HELMHOLTZ Metadata Collaboration

dkfz. GERMAN CANCER RESEARCH CENTER IN THE HELMHOLTZ ASSOCIATION
Research for a Life without Cancer

AST | E&E | ENERGY | **HEALTH** | INFORMATION | MATTER | DATA COMMONS | HMC OFFICE | HMC PROJECTS

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

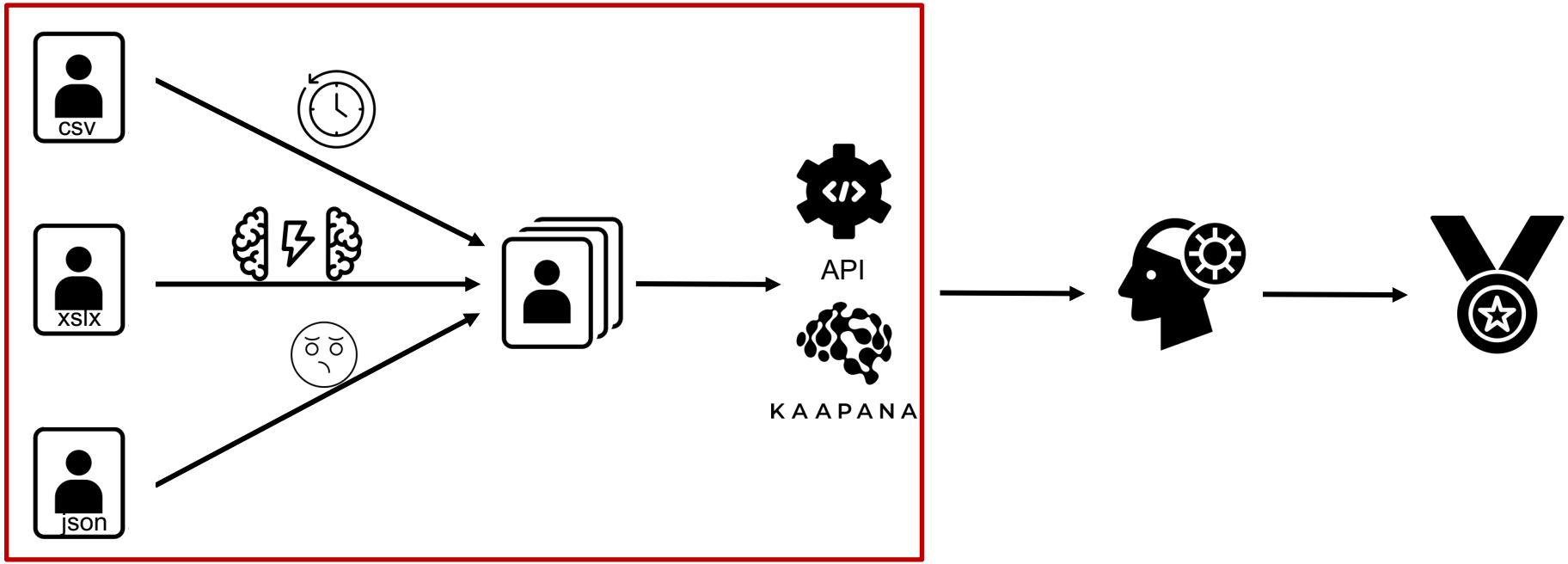
Lucas Kulla[1 2], Philipp Schader[1], Marco Nolden[1 2] and Klaus Maier-Hein[1 2]

1 Division of Medical Image Computing, German Cancer Research Center (DKFZ), Heidelberg, Germany
2 Helmholtz Metadata Collaboration (HMC), Hub Health

