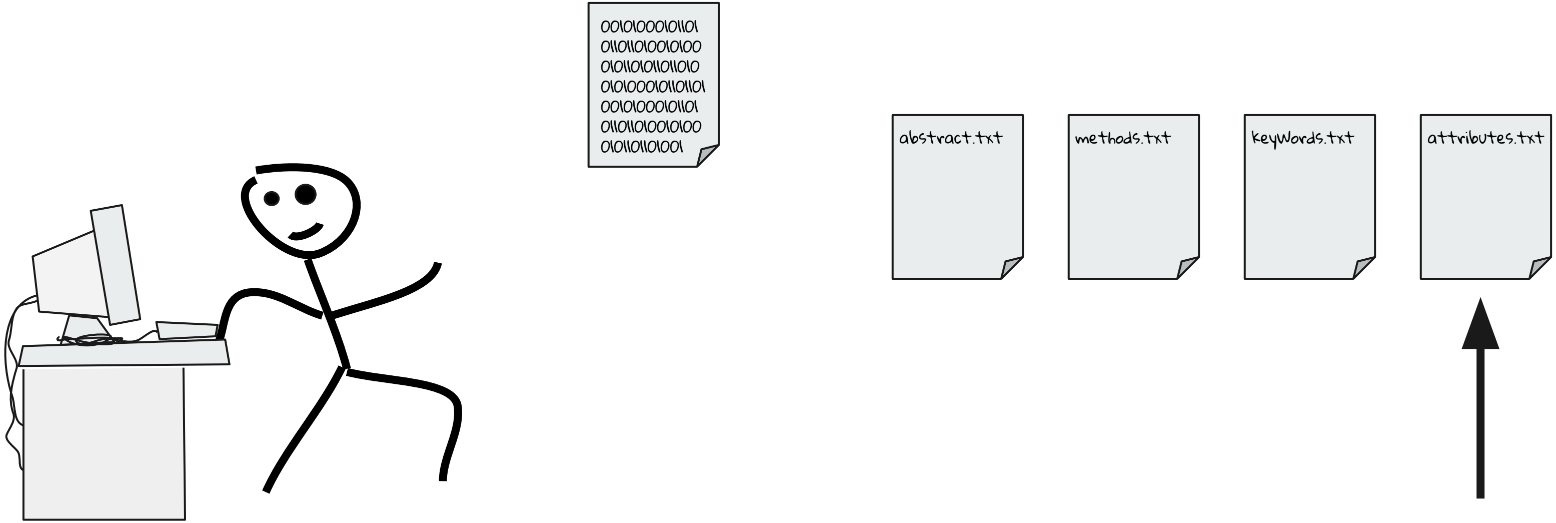
Step 3: Import templates



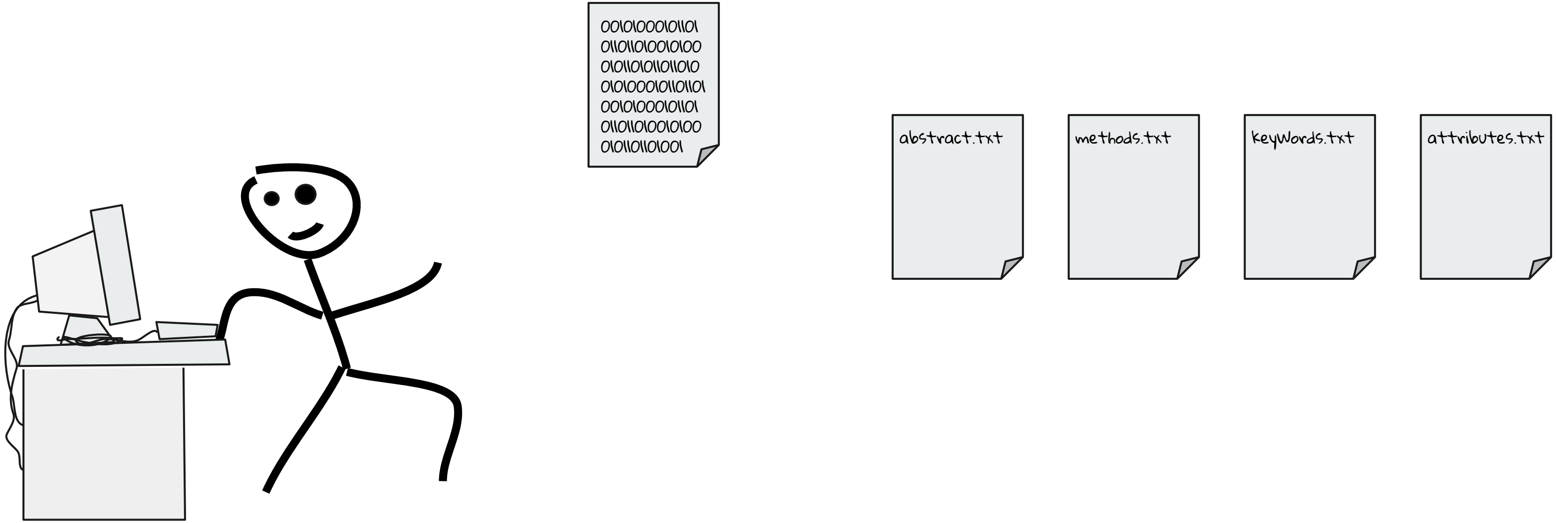
Step 3: Import templates



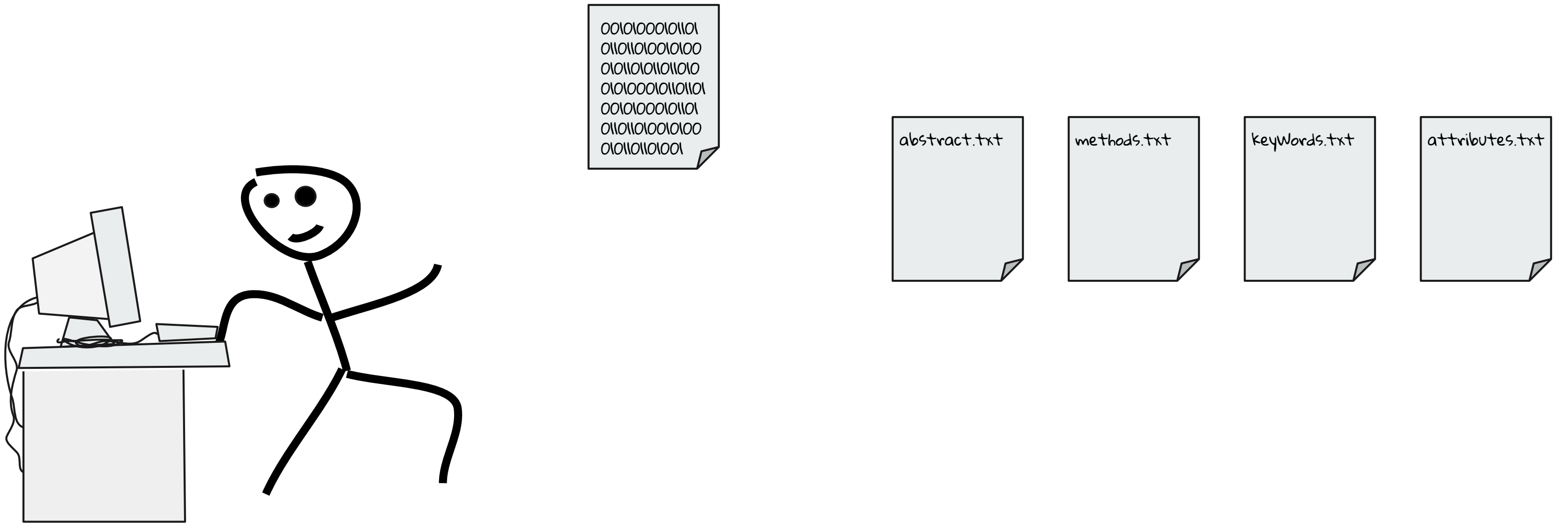
Step 3: Import templates



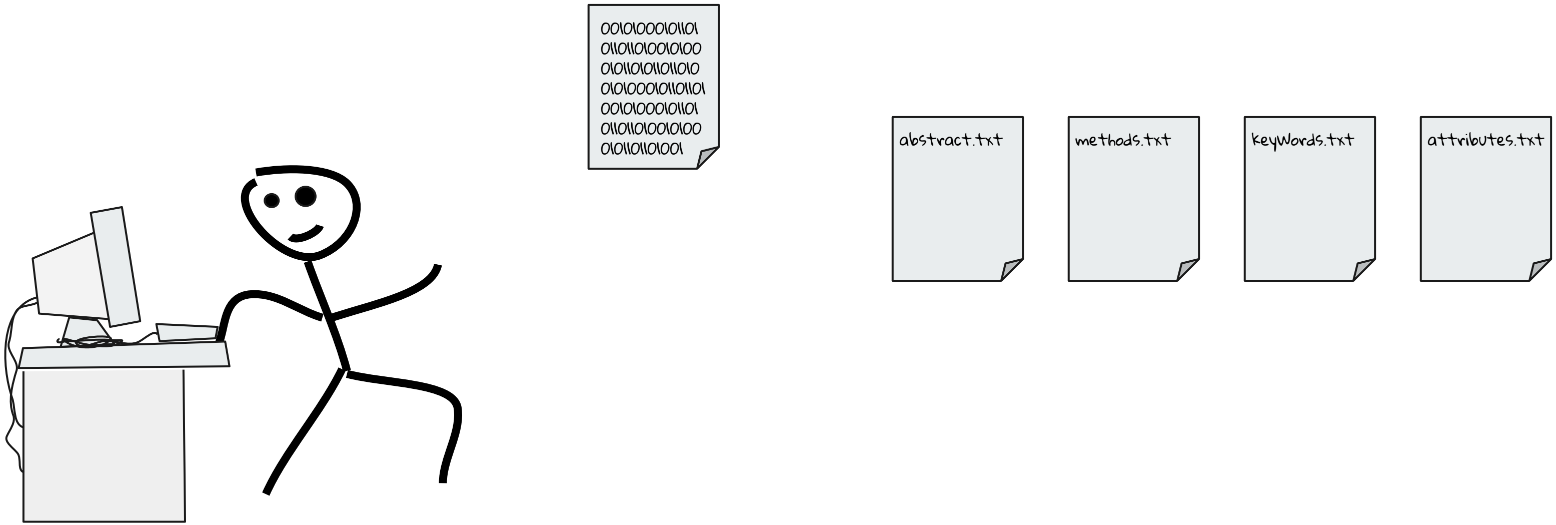
Step 3: Import templates



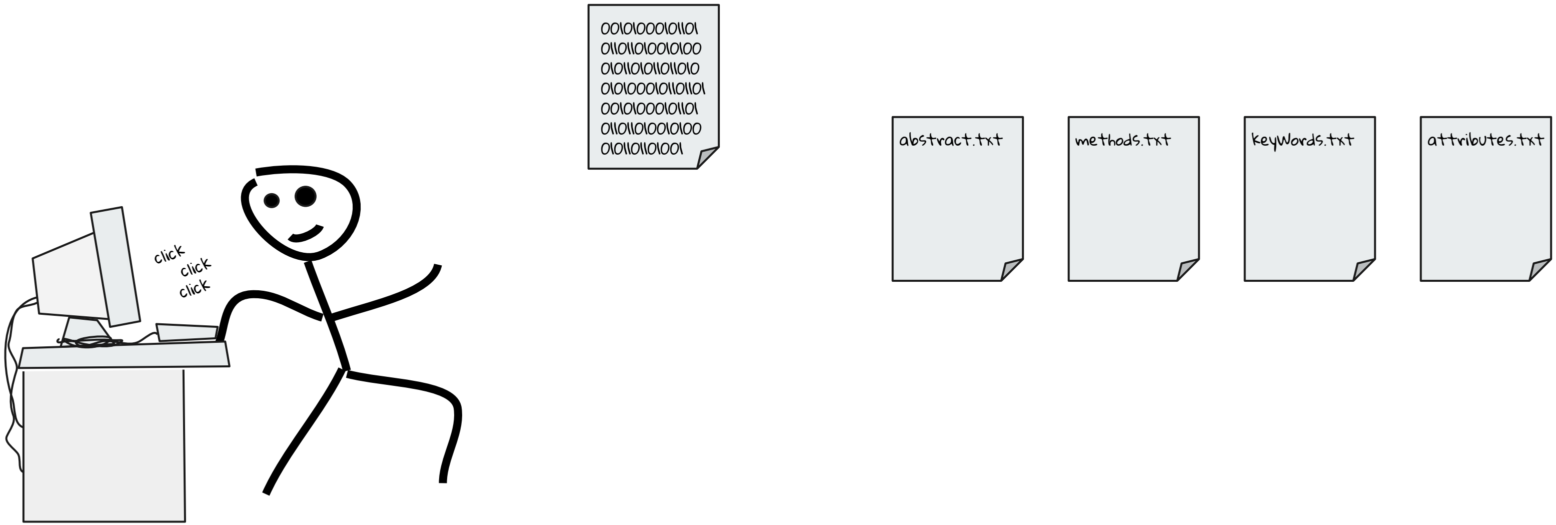
Step 4:


00101000101101
011011010010100
0101101011011010
0101000101101101
00101000101101
011011010010100
0101101101001
abstract.txt
methods.txt
keyWords.txt
attributes.txt

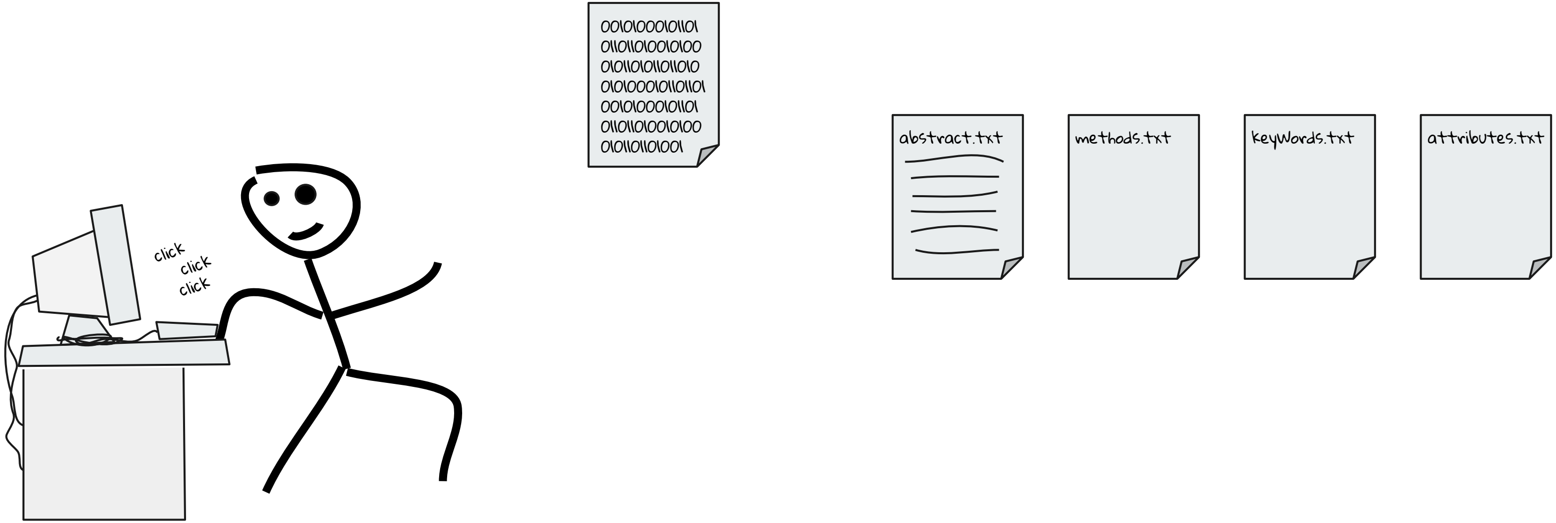
Step 4: Complete templates



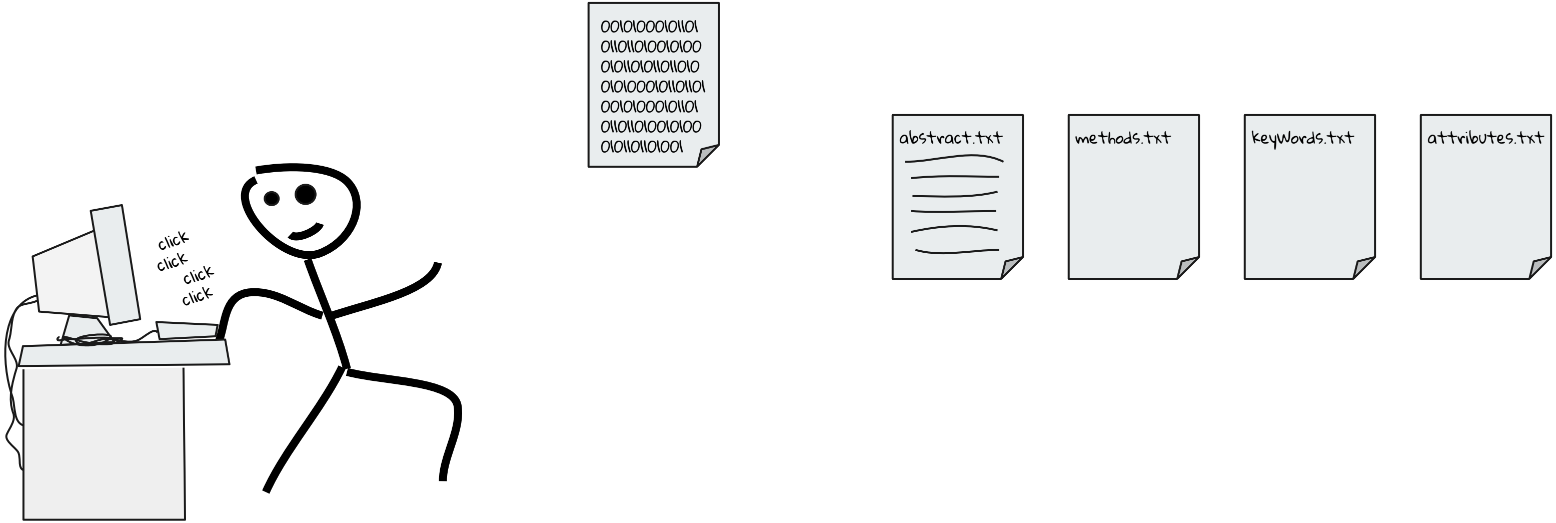
Step 4: Complete templates



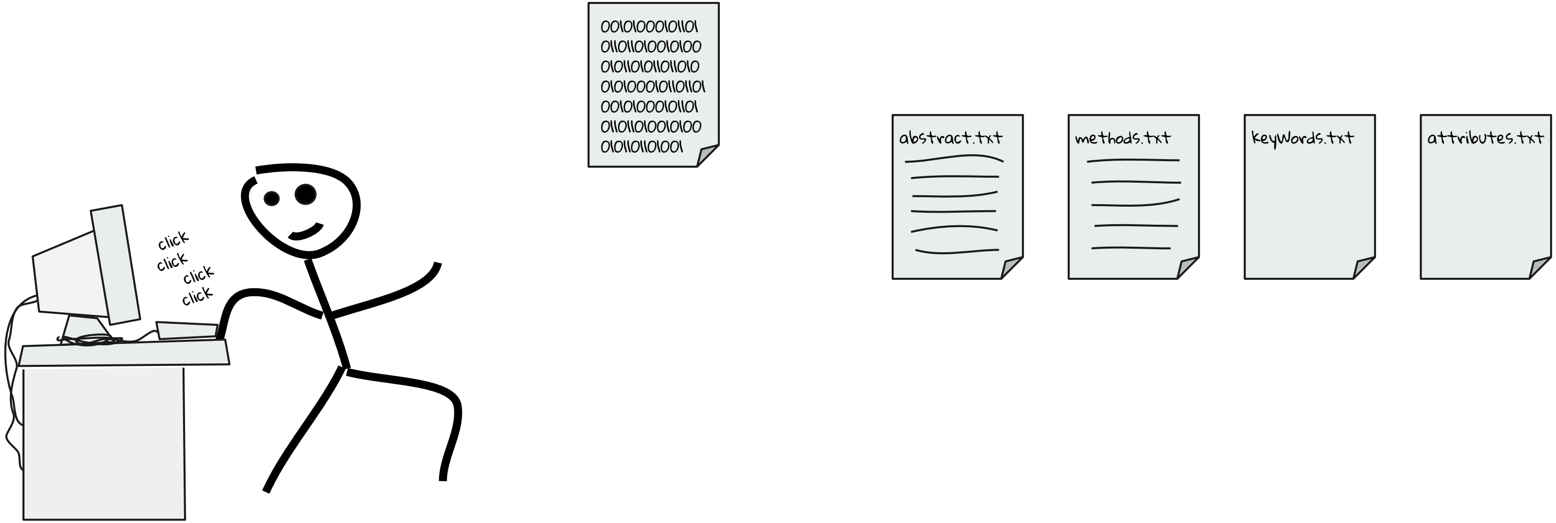
Step 4: Complete templates



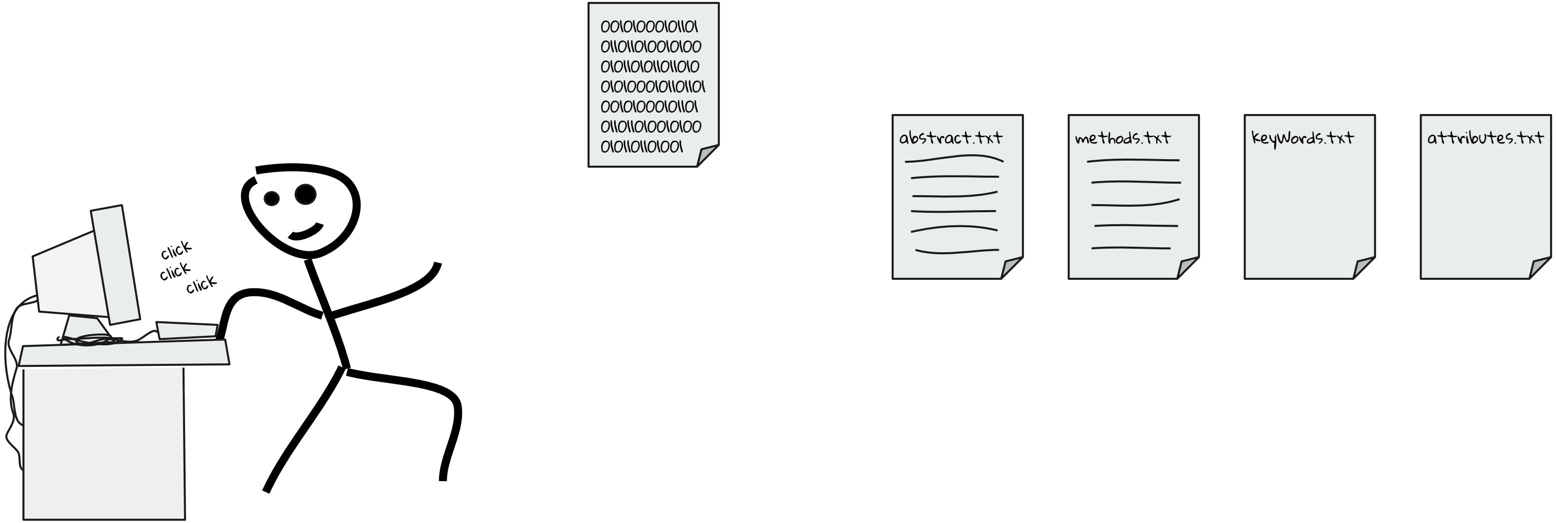
Step 4: Complete templates



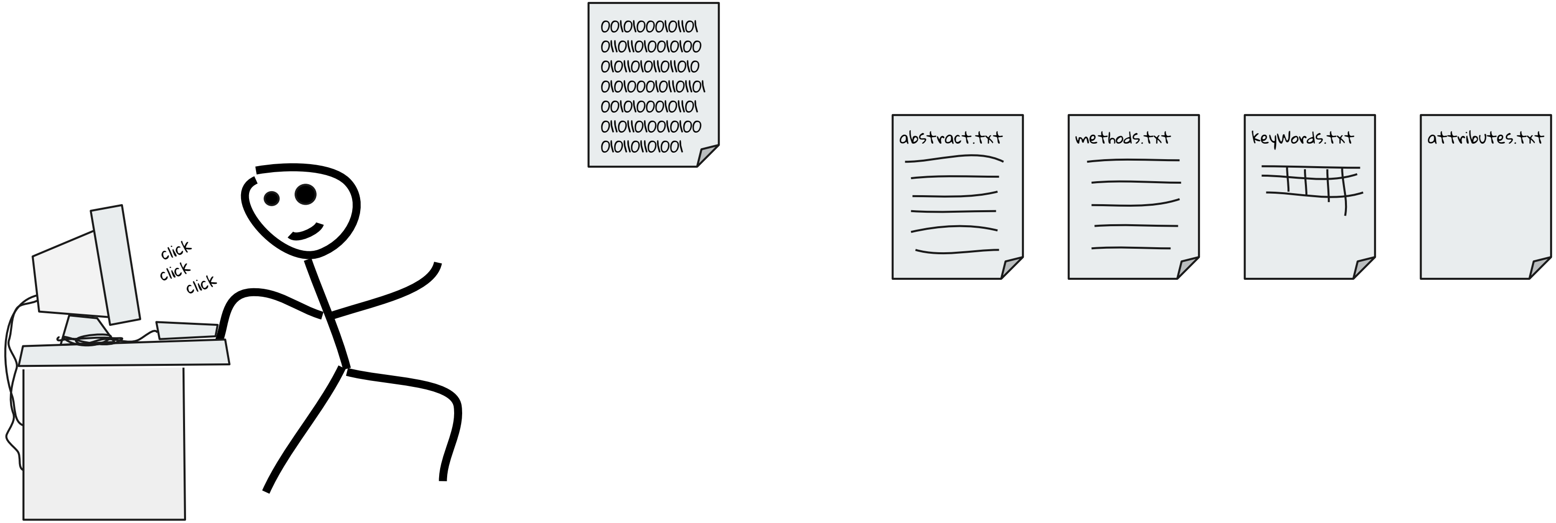
Step 4: Complete templates



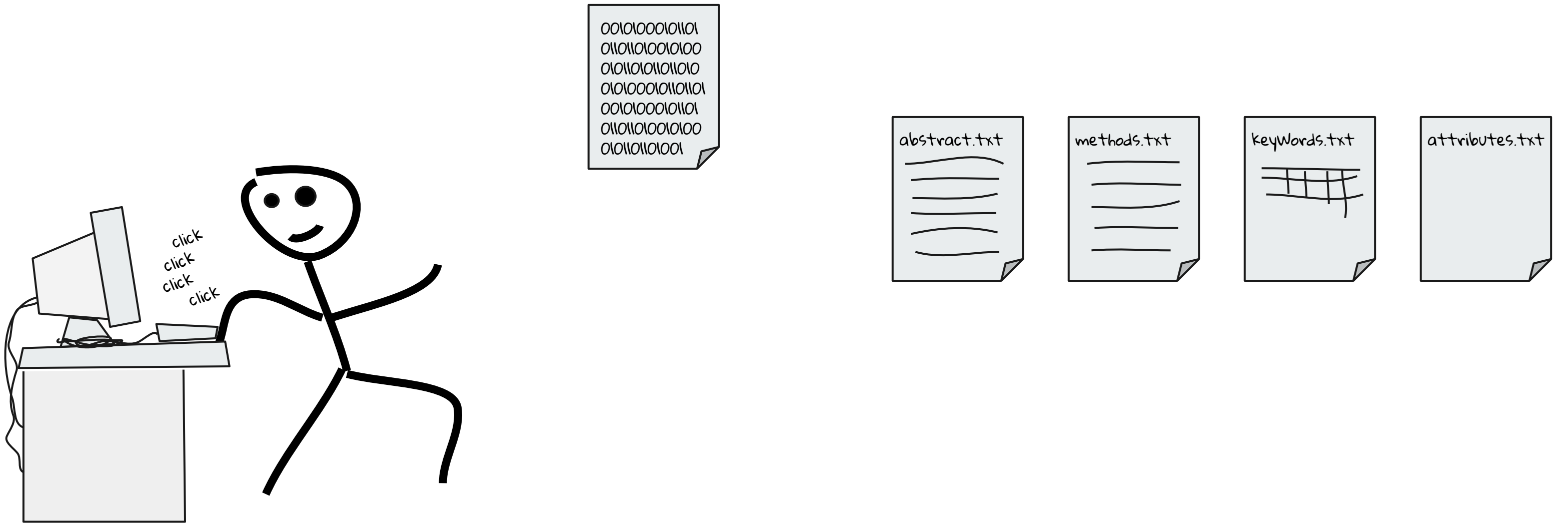
