gesis

Leibniz Institute
for the Social Sciences
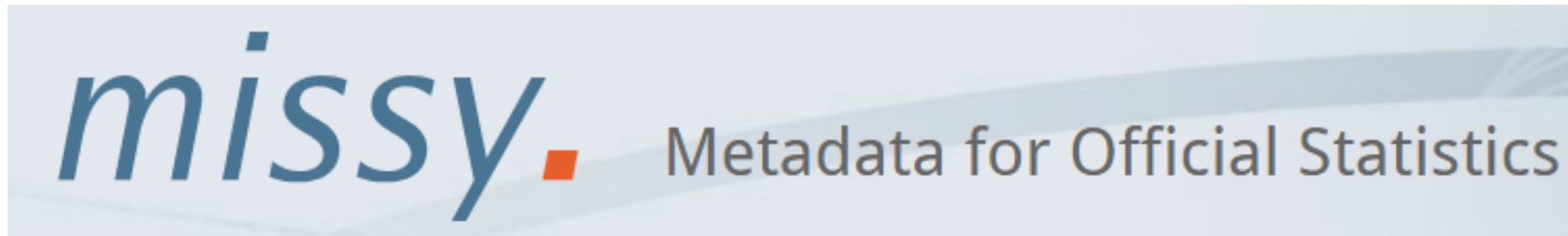
Stata export for metadata documentation

Munich, 26.05.2019

Anne Balz, Klaus Pforr, Florian Thirolf

Leibniz
Association

Motivation

- German Microdata Lab (GML) offers Metadata for various official microdata online

- Goal: extract Metadata from these Datasets automatically and import them into our database

  ▸ German Microcensus

  ▸ European Labour Force Survey

  ▸ EU-SILC (European Union Statistics on Income and Living Conditions)

Microdata-Informationsystem MISSY

- Online plattform („MISSY-web")

- Documentation of official microdata (European & national)

- Documentation on different levels:
  - study
  - question
  - variable

Microdata-Informationsystem MISSY

PB190 - Marital status (MT: 3,5=3) ▲

Basic Information **Description** **Values / Statistics**

ALL

AT BE BG CY CZ DE DK EE EL/GR ES FI FR HR

HU IT LT LU LV MT NL PL PT RO SE SI SK

PB190

Value	Value Label	Frequency	Total %	Valid %
1	never married	136531	28.2	28.5
2	married	261861	54.2	54.6
3	separated,MT:3,5=3	4834	1.0	1.0
4	widowed	44163	9.1	9.2
5	divorced	32193	6.7	6.7
	Valid Total	479582	99.2	100
(M)	Missing	3724	0.8	
	Total	**483306**	**100**	