www.helmholtz-metadaten.de

Motivation – the analysis pipeline

CANCER IMAGING ARCHIVE



Landing page

1. Import and harmonize

Start harmonizing your data

2. Modify the data or schema

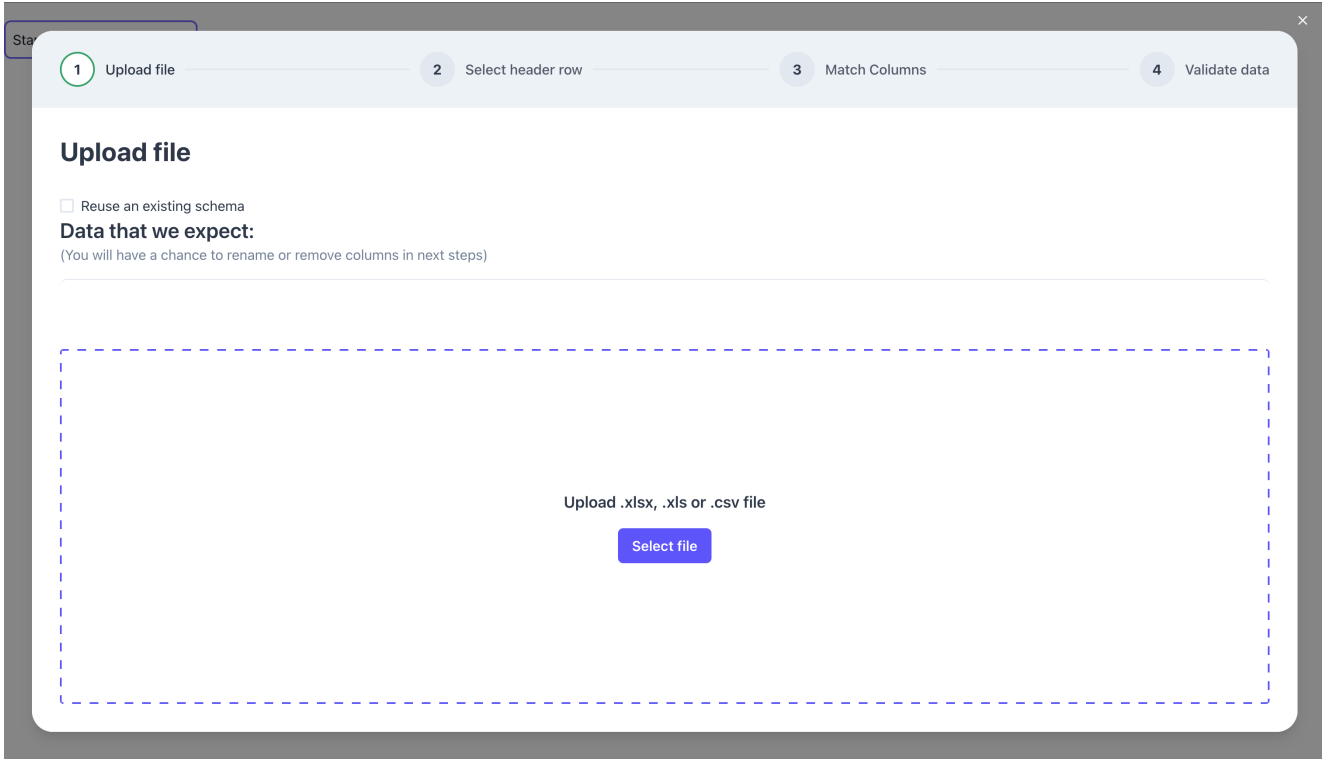
Edit data

Edit schema