Step 4: Complete templates



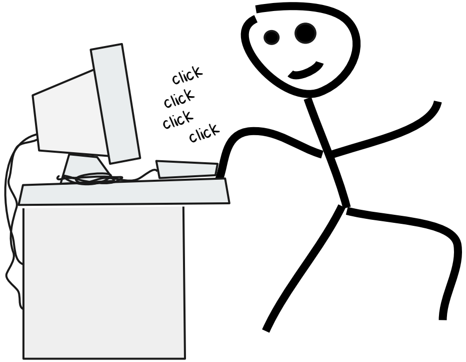
Step 4: Complete templates

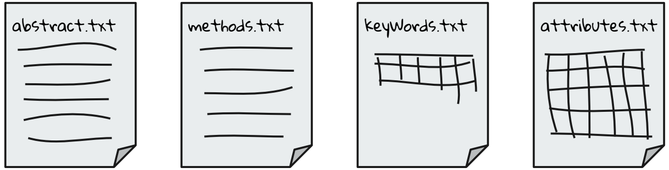


Step 4: Complete templates

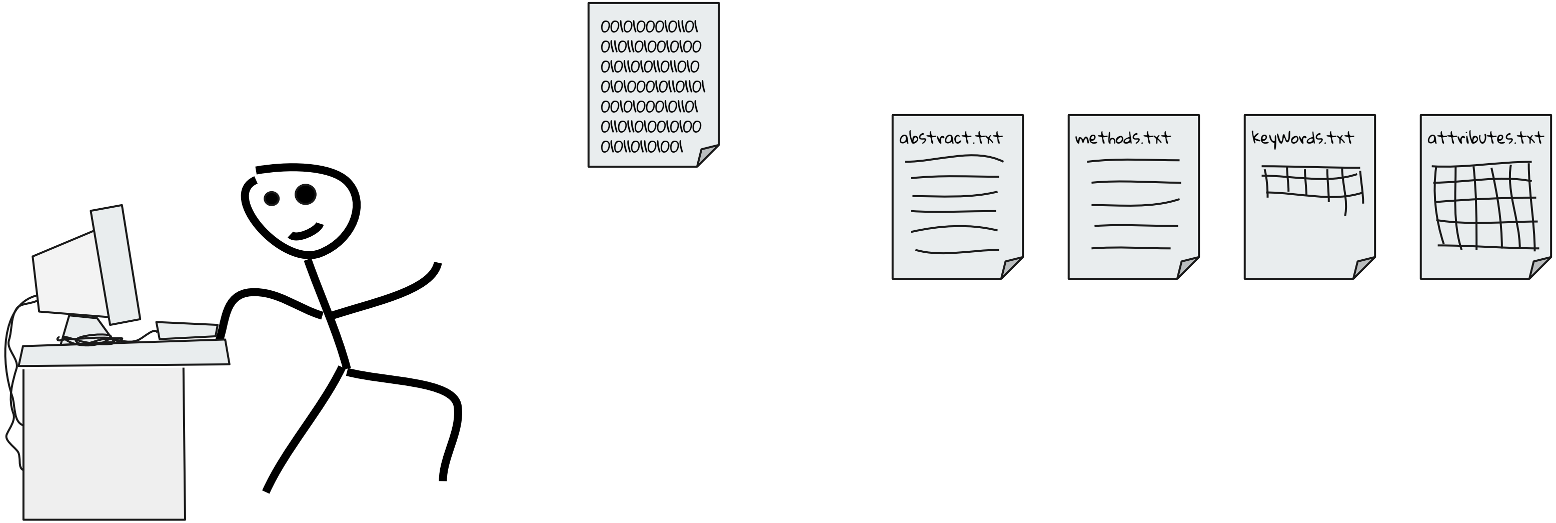


Step 4: Complete templates

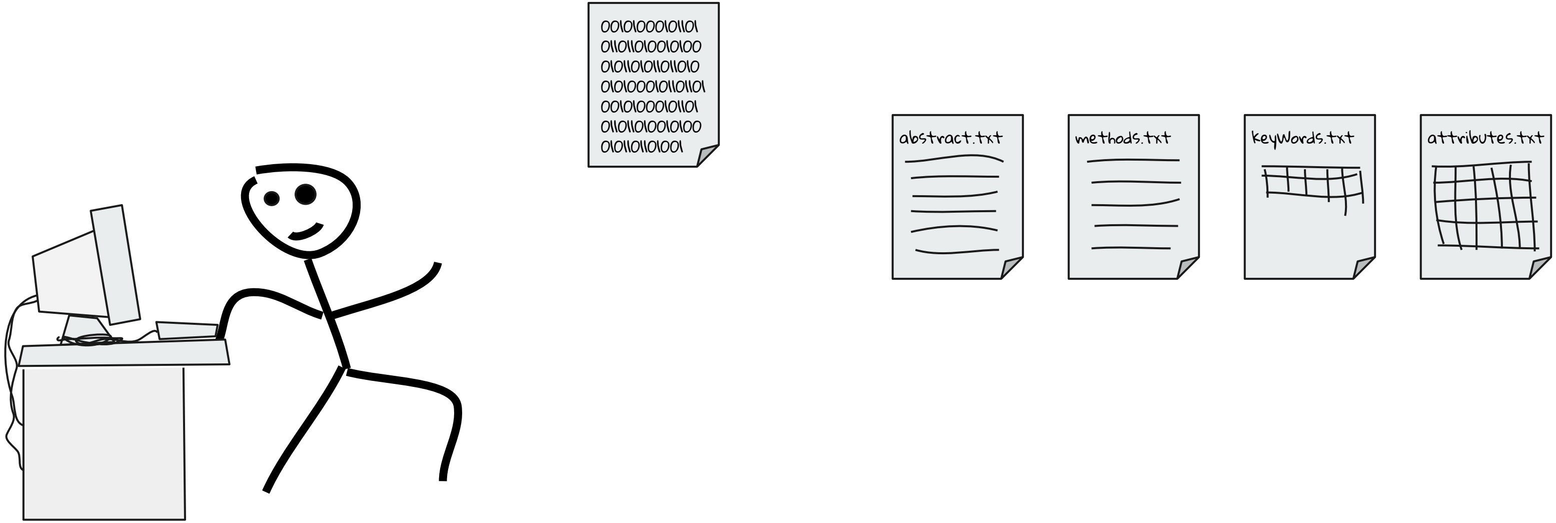




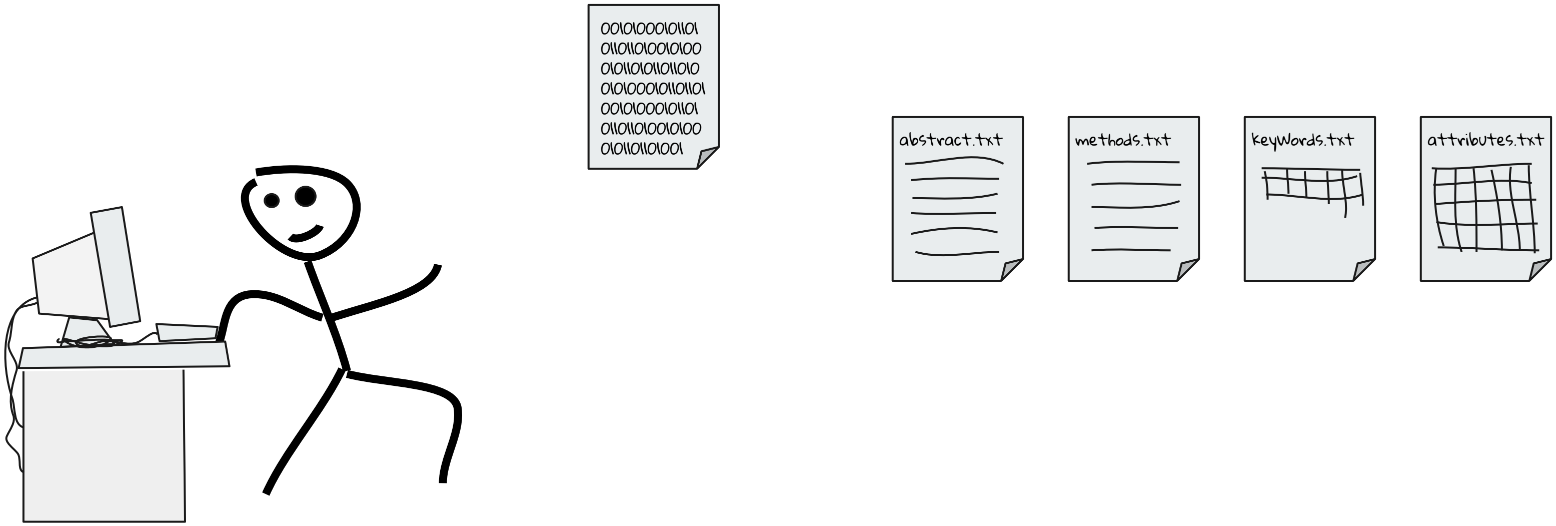
Step 4: Complete templates



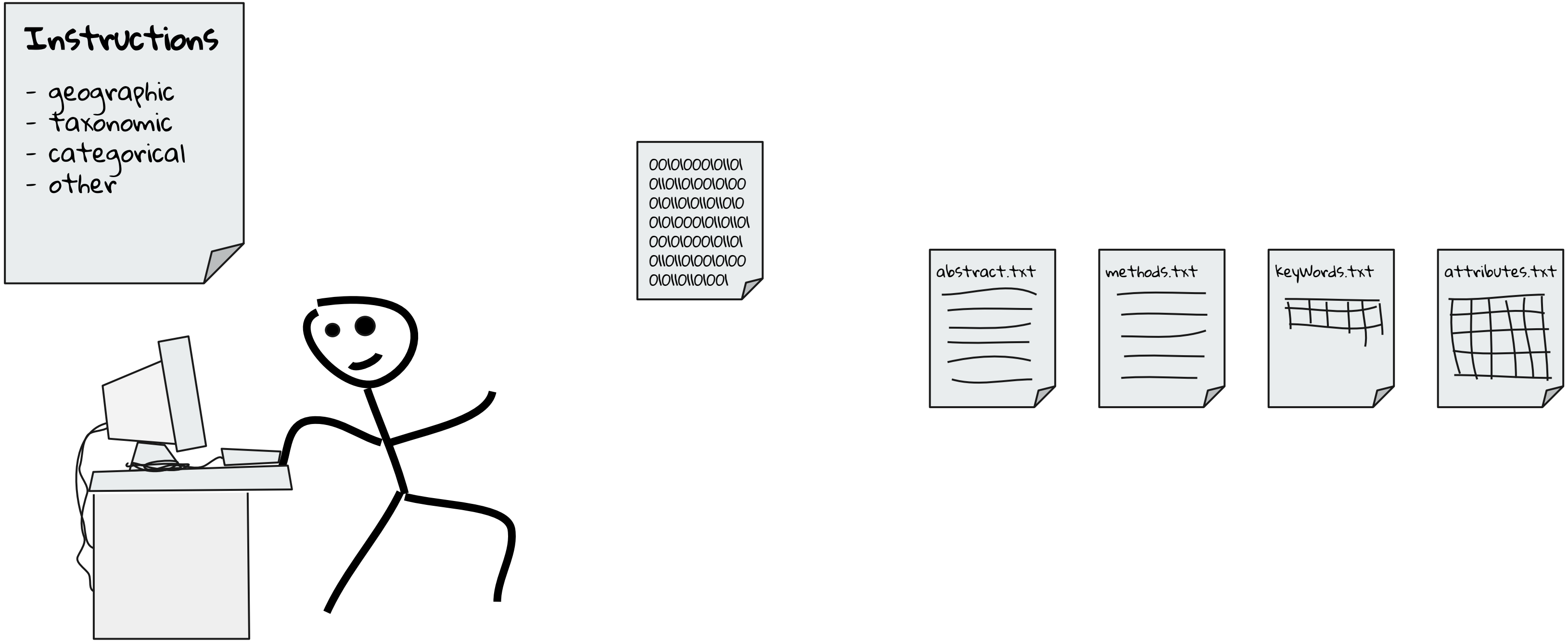
Step 5:


00101000101101
011011010010100
0101101011011010
0101000101101101
00101000101101
011011010010100
0101101101001
abstract.txt
methods.txt
keywords.txt
attributes.txt