Microdata-Informationsystem MISSY

core functionality

core functionality

*.dta

output.*
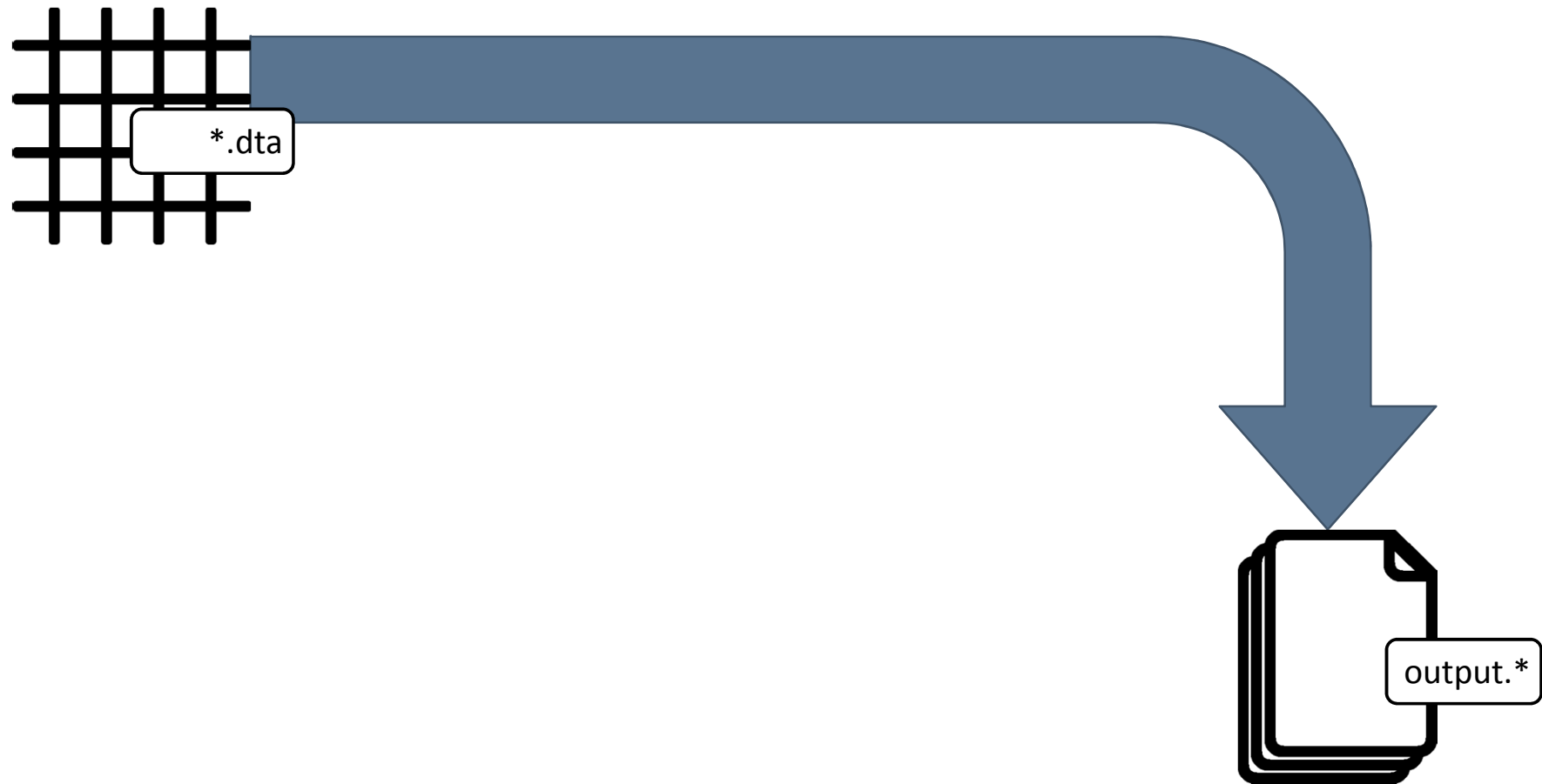
core functionality
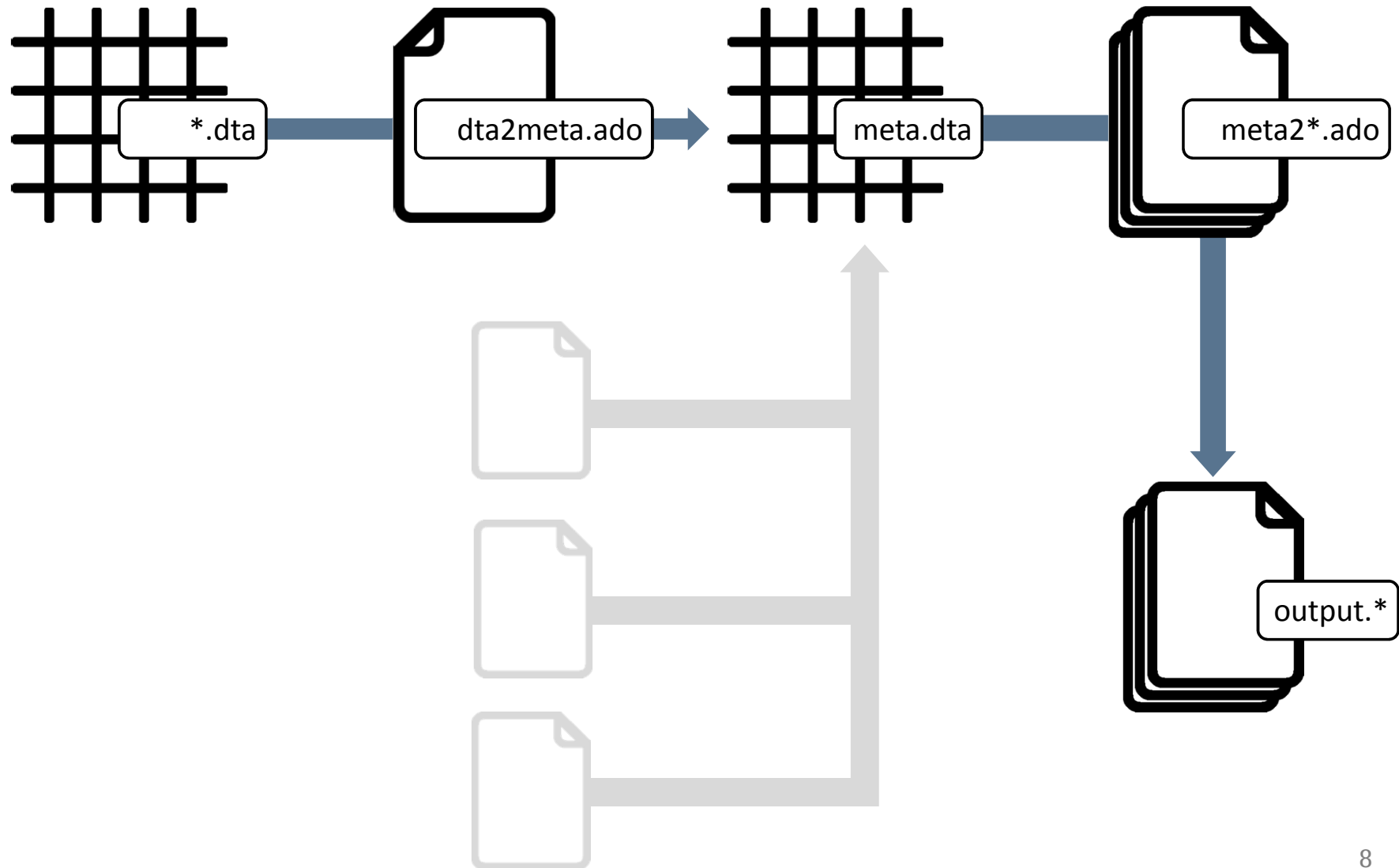
ado dta2md