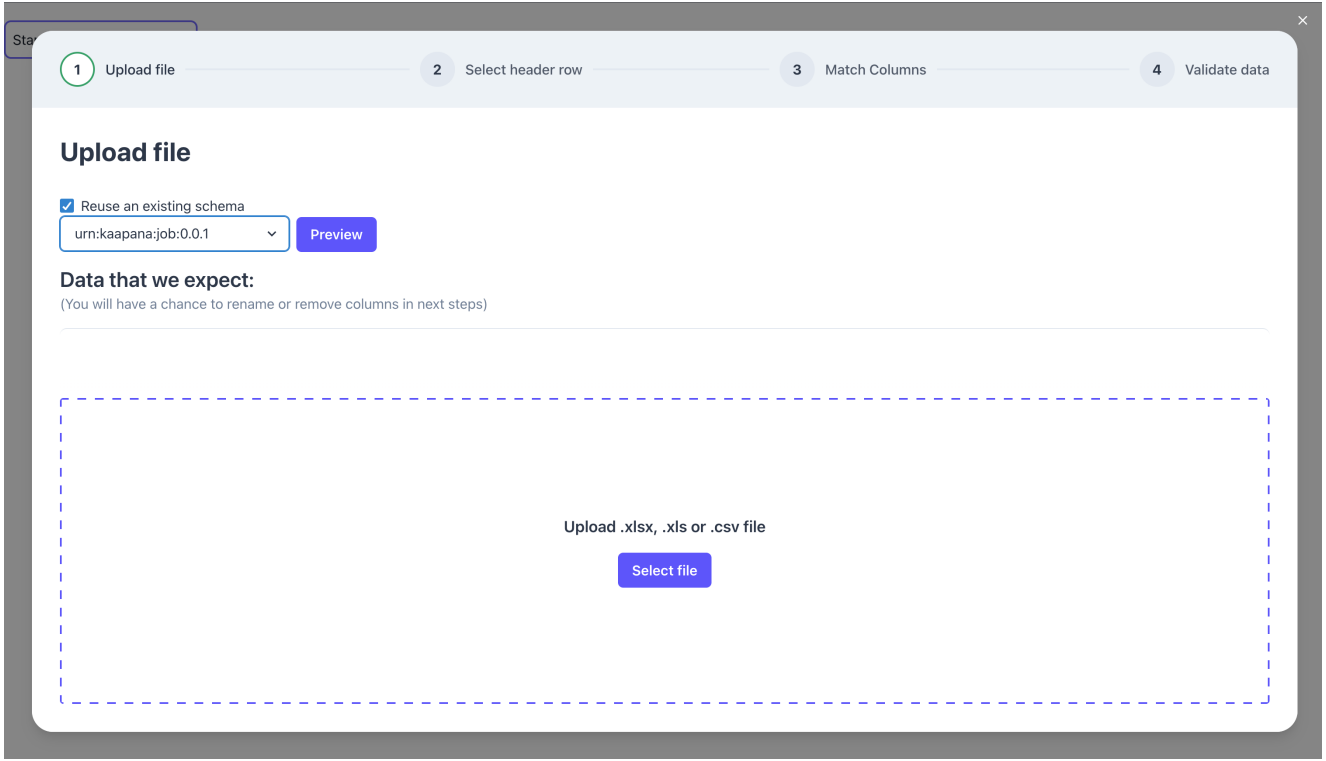
3. Export the data

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

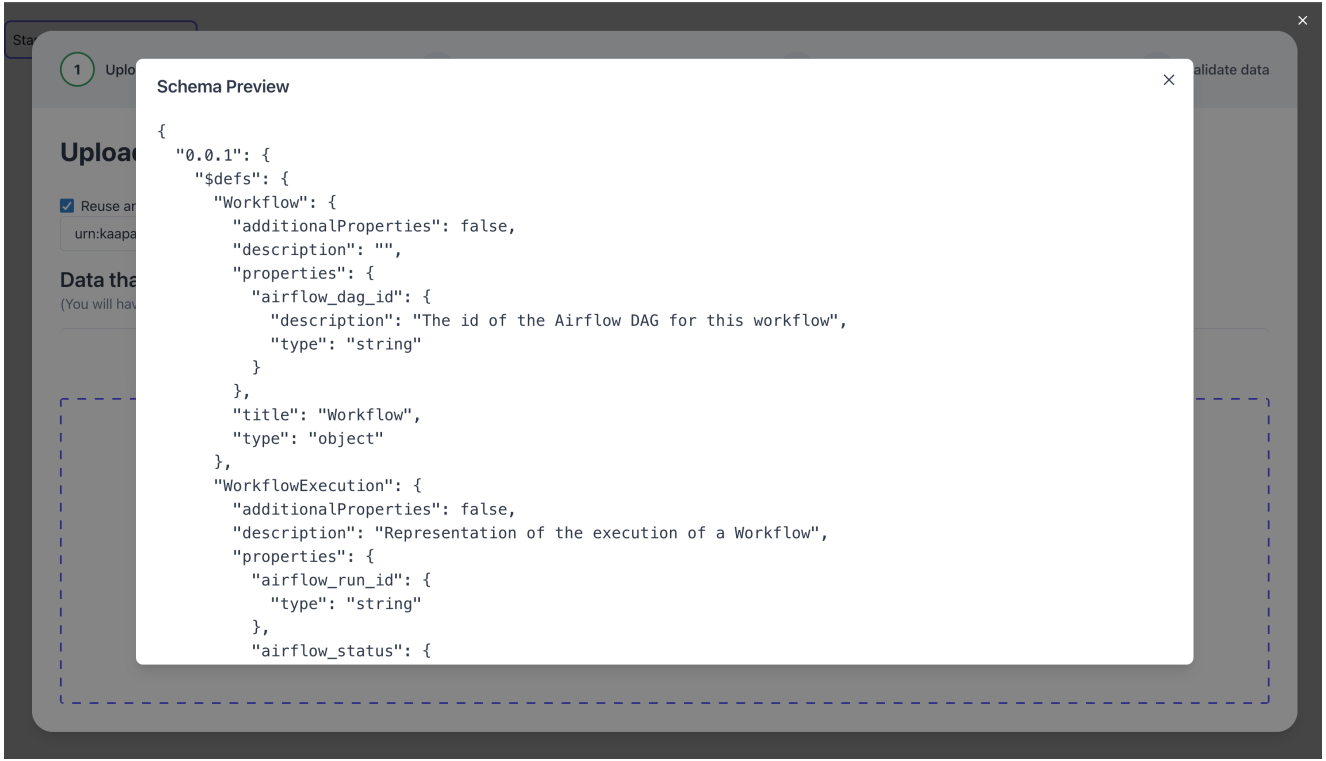
Upload data


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

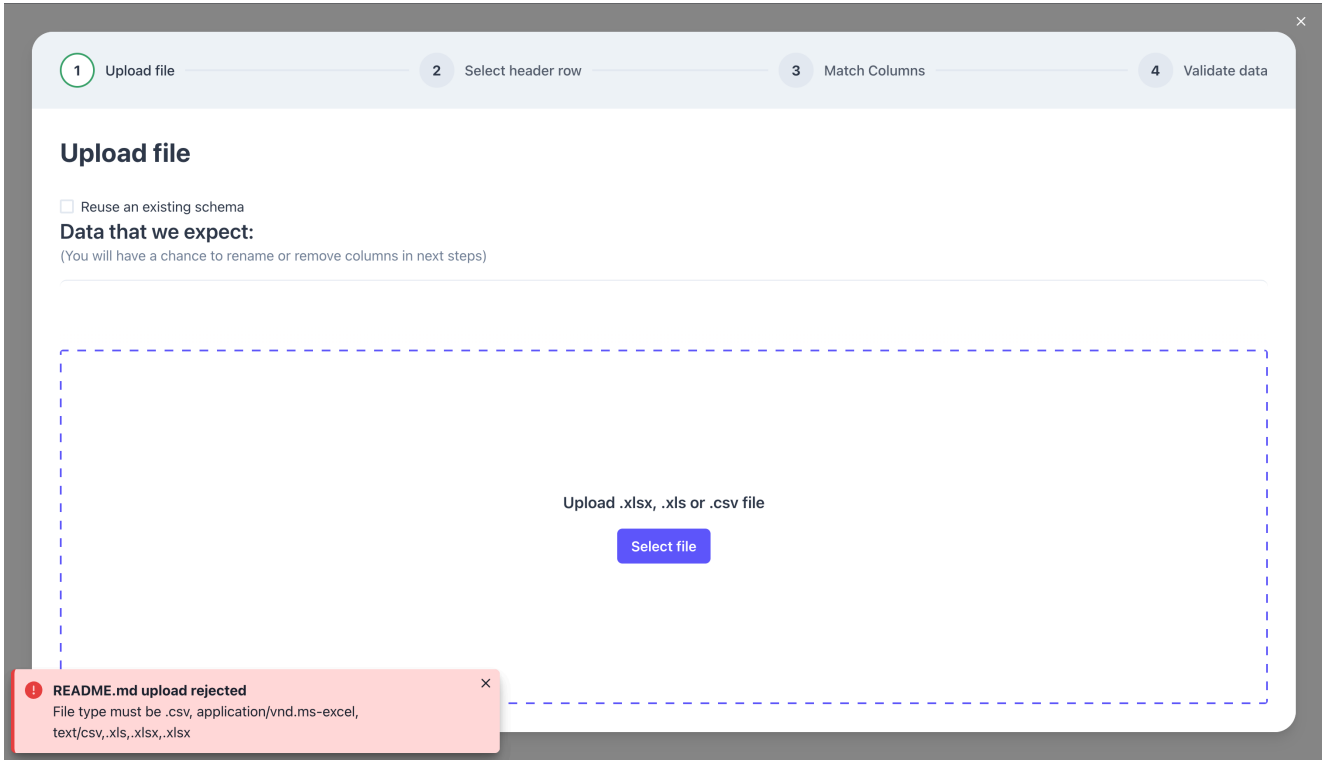