Step 5: Add custom details



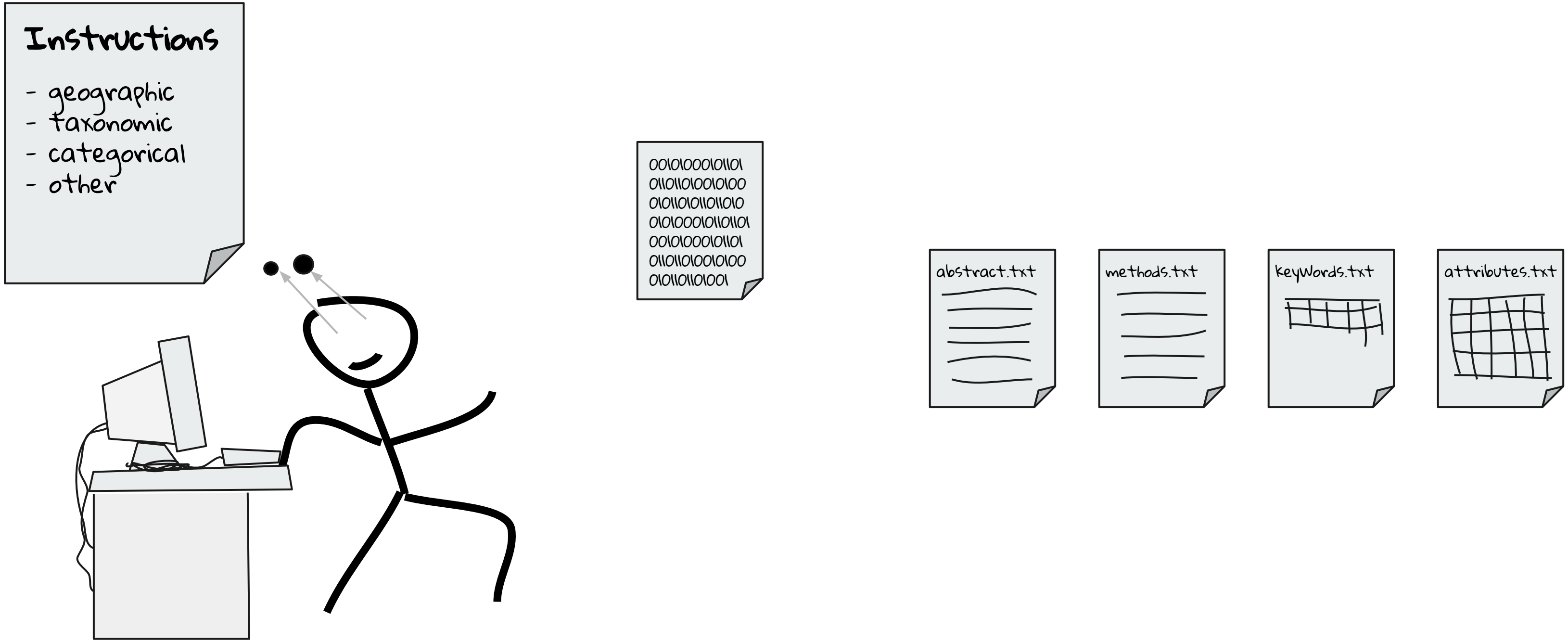
Step 5: Add custom details



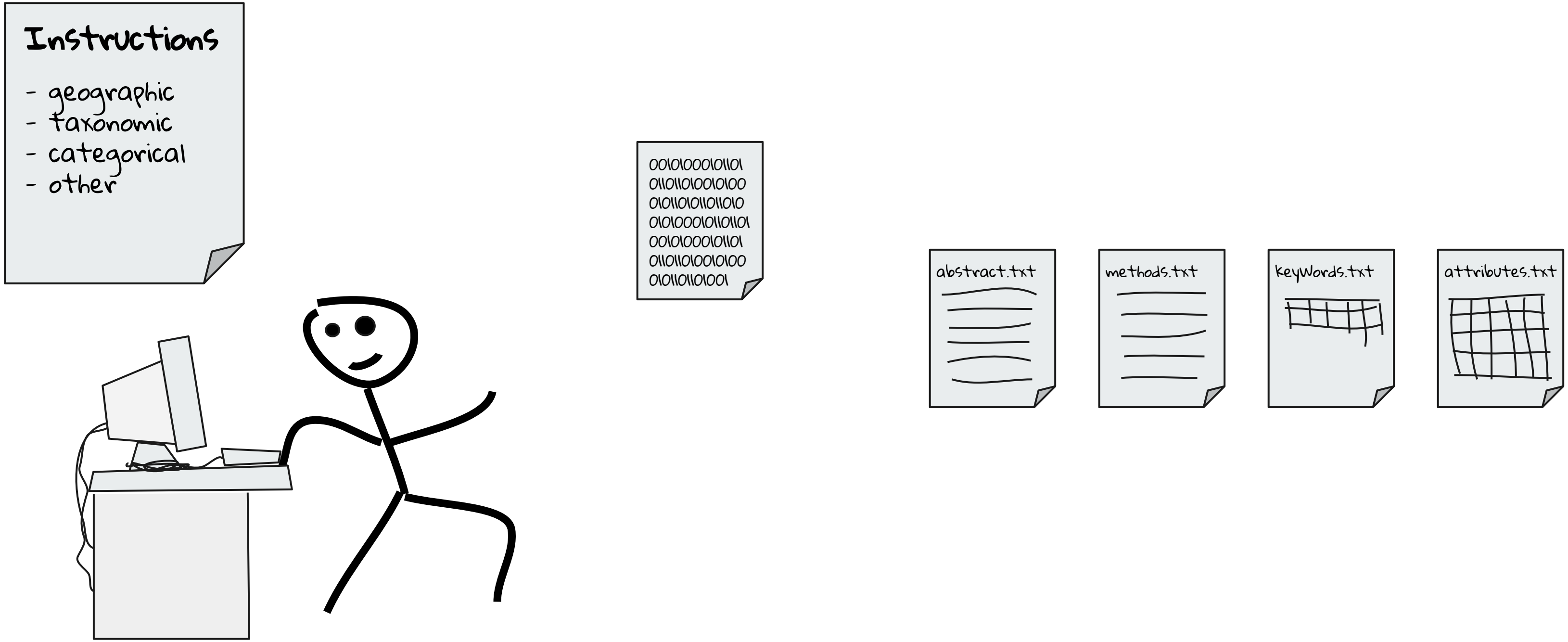
Instructions

- geographic
- taxonomic
- categorical
- other

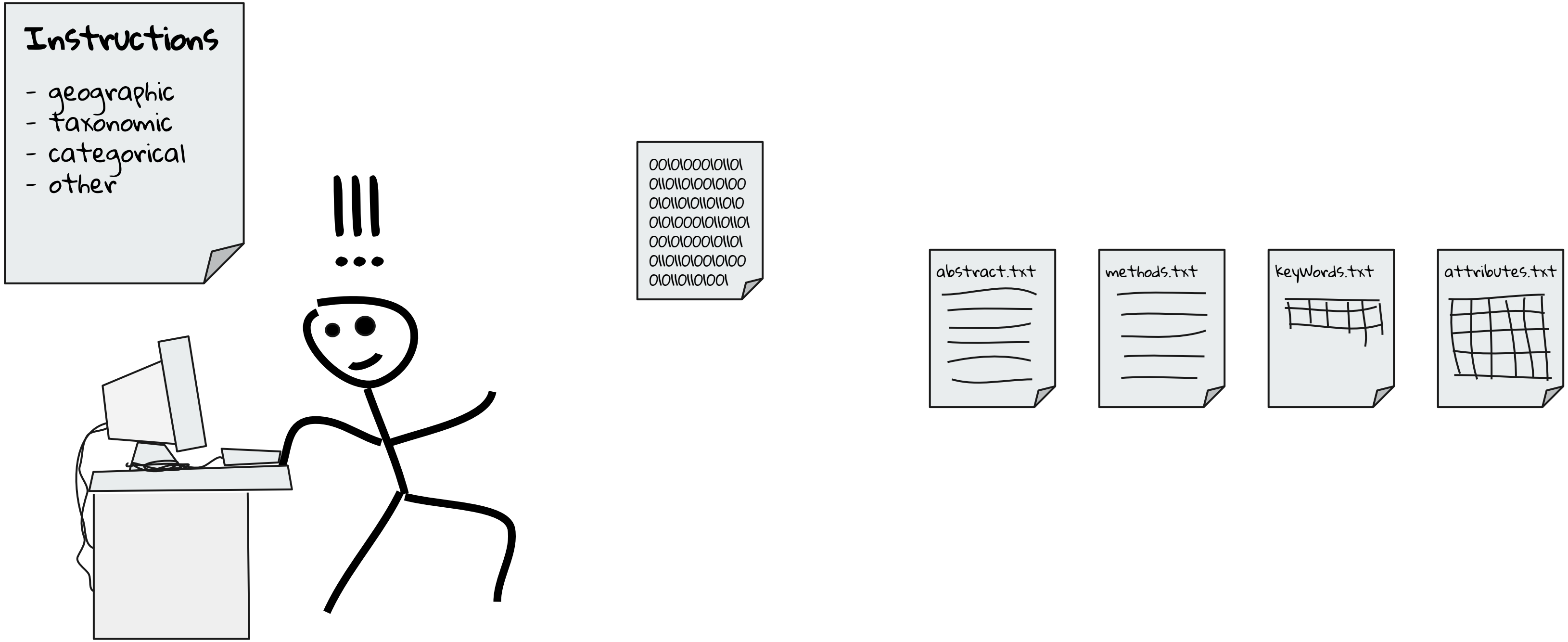
Step 5: Add custom details



Step 5: Add custom details



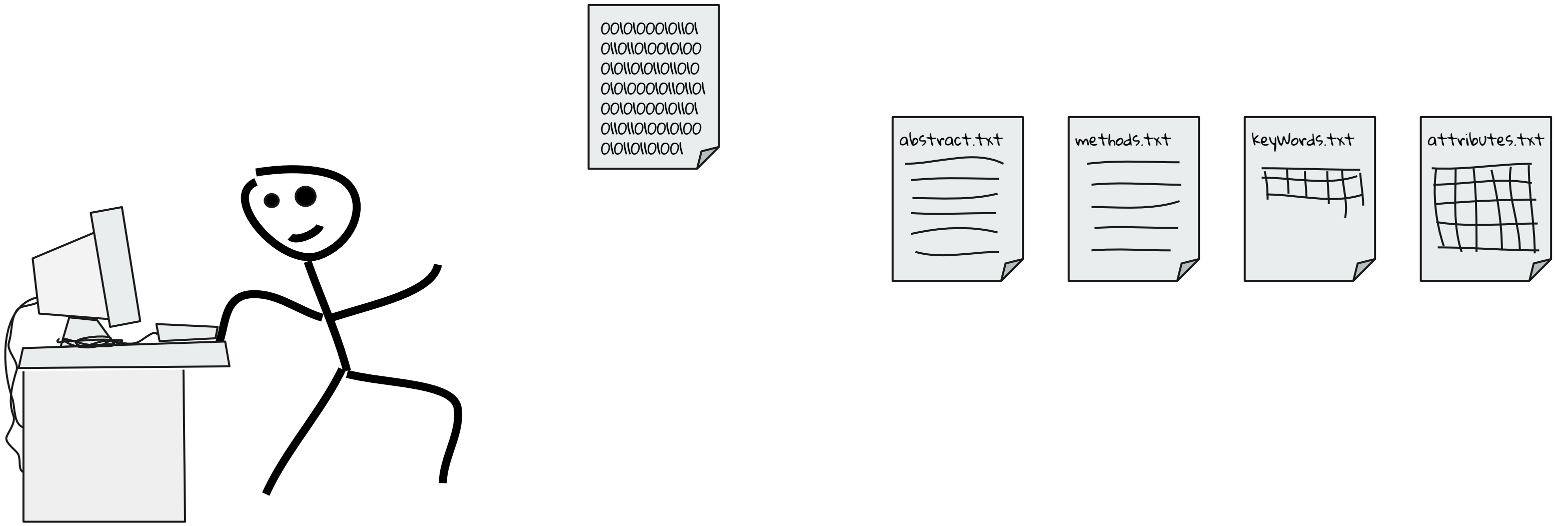
Step 5: Add custom details



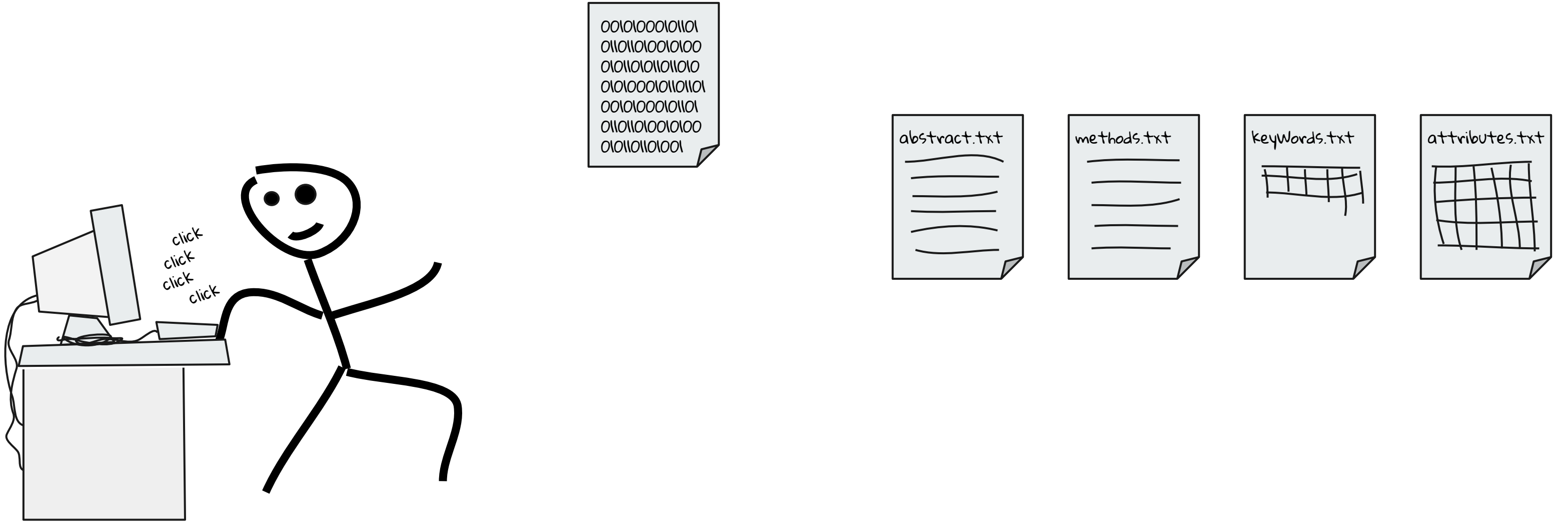
Instructions

- geographic
- taxonomic
- categorical
- other

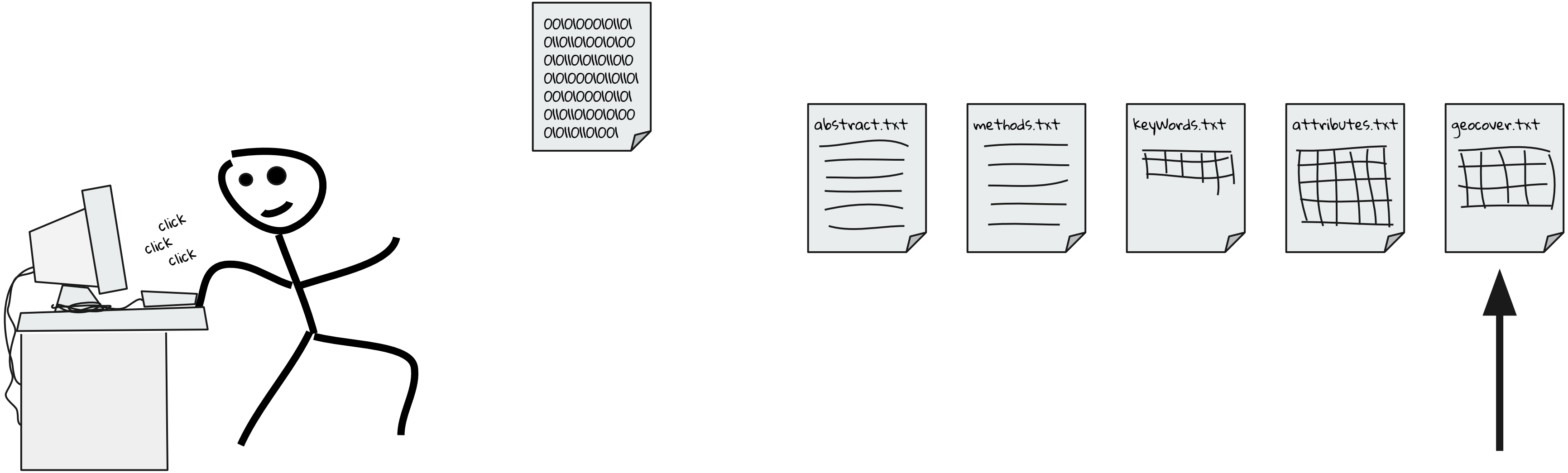