ado dta2md
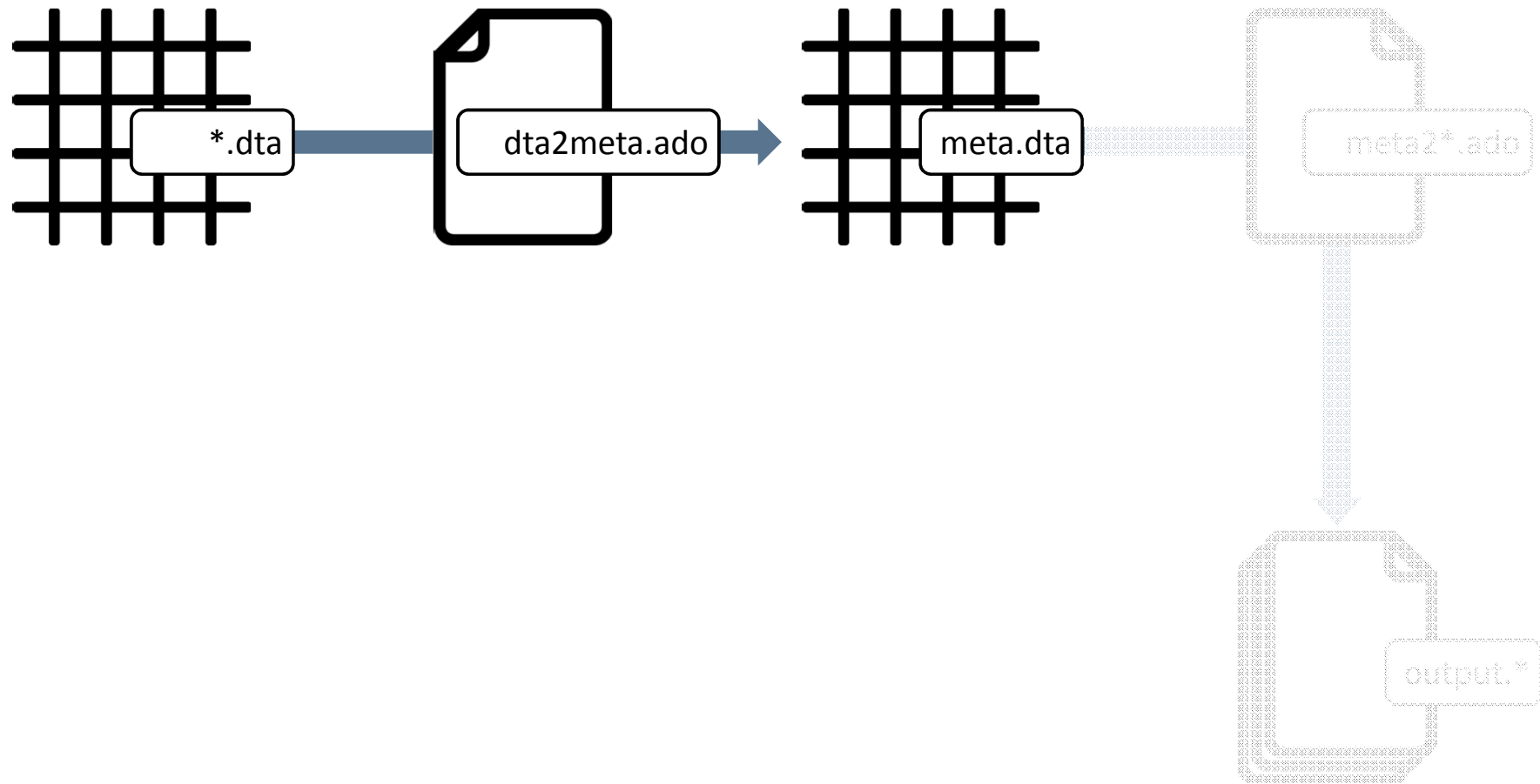
the meta-file

All necessary (meta-)information in a table format:

- Variable level
  - ▶ Varname, -label
  - ▶ Summary statistics (min, max, mean, std)

- Value level
  - ▶ Value, - label
  - ▶ Frequencies and percentages
    - Overall
    - For groups (e.g.: countries)