Upload data – select reference schema



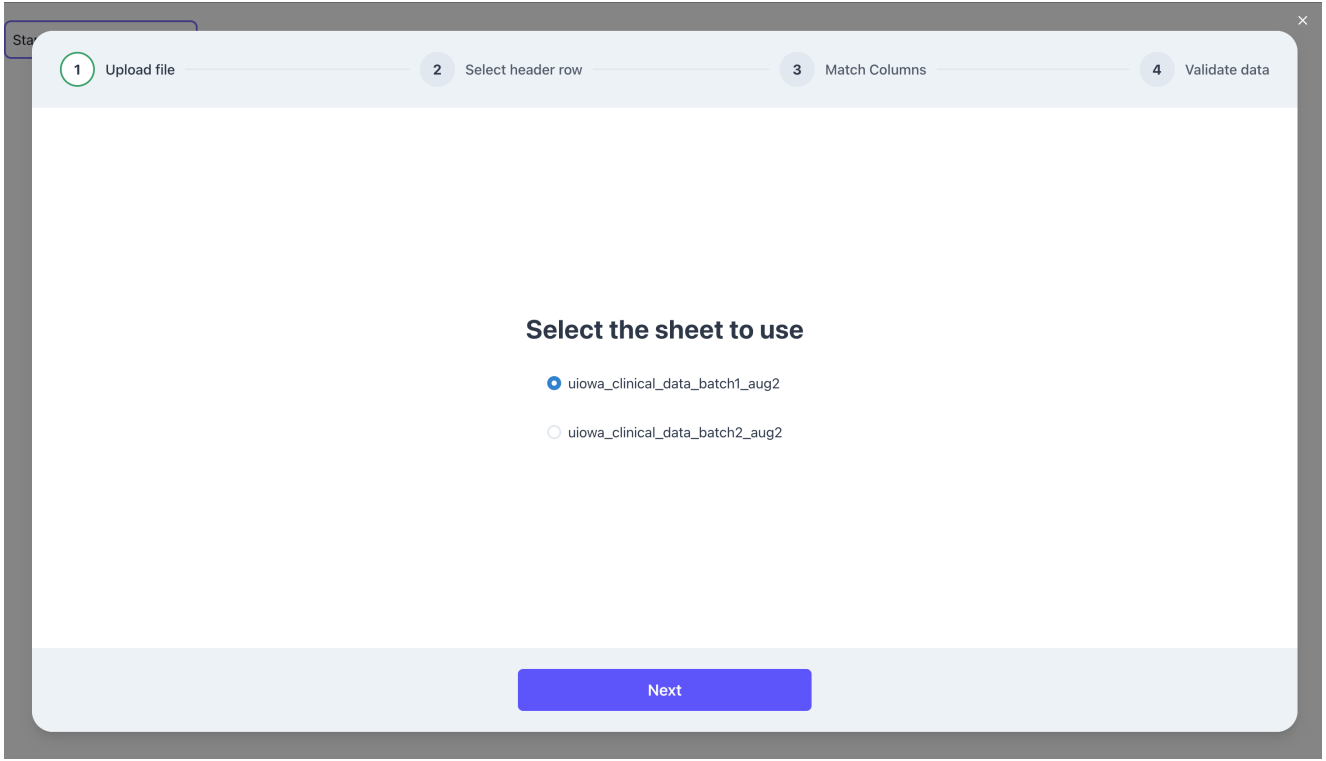
Upload data – preview reference schema



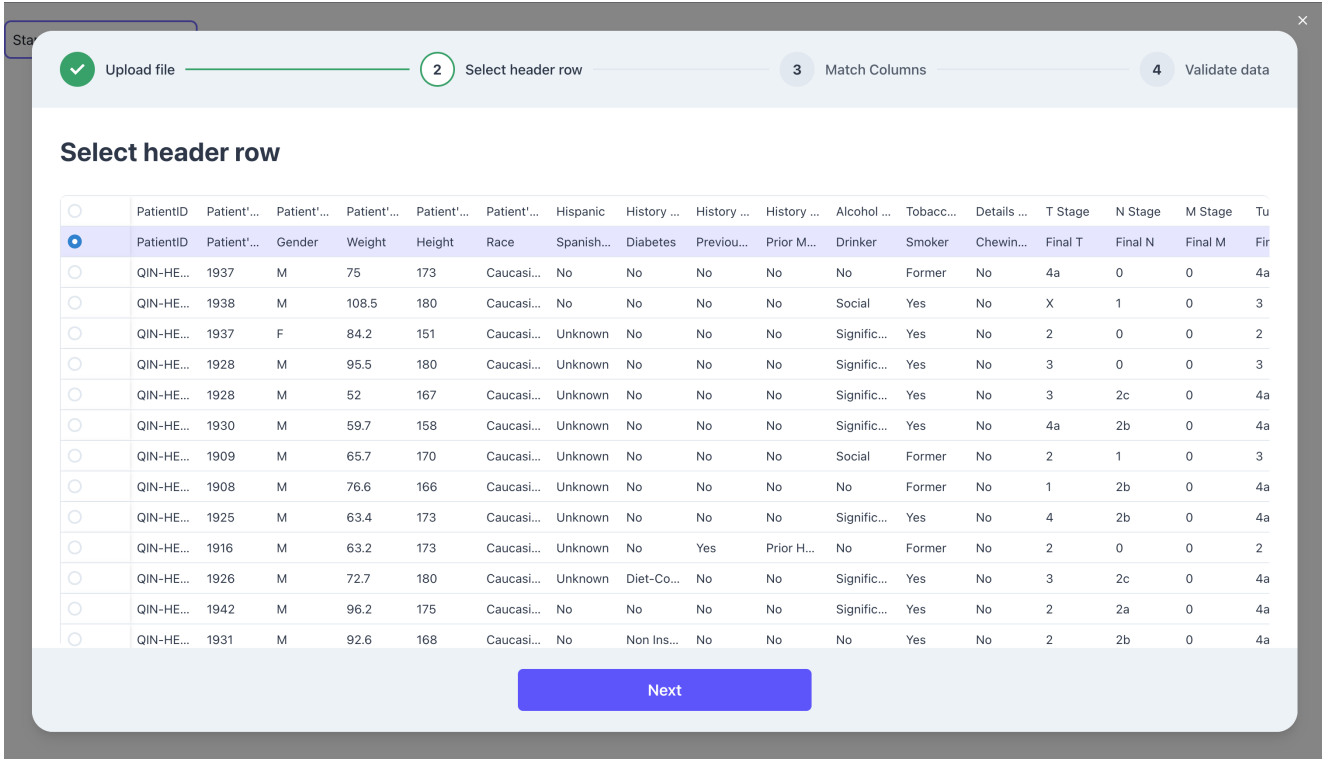
Upload data – file type not supported



Upload data – select sheet