Step 5: Add custom details



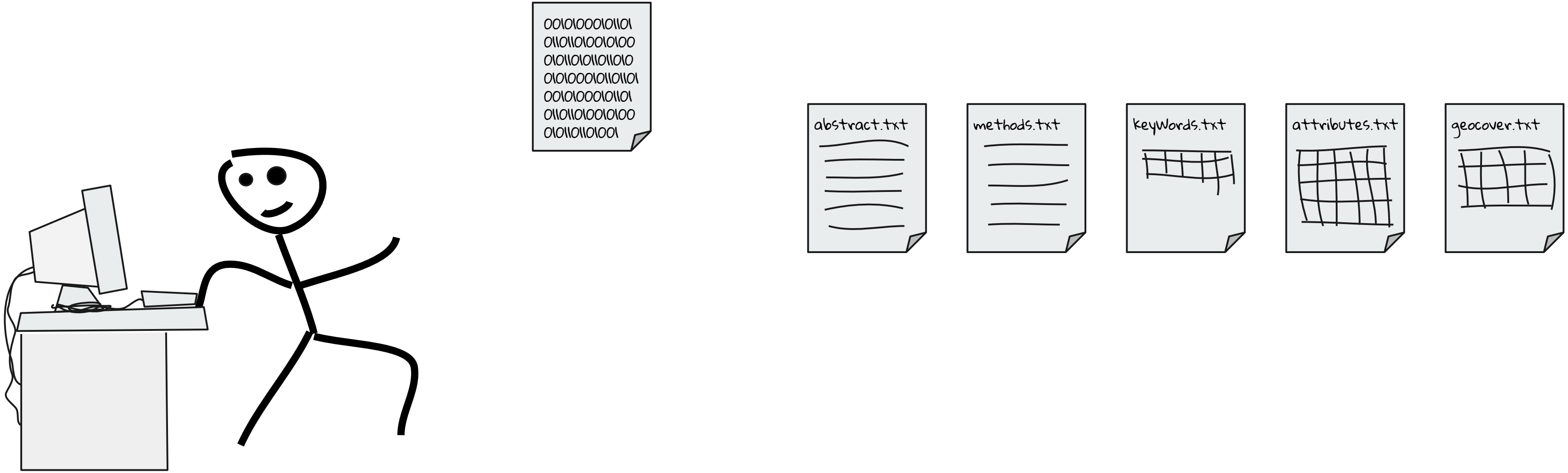
Step 5: Add custom details



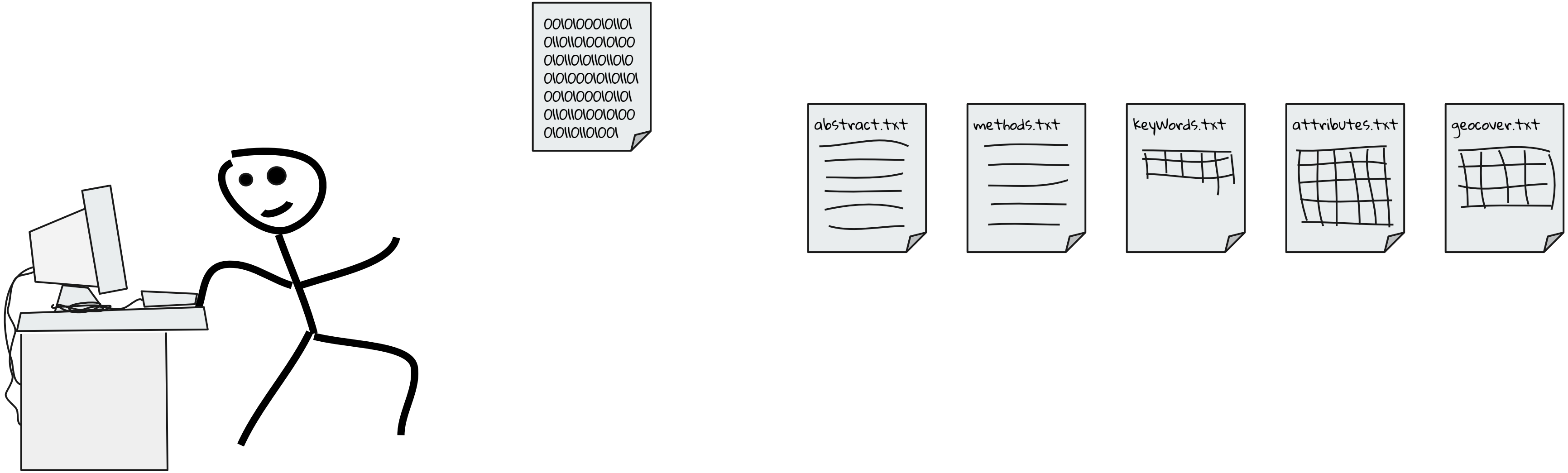
Step 5: Add custom details



Step 5: Add custom details

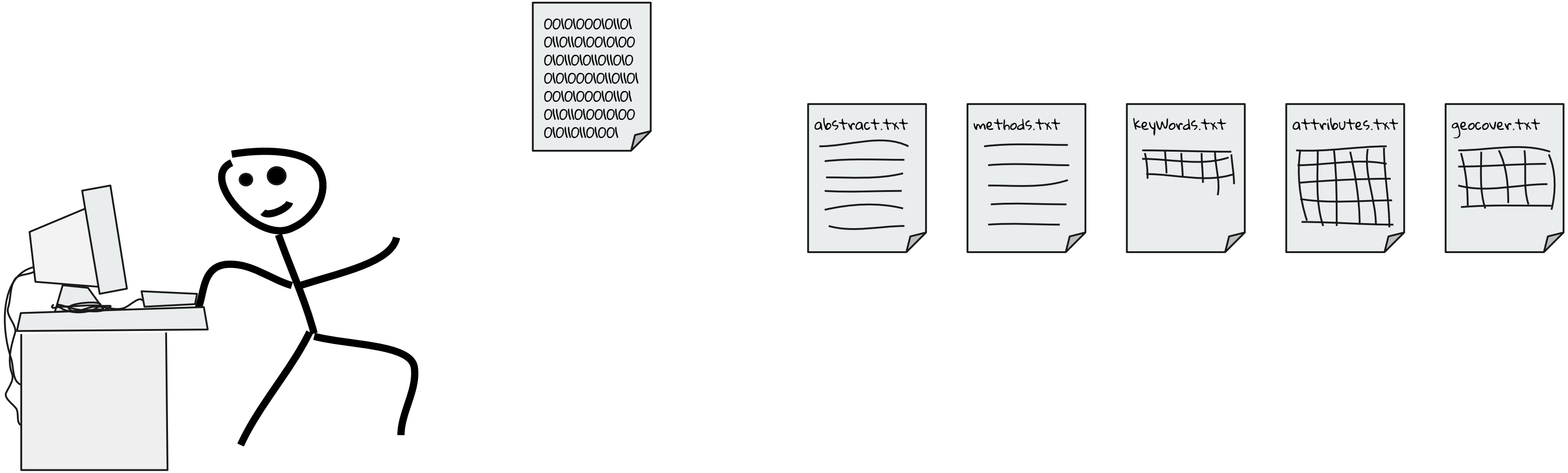


Step 6:

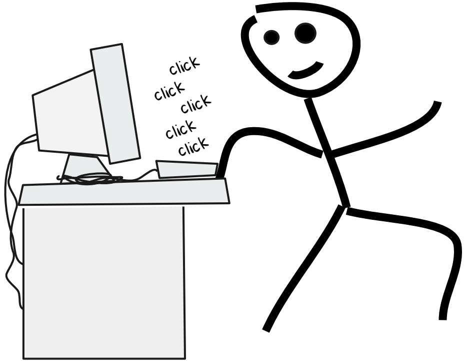


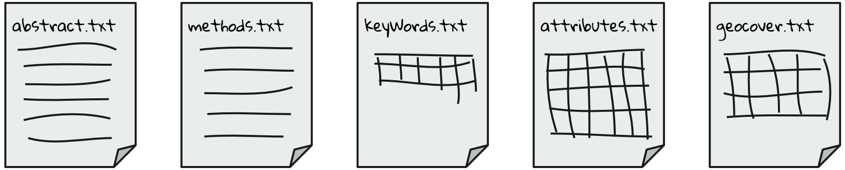
00101000101101
011011010010100
0101101011011010
0101000101101101
00101000101101
011011010010100
0101101101001
abstract.txt
methods.txt
keyWords.txt
attributes.txt
geocover.txt