ado dta2md

the meta-file

	group	computed	varName	variableLabel	total_n	total_miss~g	min	max	Mean	StandardDe~n	value
1	AT	1	PL140	Type of contract	10778	2151	1	2	1.070824	.2565458	1
2	AT	1	PL140	Type of contract	10778	2151	1	2	1.070824	.2565458	2
3	AT	1	PL140	Type of contract	10778	2151	1	2	1.070824	.2565458	.
4	BE	1	PL140	Type of contract	11352	3134	1	2	1.119372	.3242456	1
5	BE	1	PL140	Type of contract	11352	3134	1	2	1.119372	.3242456	2
6	BE	1	PL140	Type of contract	11352	3134	1	2	1.119372	.3242456	.
7	BG	1	PL140	Type of contract	15382	2944	1	2	1.069625	.2545248	1
8	BG	1	PL140	Type of contract	15382	2944	1	2	1.069625	.2545248	2
9	BG	1	PL140	Type of contract	15382	2944	1	2	1.069625	.2545248	.
10	CY	1	PL140	Type of contract	9623	2712	1	2	1.116915	.3213423	1
11	CY	1	PL140	Type of contract	9623	2712	1	2	1.116915	.3213423	2
12	CY	1	PL140	Type of contract	9623	2712	1	2	1.116915	.3213423	.
13	CZ	1	PL140	Type of contract	16286	2635	1	2	1.101531	.3020416	1
14	CZ	1	PL140	Type of contract	16286	2635	1	2	1.101531	.3020416	2
15	CZ	1	PL140	Type of contract	16286	2635	1	2	1.101531	.3020416	.
16	DE	1	PL140	Type of contract	23081	2551	1	2	1.12265	.3280428	1
17	DE	1	PL140	Type of contract	23081	2551	1	2	1.12265	.3280428	2
18	DE	1	PL140	Type of contract	23081	2551	1	2	1.12265	.3280428	.

...

	group	computed	varName	variableLabel	total_n	total_miss~g	min	max	Mean	StandardDe~n	value
79	all	1	PL140	Type of contract	483306	163899	1	2	1.143331	.3504109	1
80	all	1	PL140	Type of contract	483306	163899	1	2	1.143331	.3504109	2
81	all	1	PL140	Type of contract	483306	163899	1	2	1.143331	.3504109	.

the meta-file

	group	computed	varName	value	valueLabel	n	percent	validPercent	isValid
1	AT	1	PL140	1	Permanent job: work contract of unlimited duration	8016	74.37373	92.91759	1
2	AT	1	PL140	2	temporary job: work contract of limited duration	611	5.668955	7.082416	1
3	AT	1	PL140	.		2151	19.95732	.	0
4	BE	1	PL140	1	Permanent job: work contract of unlimited duration	7237	63.75088	88.06279	1
5	BE	1	PL140	2	temporary job: work contract of limited duration	981	8.641649	11.93721	1
6	BE	1	PL140	.		3134	27.60747	.	0
7	BG	1	PL140	1	Permanent job: work contract of unlimited duration	11572	75.23079	93.03747	1
8	BG	1	PL140	2	temporary job: work contract of limited duration	866	5.629957	6.962534	1
9	BG	1	PL140	.		2944	19.13925	.	0
10	CY	1	PL140	1	Permanent job: work contract of unlimited duration	6103	63.42097	88.30849	1
11	CY	1	PL140	2	temporary job: work contract of limited duration	808	8.39655	11.69151	1
12	CY	1	PL140	.		2712	28.18248	.	0
13	CZ	1	PL140	1	Permanent job: work contract of unlimited duration	12265	75.31008	89.8469	1
14	CZ	1	PL140	2	temporary job: work contract of limited duration	1386	8.510377	10.1531	1
15	CZ	1	PL140	.		2635	16.17954	.	0
16	DE	1	PL140	1	Permanent job: work contract of unlimited duration	18012	78.03822	87.73502	1
17	DE	1	PL140	2	temporary job: work contract of limited duration	2518	10.90941	12.26498	1
18	DE	1	PL140	.		2551	11.05238	.	0

...