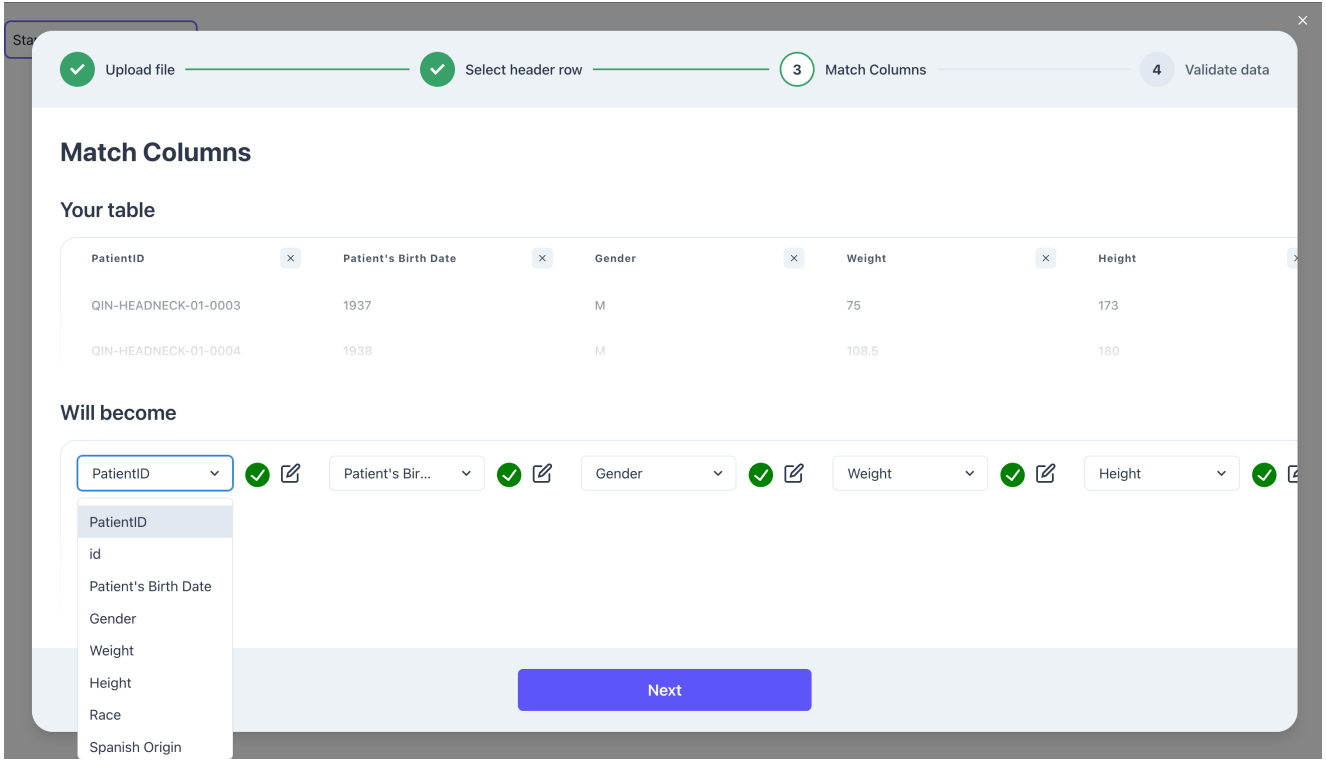


Upload data – select header row

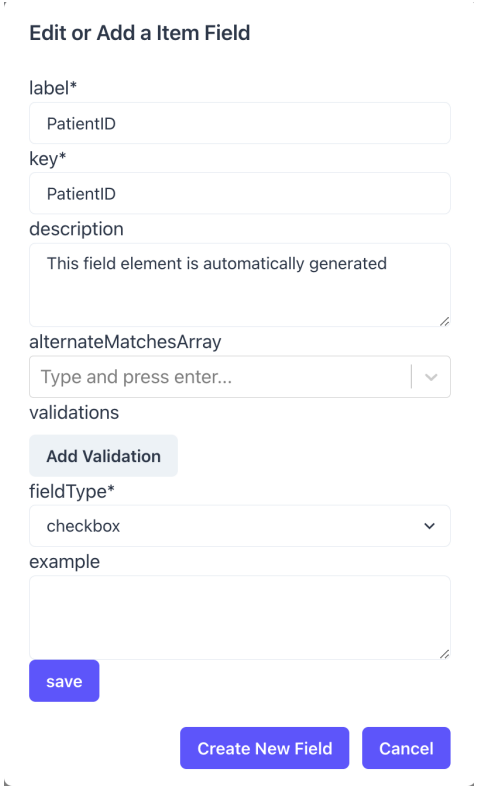


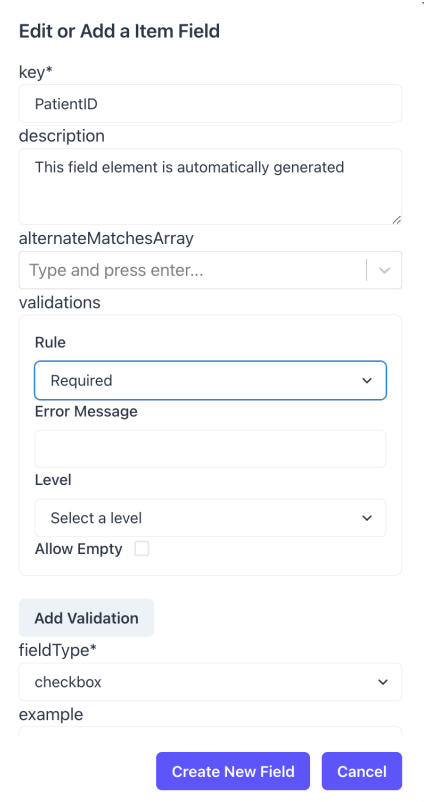
	PatientID	Patient'...	Patient'...	Patient'...	Patient'...	Patient'...	Hispanic	History ...	History ...	History ...	Alcohol ...	Tobacc...	Details ...	T Stage	N Stage	M Stage	Tu
◉	PatientID	Patient'...	Gender	Weight	Height	Race	Spanish...	Diabetes	Previou...	Prior M...	Drinker	Smoker	Chewin...	Final T	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