Step 6: Translate to EML

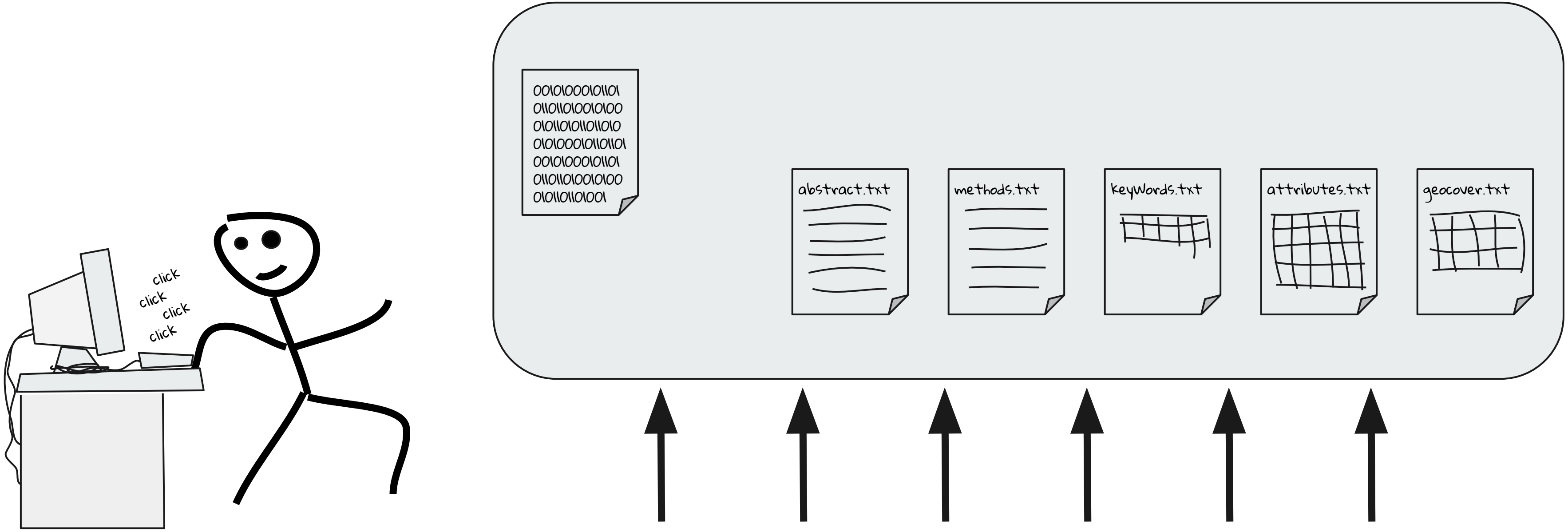


Step 6: Translate to EML

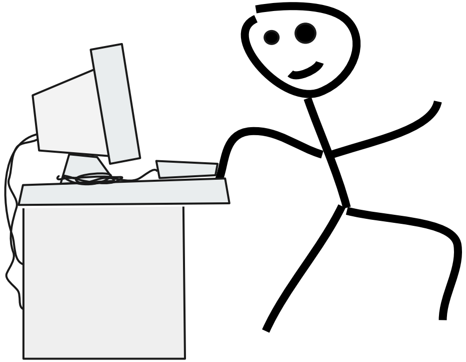


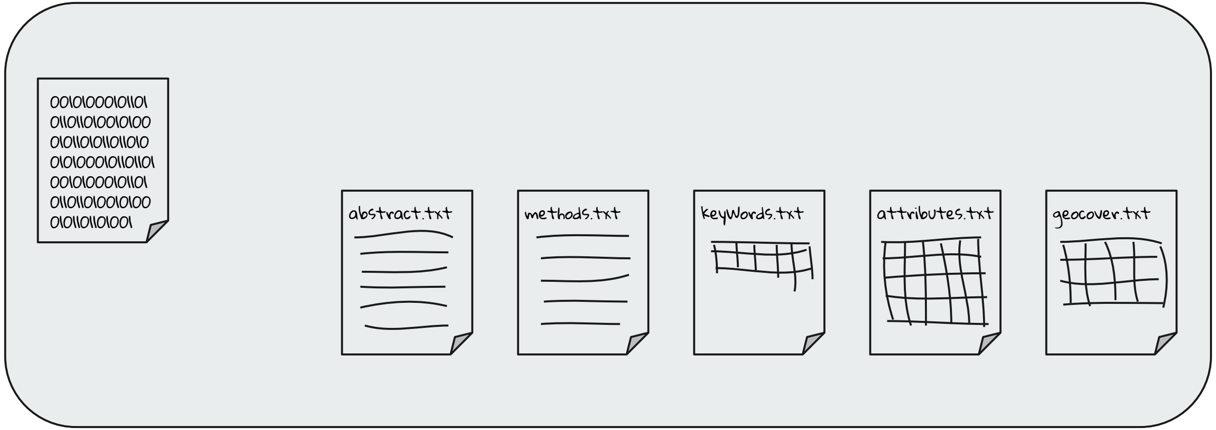


Step 6: Translate to EML

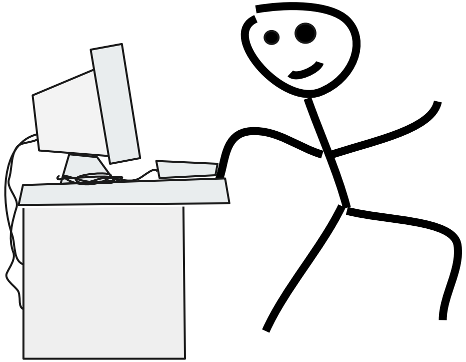


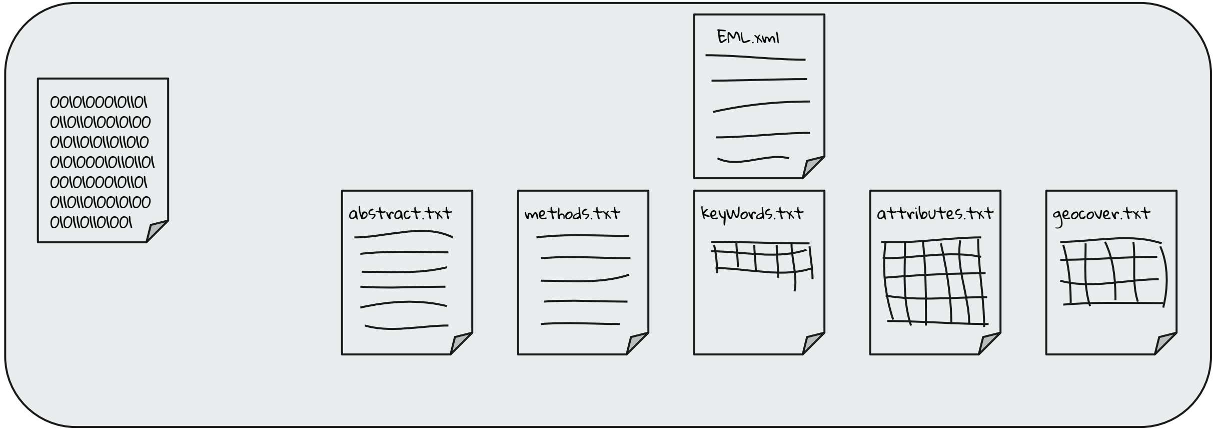
Step 6: Translate to EML



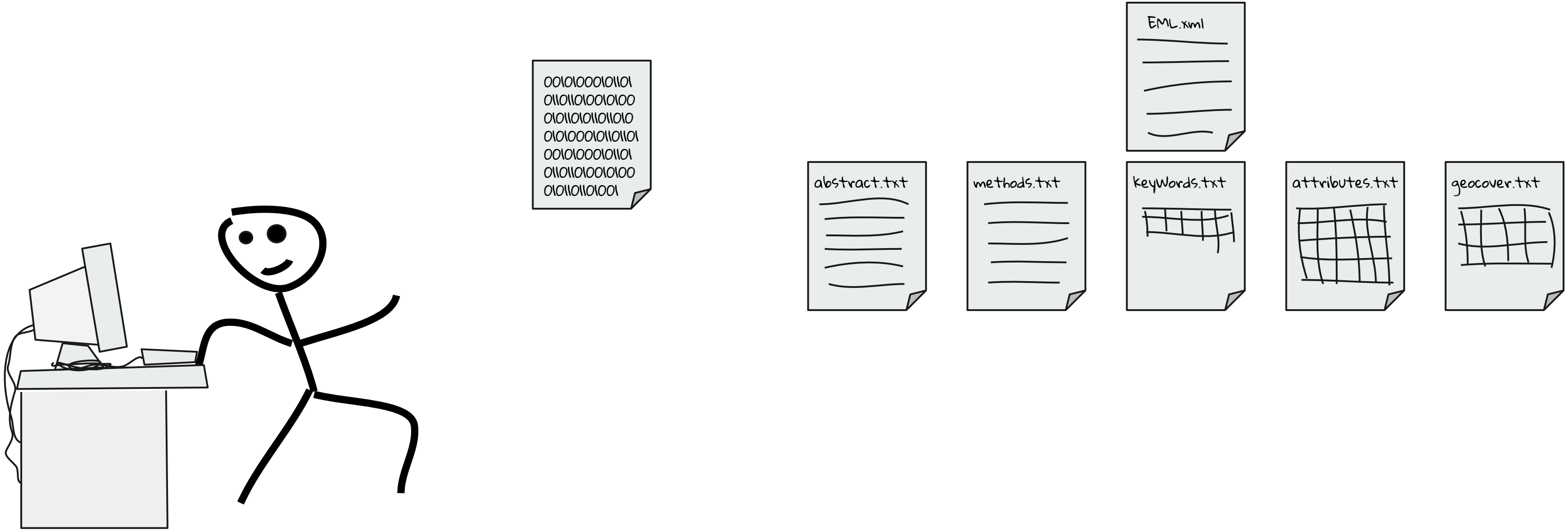


Step 6: Translate to EML

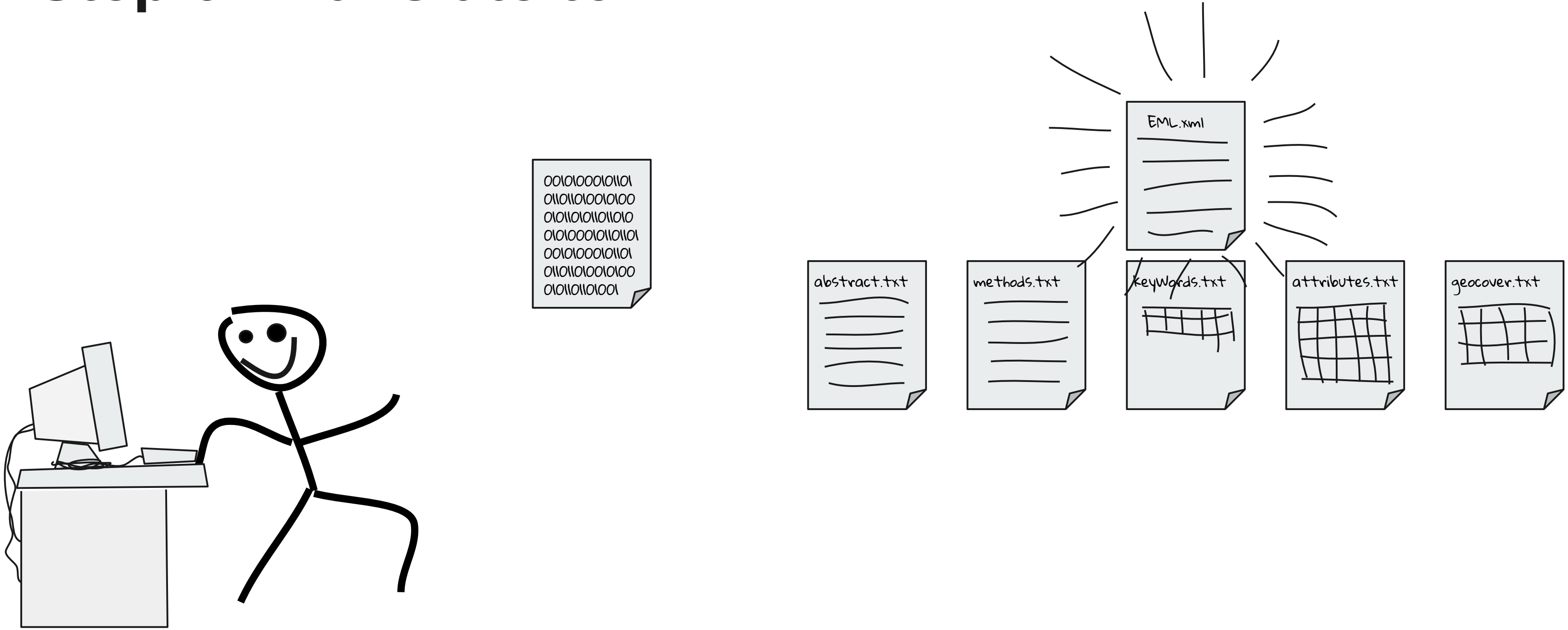




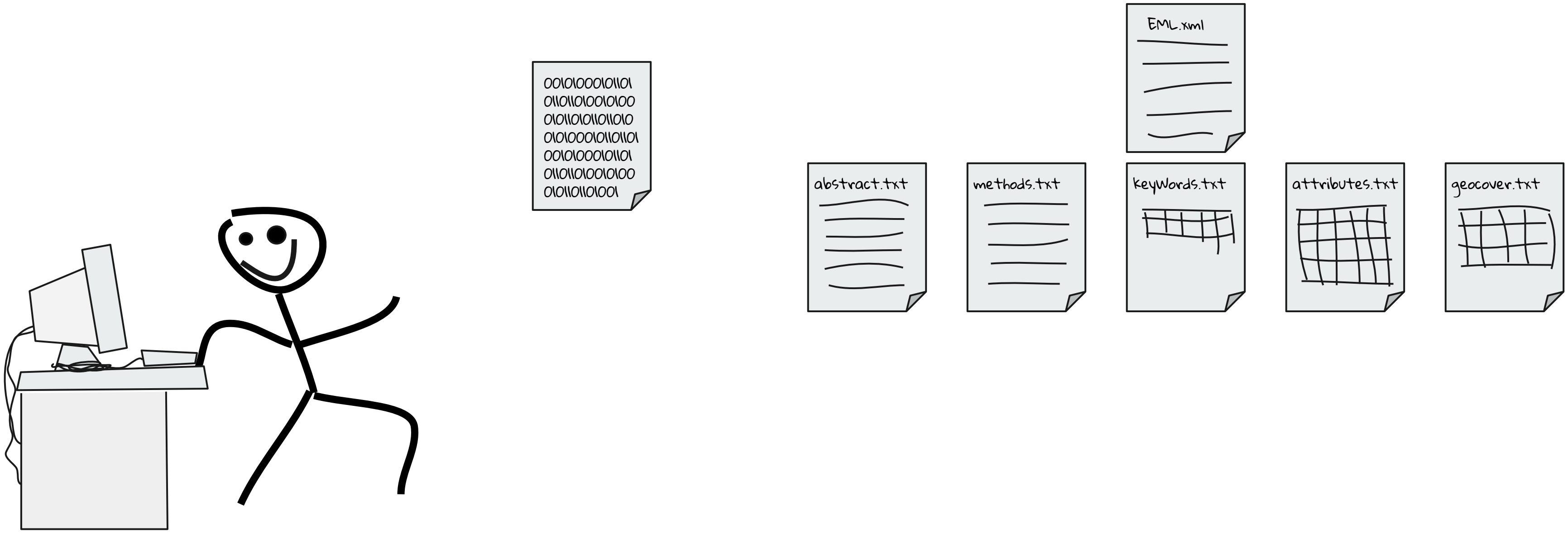
Step 6: Translate to EML



Step 6: Translate to EML



Step 6: Translate to EML

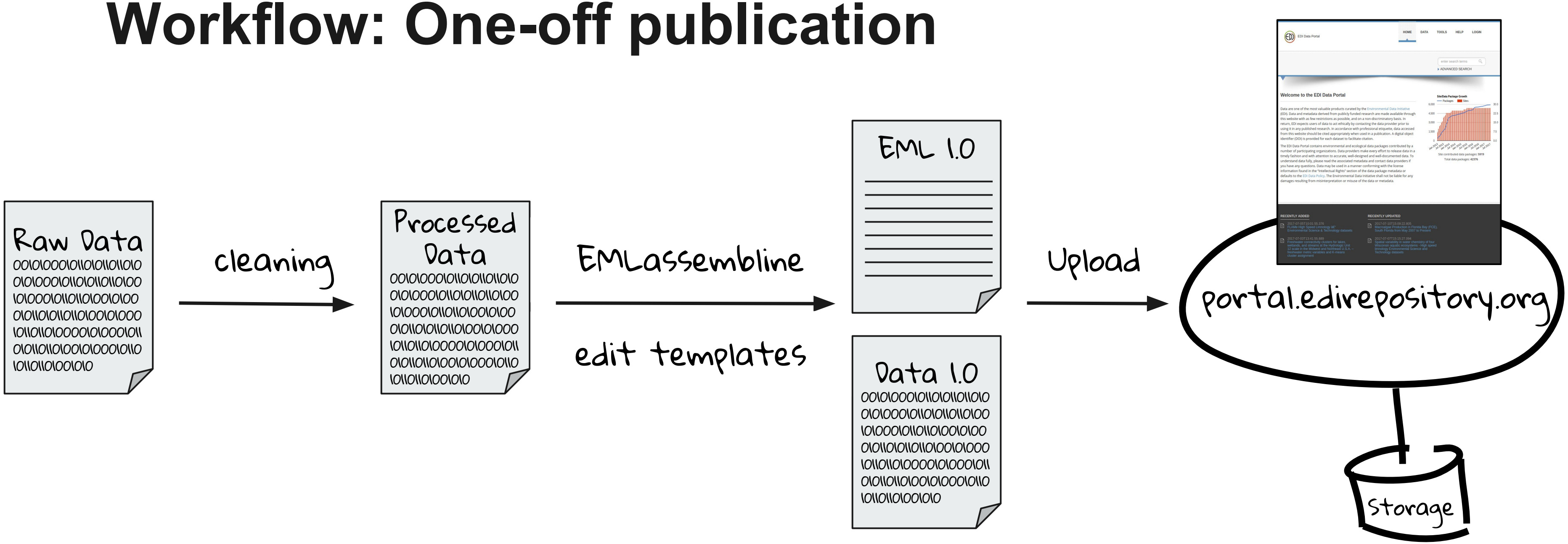


Live demo!

Workflow:

Workflow: One-off publication

Workflow: One-off publication



EDI Data Portal

HOME DATA TOOLS HELP LOGIN

enter search terms

ADVANCED SEARCH

Welcome to the EDI Data Portal

Data are one of the most valuable products curated by the Environmental Data Initiative (EDI). Data and metadata derived from publicly funded research are made available through this website with as few restrictions as possible, and on a non-discriminatory basis. In return, EDI expects users of data to act ethically by contacting the data provider prior to using it in any published research. In accordance with professional etiquette, data accessed from this website should be cited appropriately when used in a publication. A digital object identifier (DOI) is provided for each dataset to facilitate citation.

The EDI Data Portal contains environmental and ecological data packages contributed by a number of participating organizations. Data providers make every effort to release data in a timely fashion and with attention to accurate, well-designed and well-documented data. To understand data fully, please read the associated metadata and contact data providers if you have any questions. Data may be used in a manner conforming with the license information found in the "Intellectual Rights" section of the data package metadata or defaults to the EDI Data Policy. The Environmental Data Initiative shall not be liable for any damages resulting from misinterpretation or misuse of the data or metadata.

