

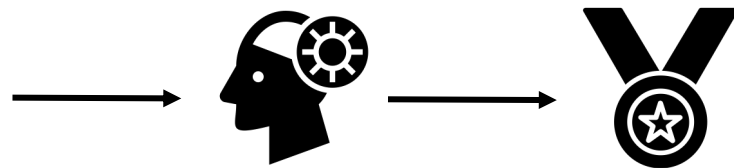
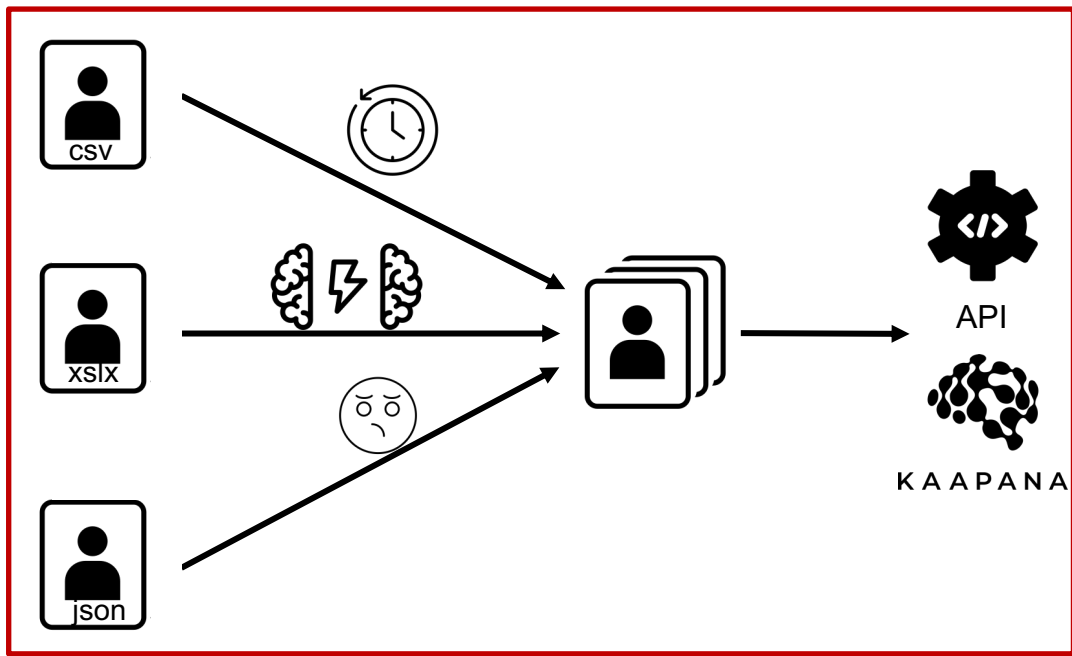
A web based tool for Import, Mapping and Harmonization of Tabular Data (applied to clinical data)

Lucas Kulla^{1 2}, Philipp Schader¹, Marco Nolden^{1 2} and Klaus Maier-Hein^{1 2}

1 Division of Medical Image Computing, German Cancer Research Center (DKFZ), Heidelberg, Germany

2 Helmholtz Metadata Collaboration (HMC), Hub Health

Motivation – the analysis pipeline



Landing page

1. Import and harmonize

[Start harmonizing your data](#)

2. Modify the data or schema

[Edit data](#)

[Edit schema](#)

3. Export the data

[Download data as .csv](#)

[Download invalid data as .csv](#)

[Open upload dialog](#)

Upload data

Star

1 Upload file

2 Select header row

3 Match Columns

4 Validate data

Upload file

☐ Reuse an existing schema

Data that we expect:
(You will have a chance to rename or remove columns in next steps)

Upload .xlsx, .xls or .csv file

Select file

Upload data – select reference schema

Star

1 Upload file

2 Select header row

3 Match Columns

4 Validate data

Upload file

☒ Reuse an existing schema

urn:kaapana:job:0.0.1

▼

Preview

Data that we expect:

(You will have a chance to rename or remove columns in next steps)

Upload .xlsx, .xls or .csv file

Select file

Upload data – preview reference schema



Upload data – file type not supported

1 Upload file

2 Select header row

3 Match Columns

4 Validate data

Upload file

☐ Reuse an existing schema

Data that we expect:
(You will have a chance to rename or remove columns in next steps)

Upload .xlsx, .xls or .csv file

Select file

README.md upload rejected
File type must be .csv, application/vnd.ms-excel,
text/csv,xls,.xlsx,.xlsx

Upload data – select sheet

Star

1 Upload file

2 Select header row

3 Match Columns

4 Validate data

Select the sheet to use

☒ uiowa_clinical_data_batch1_aug2

☐ uiowa_clinical_data_batch2_aug2

Next

Upload data – select header row

✓ Upload file

2 Select header row

3 Match Columns

4 Validate data

Select header row

☐	PatientID	Patient'...	Patient'...	Patient'...	Patient'...	Patient'...	Hispanic	History ...	History ...	History ...	Alcohol ...	Tobacc...	Details ...	T Stage	N Stage	M Stage	Tu
☒	PatientID	Patient'...	Gender	Weight	Height	Race	Spanish...	Diabetes	Previou...	Prior M...	Drinker	Smoker	Chewin...	Final T	Final N	Final M	Fir
☐	QIN-HE...	1937	M	75	173	Caucasi...	No	No	No	No	No	Former	No	4a	0	0	4a
☐	QIN-HE...	1938	M	108.5	180	Caucasi...	No	No	No	No	Social	Yes	No	X	1	0	3
☐	QIN-HE...	1937	F	84.2	151	Caucasi...	Unknown	No	No	No	Signific...	Yes	No	2	0	0	2
☐	QIN-HE...	1928	M	95.5	180	Caucasi...	Unknown	No	No	No	Signific...	Yes	No	3	0	0	3
☐	QIN-HE...	1928	M	52	167	Caucasi...	Unknown	No	No	No	Signific...	Yes	No	3	2c	0	4a
☐	QIN-HE...	1930	M	59.7	158	Caucasi...	Unknown	No	No	No	Signific...	Yes	No	4a	2b	0	4a
☐	QIN-HE...	1909	M	65.7	170	Caucasi...	Unknown	No	No	No	Social	Former	No	2	1	0	3
☐	QIN-HE...	1908	M	76.6	166	Caucasi...	Unknown	No	No	No	No	Former	No	1	2b	0	4a
☐	QIN-HE...	1925	M	63.4	173	Caucasi...	Unknown	No	No	No	Signific...	Yes	No	4	2b	0	4a
☐	QIN-HE...	1916	M	63.2	173	Caucasi...	Unknown	No	Yes	Prior H...	No	Former	No	2	0	0	2
☐	QIN-HE...	1926	M	72.7	180	Caucasi...	Unknown	Diet-Co...	No	No	Signific...	Yes	No	3	2c	0	4a
☐	QIN-HE...	1942	M	96.2	175	Caucasi...	No	No	No	No	Signific...	Yes	No	2	2a	0	4a
☐	QIN-HE...	1931	M	92.6	168	Caucasi...	No	Non Ins...	No	No	No	Yes	No	2	2b	0	4a

Next

Data harmonization

✓ Upload file

✓ Select header row

3 Match Columns

4 Validate data

Match Columns

Your table

PatientID	Patient's Birth Date	Gender	Weight	Height
QIN-HEADNECK-01-0003	1937	M	75	173
QIN-HEADNECK-01-0004	1938	M	108.5	180

Will become

PatientID

✓

✎

Patient's Bir...