Data harmonization

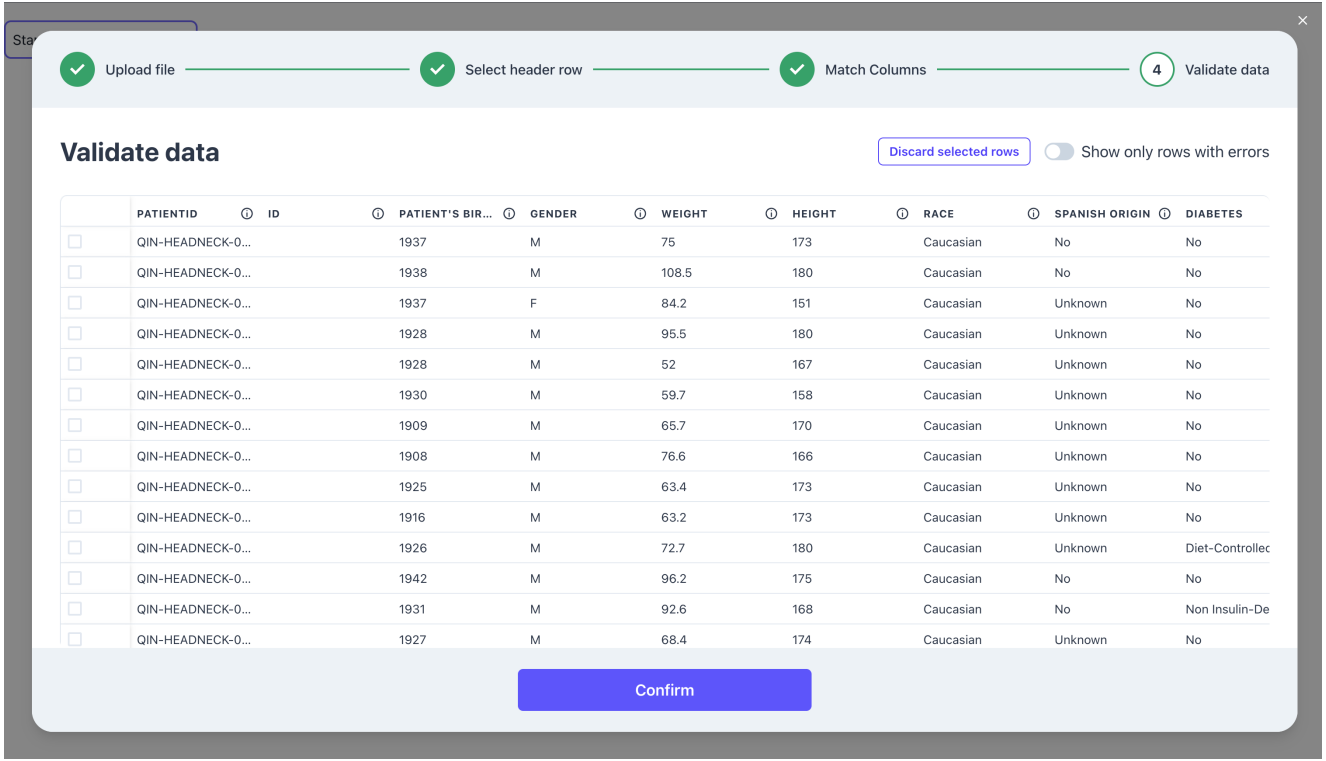


Data harmonization – add a Field





Validate harmonized dataset



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

Landing page

1. Import and harmonize

Start harmonizing your data

2. Modify the data or schema

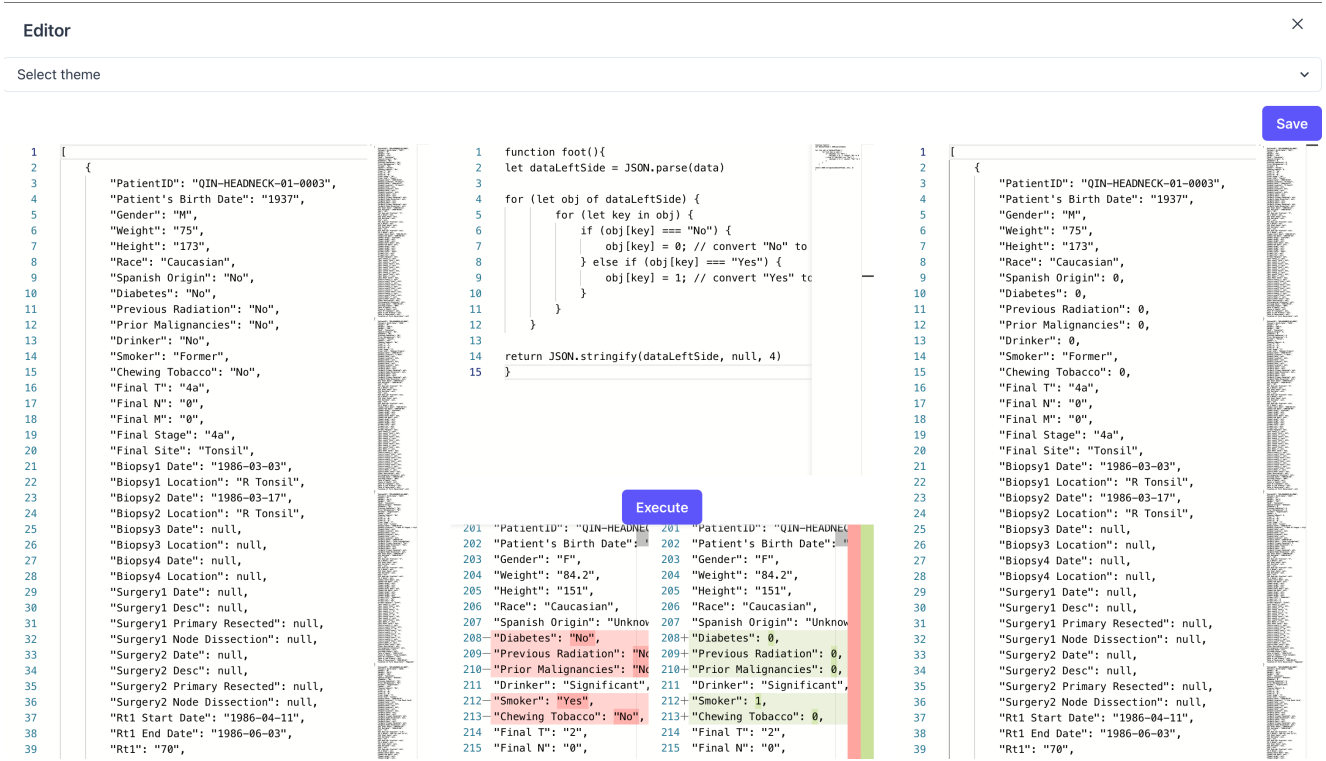
Edit data

Edit schema

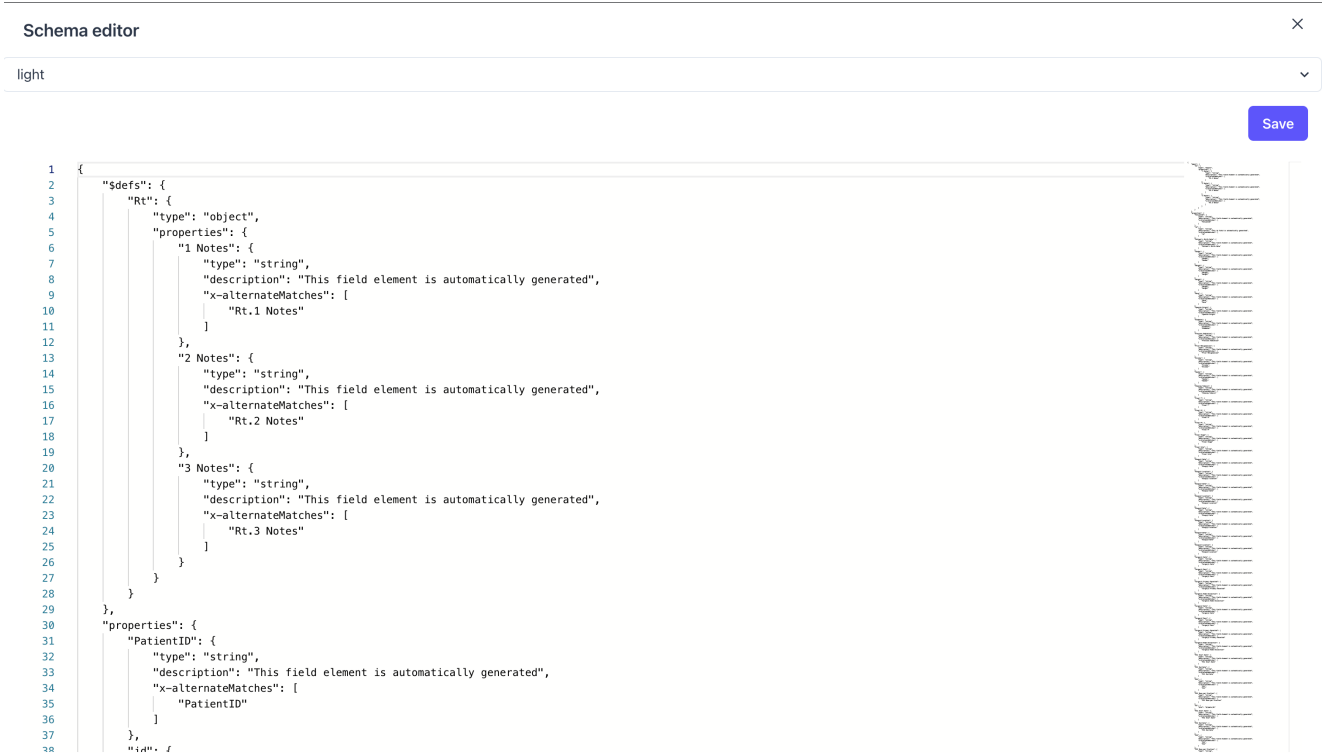
3. Export the data

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

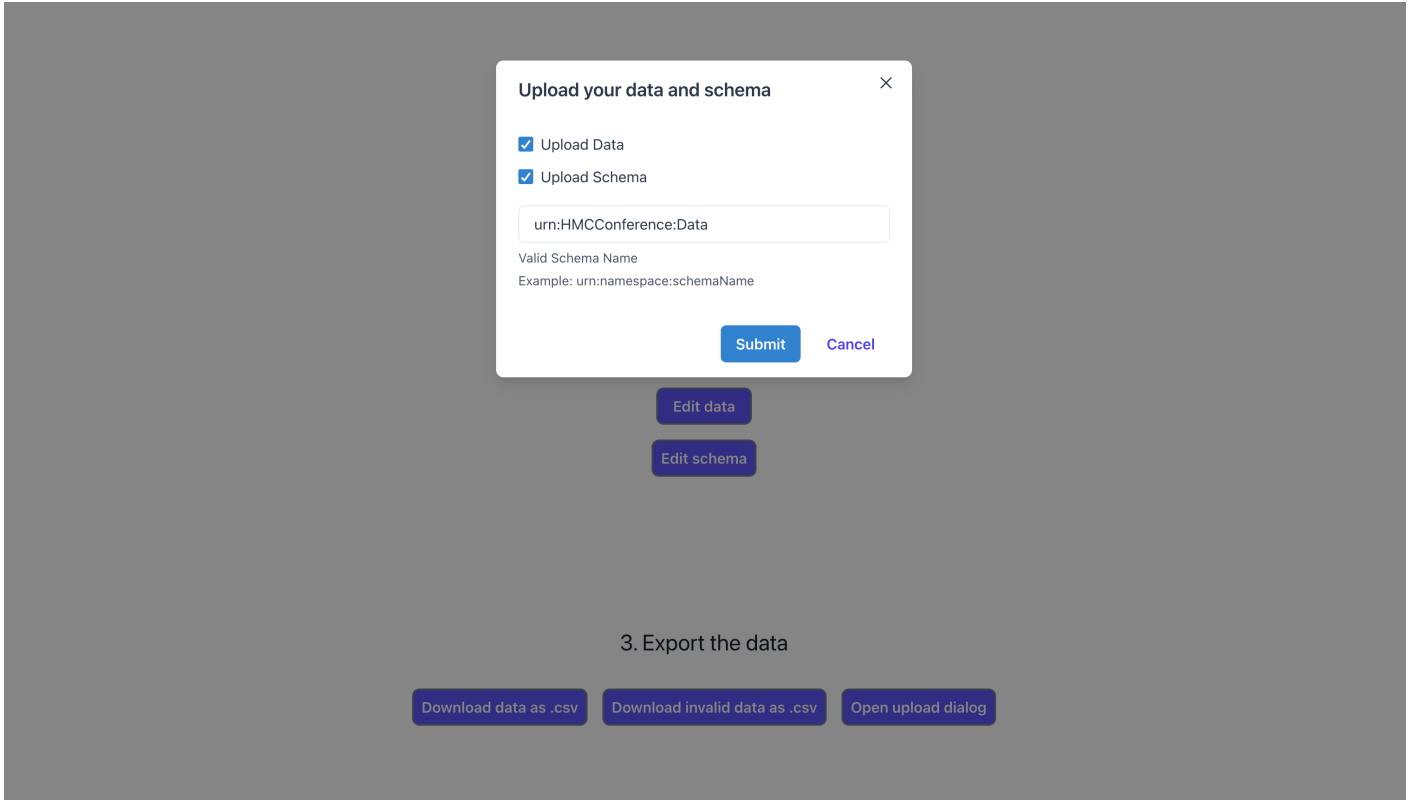
Data editor – complex harmonization



Schema editor



Upload data / schema to an API



HELMHOLTZ Metadata Collaboration

dkfz. GERMAN CANCER RESEARCH CENTER IN THE HELMHOLTZ ASSOCIATION
Research for a Life without Cancer

AST | E&E | ENERGY | **HEALTH** | INFORMATION | MATTER | DATA COMMONS | HMC OFFICE | HMC PROJECTS

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

Lucas Kulla[1 2], Philipp Schader[1], 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

www.helmholtz-metadaten.de