	group	computed	varName	value	valueLabel	n	percent	validPercent	isValid
79	all	1	PL140	1	Permanent job: work contract of unlimited duration	273626	56.61548	85.66688	1
80	all	1	PL140	2	temporary job: work contract of limited duration	45781	9.472466	14.33312	1
81	all	1	PL140	.		163899	33.91206	.	0

ado dta2md

Syntax

    dta2md output(filename) freqvarlist(varlist) input(filename) [group(varname) replace missingdef(exp) smissingdef(exp)]

| Arguments	Description
output	specifies where the metadata file is saved.
freqvarlist	list of variables for which information on value level is computed.
input	specifies where the Stata system file which is converted.
group	specifies group variable across which meta data are stratified.
replace	specifies if output file replaces existing file.
missingdeff	specifies definition which non-system-missing are considered as .
smissingdeff	specifies where the Stata system file which is converted.
relation	specifies where the Stata system file which is converted.
matrix(name[, subopts])	tabulate matrix name
e(name[, subopts])	tabulate matrix e(name)
r(name[, subopts])	tabulate matrix r(name)

Description

    dta2md converts a Stata system file to metadata.
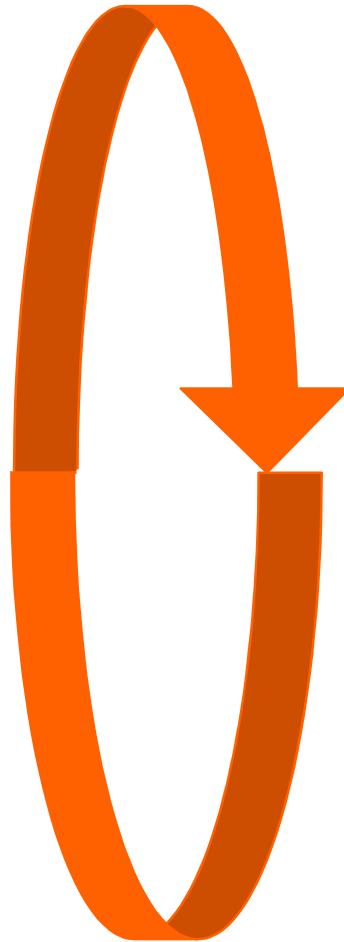
ado dta2md

```
dta2md input(filename) output(filename) //
        freqvarlist(varlist) //
        [group(varname) //
        missingdef(string) smissingdef(string) //
        replace ]


dta2md input($path/micro_file.dta) output($path/meta_file.dta)//
        freqvarlist(var1 var2 var3) //
        group(country) //
        missing("X<0") //
        smissingdef(`"X="invalid answer"| X="did not understand""') //
        replace
```
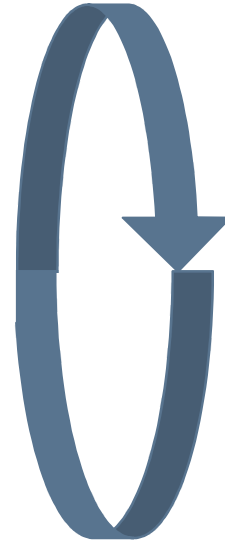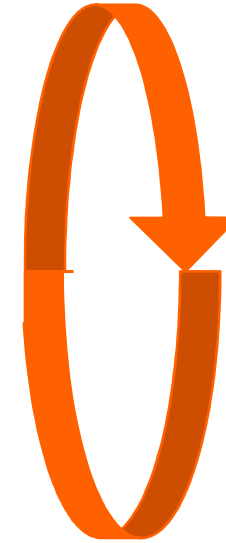
ado dta2md
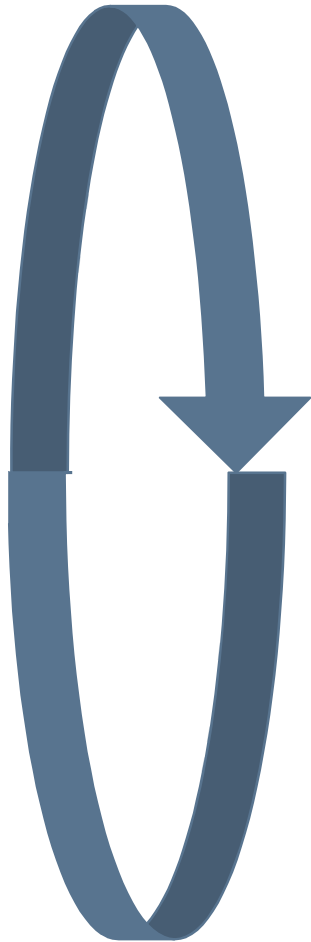
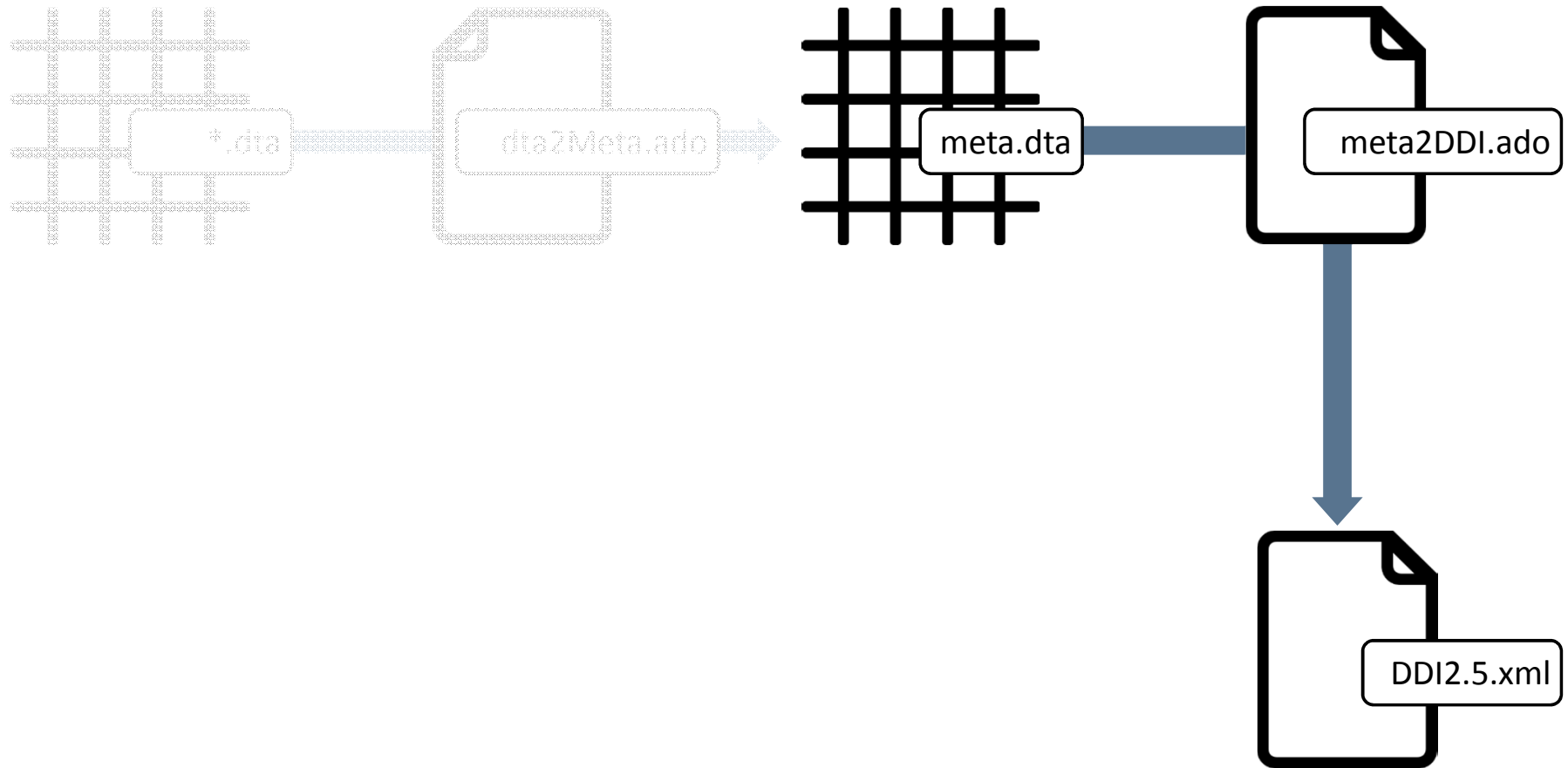Loop over all vars