Site contributed data packages: **5919**

Total data packages: **42376**

RECENTLY ADDED

2017-07-05T10:01:55.376
FLAMe High Speed Limnology â€“ Environmental Science & Technology datasets

2017-07-03T13:41:55.889
Freshwater connectivity clusters for lakes, wetlands, and streams at the Hydrologic Unit 12 scale in the Midwest and Northeast U.S.A. – freshwater metric variables and K-means cluster assignment

RECENTLY UPDATED

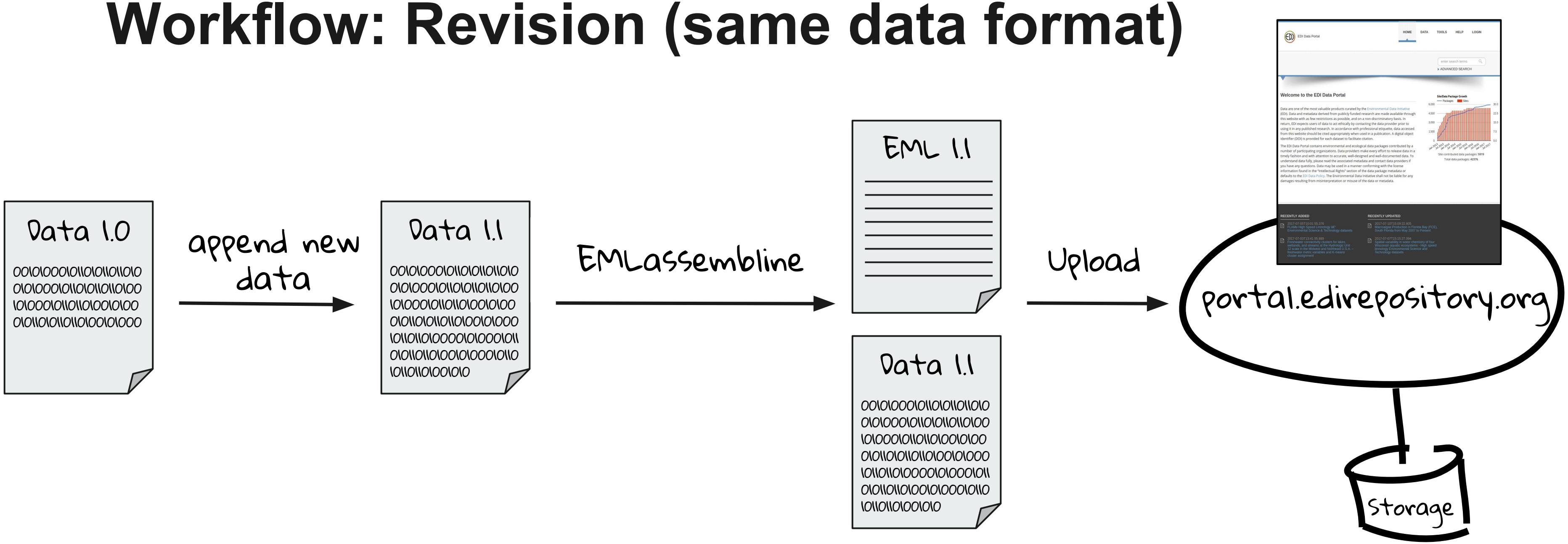
2017-07-10T15:09:22.805
Macroalgae Production in Florida Bay (FCE), South Florida from May 2007 to Present

2017-07-07T15:15:27.094
Spatial variability in water chemistry of four Wisconsin aquatic ecosystems - High speed limnology Environmental Science and Technology datasets

Workflow:

Workflow: Revision (same data format)

Workflow: Revision (same data format)



Welcome to the EDI Data Portal

Data are one of the most valuable products curated by the Environmental Data Initiative (EDI). Data and metadata derived from publicly funded research are made available through this website with as few restrictions as possible, and on a non-discriminatory basis. In return, EDI expects users of data to act ethically by contacting the data provider prior to using it in any published research. In accordance with professional etiquette, data accessed from this website should be cited appropriately when used in a publication. A digital object identifier (DOI) is provided for each dataset to facilitate citation.

The EDI Data Portal contains environmental and ecological data packages contributed by a number of participating organizations. Data providers make every effort to release data in a timely fashion and with attention to accurate, well-designed and well-documented data. To understand data fully, please read the associated metadata and contact data providers if you have any questions. Data may be used in a manner conforming with the license information found in the "Intellectual Rights" section of the data package metadata or defaults to the EDI Data Policy. The Environmental Data Initiative shall not be liable for any damages resulting from misinterpretation or misuse of the data or metadata.

Site contributed data packages: **5919**

Total data packages: **42376**

RECENTLY ADDED

2017-07-05T10:01:55.376
FLAMe High Speed Limnology â€" Environmental Science & Technology datasets

2017-07-03T13:41:55.889
Freshwater connectivity clusters for lakes, wetlands, and streams at the Hydrologic Unit 12 scale in the Midwest and Northeast U.S.A. – freshwater metric variables and K-means cluster assignment

RECENTLY UPDATED

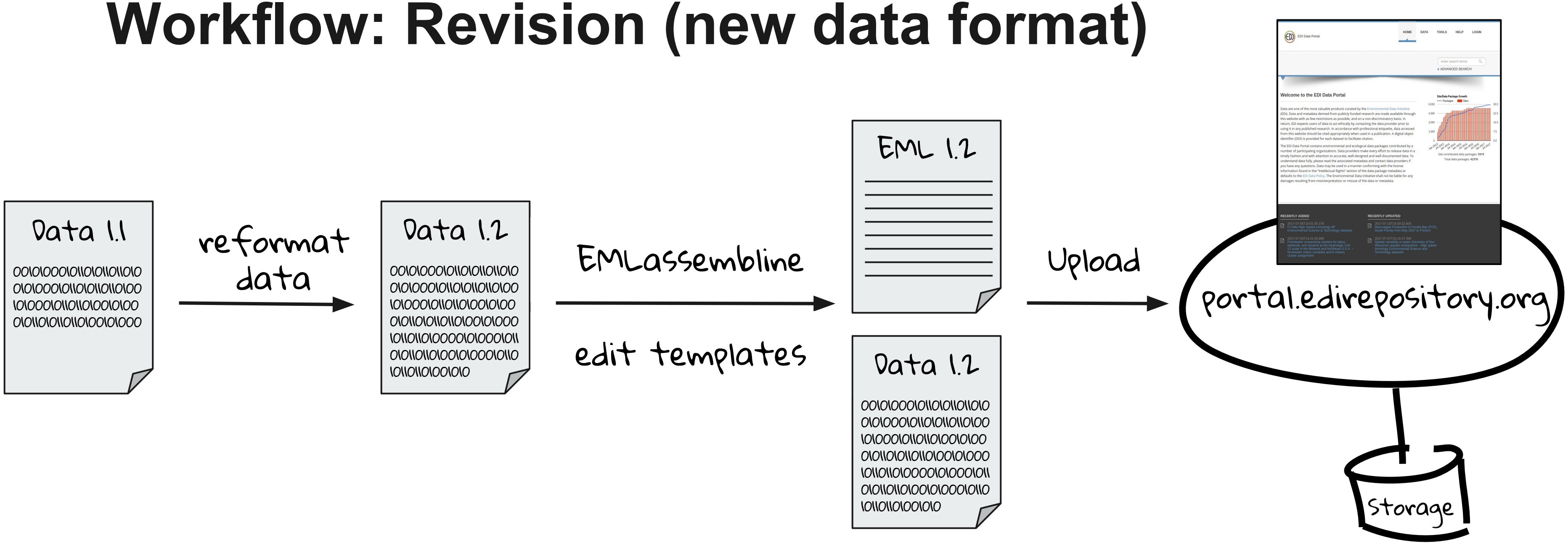
2017-07-10T15:09:22.805
Macroalgae Production in Florida Bay (FCE), South Florida from May 2007 to Present

2017-07-07T15:15:27.094
Spatial variability in water chemistry of four Wisconsin aquatic ecosystems - High speed limnology Environmental Science and Technology datasets

Workflow:

Workflow: Revision (new data format)

Workflow: Revision (new data format)



EDI Data Portal

HOME DATA TOOLS HELP LOGIN

enter search terms

ADVANCED SEARCH

Welcome to the EDI Data Portal

Data are one of the most valuable products curated by the Environmental Data Initiative (EDI). Data and metadata derived from publicly funded research are made available through this website with as few restrictions as possible, and on a non-discriminatory basis. In return, EDI expects users of data to act ethically by contacting the data provider prior to using it in any published research. In accordance with professional etiquette, data accessed from this website should be cited appropriately when used in a publication. A digital object identifier (DOI) is provided for each dataset to facilitate citation.

The EDI Data Portal contains environmental and ecological data packages contributed by a number of participating organizations. Data providers make every effort to release data in a timely fashion and with attention to accurate, well-designed and well-documented data. To understand data fully, please read the associated metadata and contact data providers if you have any questions. Data may be used in a manner conforming with the license information found in the "Intellectual Rights" section of the data package metadata or defaults to the EDI Data Policy. The Environmental Data Initiative shall not be liable for any damages resulting from misinterpretation or misuse of the data or metadata.

Site contributed data packages: 5919

Total data packages: 42376

RECENTLY ADDED

2017-07-05T10:01:55.376
FLAMe High Speed Limnology â€“ Environmental Science & Technology datasets

2017-07-03T13:41:55.889
Freshwater connectivity clusters for lakes, wetlands, and streams at the Hydrologic Unit 12 scale in the Midwest and Northeast U.S.A. – freshwater metric variables and K-means cluster assignment

RECENTLY UPDATED

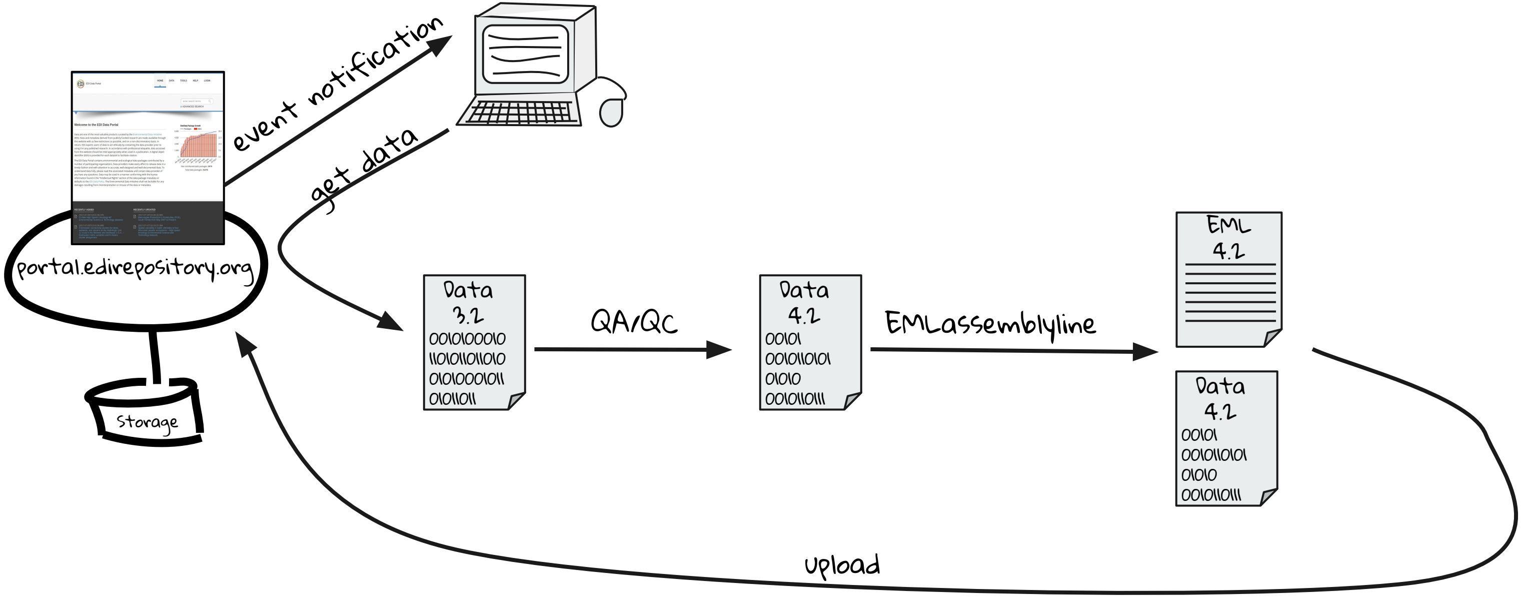
2017-07-10T15:09:22.805
Macroalgae Production in Florida Bay (FCE), South Florida from May 2007 to Present

2017-07-07T15:15:27.094
Spatial variability in water chemistry of four Wisconsin aquatic ecosystems - High speed limnology Environmental Science and Technology datasets

Workflow:

Workflow: Automation (derived data)

Workflow: Automation (derived data)



Road map

- Where is the project going?
  - High-level functions
  - Auto-extracted metadata from data entities
  - Support for more data types
- How to get involved?
  - https://github.com/EDIorg/EMLassemblyline
  - colin.smith@wisc.edu

EDI Resources

- **EDI website on “5 phases of data publishing”**
  - environmentaldatainitiative.org/resources/assemble-data-and-metadata
- **Contact EDI’s data curation team**
  - info@environmentaldatainitiative.org
- **Data Portal**
  - portal.edirepository.org/nis/home.jsp
- **GitHub**
  - github.com/EDIorg
- **Twitter**
  - @EDIgotdata
- **Slack**
  - edi-got-data.slack.com