✓

✎

Gender

✓

✎

Weight

✓

✎

Height

✓

✎

PatientID

id

Patient's Birth Date

Gender

Weight

Height

Race

Spanish Origin

Next

Data harmonization – add a Field

Edit or Add a Item Field

label*

key*

description

alternateMatchesArray

validations
Add Validation

fieldType*

example

Edit or Add a Item Field

key*

description

alternateMatchesArray

validations

Rule

Error Message

Level

Allow Empty ☐

Add Validation

fieldType*

example

Validate harmonized dataset

Star

✓ Upload file

✓ Select header row

✓ Match Columns

4 Validate data

Validate data

Discard selected rows

Show only rows with errors

	PATIENTID	ID	PATIENT'S BIR...	GENDER	WEIGHT	HEIGHT	RACE	SPANISH ORIGIN	DIABETES
☐	QIN-HEADNECK-0...		1937	M	75	173	Caucasian	No	No
☐	QIN-HEADNECK-0...		1938	M	108.5	180	Caucasian	No	No
☐	QIN-HEADNECK-0...		1937	F	84.2	151	Caucasian	Unknown	No
☐	QIN-HEADNECK-0...		1928	M	95.5	180	Caucasian	Unknown	No
☐	QIN-HEADNECK-0...		1928	M	52	167	Caucasian	Unknown	No
☐	QIN-HEADNECK-0...		1930	M	59.7	158	Caucasian	Unknown	No
☐	QIN-HEADNECK-0...		1909	M	65.7	170	Caucasian	Unknown	No
☐	QIN-HEADNECK-0...		1908	M	76.6	166	Caucasian	Unknown	No
☐	QIN-HEADNECK-0...		1925	M	63.4	173	Caucasian	Unknown	No
☐	QIN-HEADNECK-0...		1916	M	63.2	173	Caucasian	Unknown	No
☐	QIN-HEADNECK-0...		1926	M	72.7	180	Caucasian	Unknown	Diet-Controllec
☐	QIN-HEADNECK-0...		1942	M	96.2	175	Caucasian	No	No
☐	QIN-HEADNECK-0...		1931	M	92.6	168	Caucasian	No	Non Insulin-De
☐	QIN-HEADNECK-0...		1927	M	68.4	174	Caucasian	Unknown	No

Confirm

Landing page

1. Import and harmonize

[Start harmonizing your data](#)

2. Modify the data or schema

[Edit data](#)

[Edit schema](#)

3. Export the data

[Download data as .csv](#)

[Download invalid data as .csv](#)

[Open upload dialog](#)

Data editor – complex harmonization

Editor

Select theme

Save

```
1 {
2   {
3     "PatientID": "QIN-HEADNECK-01-0003",
4     "Patient's Birth Date": "1937",
5     "Gender": "M",
6     "Weight": "75",
7     "Height": "173",
8     "Race": "Caucasian",
9     "Spanish Origin": "No",
10    "Diabetes": "No",
11    "Previous Radiation": "No",
12    "Prior Malignancies": "No",
13    "Drinker": "No",
14    "Smoker": "Former",
15    "Chewing Tobacco": "No",
16    "Final T": "4a",
17    "Final N": "0",
18    "Final M": "0",
19    "Final Stage": "4a",
20    "Final Site": "Tonsil",
21    "Biopsy1 Date": "1986-03-03",
22    "Biopsy1 Location": "R Tonsil",
23    "Biopsy2 Date": "1986-03-17",
24    "Biopsy2 Location": "R Tonsil",
25    "Biopsy3 Date": null,
26    "Biopsy3 Location": null,
27    "Biopsy4 Date": null,
28    "Biopsy4 Location": null,
29    "Surgery1 Date": null,
30    "Surgery1 Desc": null,
31    "Surgery1 Primary Resected": null,
32    "Surgery1 Node Dissection": null,
33    "Surgery2 Date": null,
34    "Surgery2 Desc": null,
35    "Surgery2 Primary Resected": null,
36    "Surgery2 Node Dissection": null,
37    "Rt1 Start Date": "1986-04-11",
38    "Rt1 End Date": "1986-06-03",
39    "Rt1": "70",
```

Execute

```
1 function foot(){
2   let dataLeftSide = JSON.parse(data)
3
4   for (let obj of dataLeftSide) {
5     for (let key in obj) {
6       if (obj[key] === "No") {
7         obj[key] = 0; // convert "No" to
8       } else if (obj[key] === "Yes") {
9         obj[key] = 1; // convert "Yes" to
10      }
11    }
12  }
13
14  return JSON.stringify(dataLeftSide, null, 4)
15 }
```

```
1 {
2   {
3     "PatientID": "QIN-HEADNECK-01-0003",
4     "Patient's Birth Date": "1937",
5     "Gender": "M",
6     "Weight": "75",
7     "Height": "173",
8     "Race": "Caucasian",
9     "Spanish Origin": 0,
10    "Diabetes": 0,
11    "Previous Radiation": 0,
12    "Prior Malignancies": 0,
13    "Drinker": 0,
14    "Smoker": "Former",
15    "Chewing Tobacco": 0,
16    "Final T": "4a",
17    "Final N": "0",
18    "Final M": "0",
19    "Final Stage": "4a",
20    "Final Site": "Tonsil",
21    "Biopsy1 Date": "1986-03-03",
22    "Biopsy1 Location": "R Tonsil",
23    "Biopsy2 Date": "1986-03-17",
24    "Biopsy2 Location": "R Tonsil",
25    "Biopsy3 Date": null,
26    "Biopsy3 Location": null,
27    "Biopsy4 Date": null,
28    "Biopsy4 Location": null,
29    "Surgery1 Date": null,
30    "Surgery1 Desc": null,
31    "Surgery1 Primary Resected": null,
32    "Surgery1 Node Dissection": null,
33    "Surgery2 Date": null,
34    "Surgery2 Desc": null,
35    "Surgery2 Primary Resected": null,
36    "Surgery2 Node Dissection": null,
37    "Rt1 Start Date": "1986-04-11",
38    "Rt1 End Date": "1986-06-03",
39    "Rt1": "70",
```

Schema editor

Schema editor ×

light ▼

Save

```
1 {
2   "$defs": {
3     "Rt": {
4       "type": "object",
5       "properties": {
6         "1 Notes": {
7           "type": "string",
8           "description": "This field element is automatically generated",
9           "x-alternateMatches": [
10            "Rt.1 Notes"
11          ]
12        },
13        "2 Notes": {
14          "type": "string",
15          "description": "This field element is automatically generated",
16          "x-alternateMatches": [
17            "Rt.2 Notes"
18          ]
19        },
20        "3 Notes": {
21          "type": "string",
22          "description": "This field element is automatically generated",
23          "x-alternateMatches": [
24            "Rt.3 Notes"
25          ]
26        }
27      }
28    },
29    "properties": {
30      "PatientID": {
31        "type": "string",
32        "description": "This field element is automatically generated",
33        "x-alternateMatches": [
34          "PatientID"
35        ]
36      },
37      "id": {
38        "type": "string",
39        "description": "This field element is automatically generated",
40        "x-alternateMatches": [
41          "id"
42        ]
43      }
44    }
45  }
46 }
```

Upload data / schema to an API

Upload your data and schema

☒ Upload Data

☒ Upload Schema

urn:HMCConference:Data

Valid Schema Name
Example: urn:namespace:schemaName

Submit Cancel

Edit data

Edit schema

3. Export the data

Download data as .csv Download invalid data as .csv Open upload dialog

A web based tool for Import, Mapping and Harmonization of Tabular Data (applied to clinical data)

Lucas Kulla^{1 2}, Philipp Schader¹, Marco Nolden^{1 2} and Klaus Maier-Hein^{1 2}

1 Division of Medical Image Computing, German Cancer Research Center (DKFZ), Heidelberg, Germany

2 Helmholtz Metadata Collaboration (HMC), Hub Health