If group specified:
Loop over all groups

If computed:

Loop over all levels
(within all vars)

If group specified:
Loop over all groups
(within levels of vars)

ado meta2DDI

ado meta2DDI

ado meta2DDI

- Uses the *„file'* command

- *„forvalues'* to run through all categories

- variables of the meta-file are used to form hierarchical output

- example:

  ▸ *„first'* (0/1) tags first category of a variable

  → used to generate output on variable level

```
    if (first[`i']) == 1 {
quietly file write ddiOutputLog`"          <var ID=""'
quietly file write ddiOutputLog (varName[`i'])
quietly file write ddiOutputLog"-"
quietly file write ddiOutputLog (group[`i'])
quietly file write ddiOutputLog`"" name=""'
quietly file write ddiOutputLog (varName[`i'])
quietly file write ddiOutputLog`"" intrvl=""'
    if (computed[`i']) == 1 {
quietly file write ddiOutputLog"discrete"
    }
    else {
quietly file write ddiOutputLog"contin"
    }
quietly file write ddiOutputLog`"">"'   _n
* Label
quietly file write ddiOutputLog"          <labl>"
quietly file write ddiOutputLog (variableLabel[`i'])
quietly file write ddiOutputLog"</labl>"_n
```

ado meta2DDI

group	varName	variableLabel	computed	total_n	total_miss~g	Mean	StandardDe~n	first
AT	PL140	Type of contract	1	10778	2151	1.070824	.2565458	1
AT	PL140	Type of contract	1	10778	2151	1.070824	.2565458	0
AT	PL140	Type of contract	1	10778	2151	1.070824	.2565458	0

```xml
<var ID="PL140-AT" name="PL140" intrvl="discrete">
    <labl>Type of contract</labl>
    <universe>AT</universe>
    <sumStat type="vald">8627</sumStat>
    <sumStat type="invd">2151</sumStat>
    <sumStat type="mean">1.0708241</sumStat>
    <sumStat type="stdev">.25654575</sumStat>
    <catgry missing="N">
        <catValu>1</catValu>
        <labl>Permanent job: work contract of unlimited duration</labl>
        <catStat type="freq">8016</catStat>
        <catStat type="percent">74.373726</catStat>
        <catStat type="other" otherType="validPercent">92.917587</catStat>
    </catgry>
    <catgry missing="N">
```

ado meta2DDI

isValid	value	valueLabel	n	percent	validPercent
1	1	Permanent job: work contract of unlimited duration	8016	74.37373	92.91759
1	2	temporary job: work contract of limited duration	611	5.668955	7.082416
0	.		2151	19.95732	.

```xml
<var ID="PL140-AT" name="PL140" intrvl="discrete">
    <labl>Type of contract</labl>
    <universe>AT</universe>
    <sumStat type="vald">8627</sumStat>
    <sumStat type="invd">2151</sumStat>
    <sumStat type="mean">1.0708241</sumStat>
    <sumStat type="stdev">.25654575</sumStat>
    <catgry missing="N">
        <catValu>1</catValu>
        <labl>Permanent job: work contract of unlimited duration</labl>
        <catStat type="freq">8016</catStat>
        <catStat type="percent">74.373726</catStat>
        <catStat type="other" otherType="validPercent">92.917587</catStat>
    </catgry>
    <catgry missing="N">
        <catValu>2</catValu>
```

usecase MISSY

Usecase MISSY
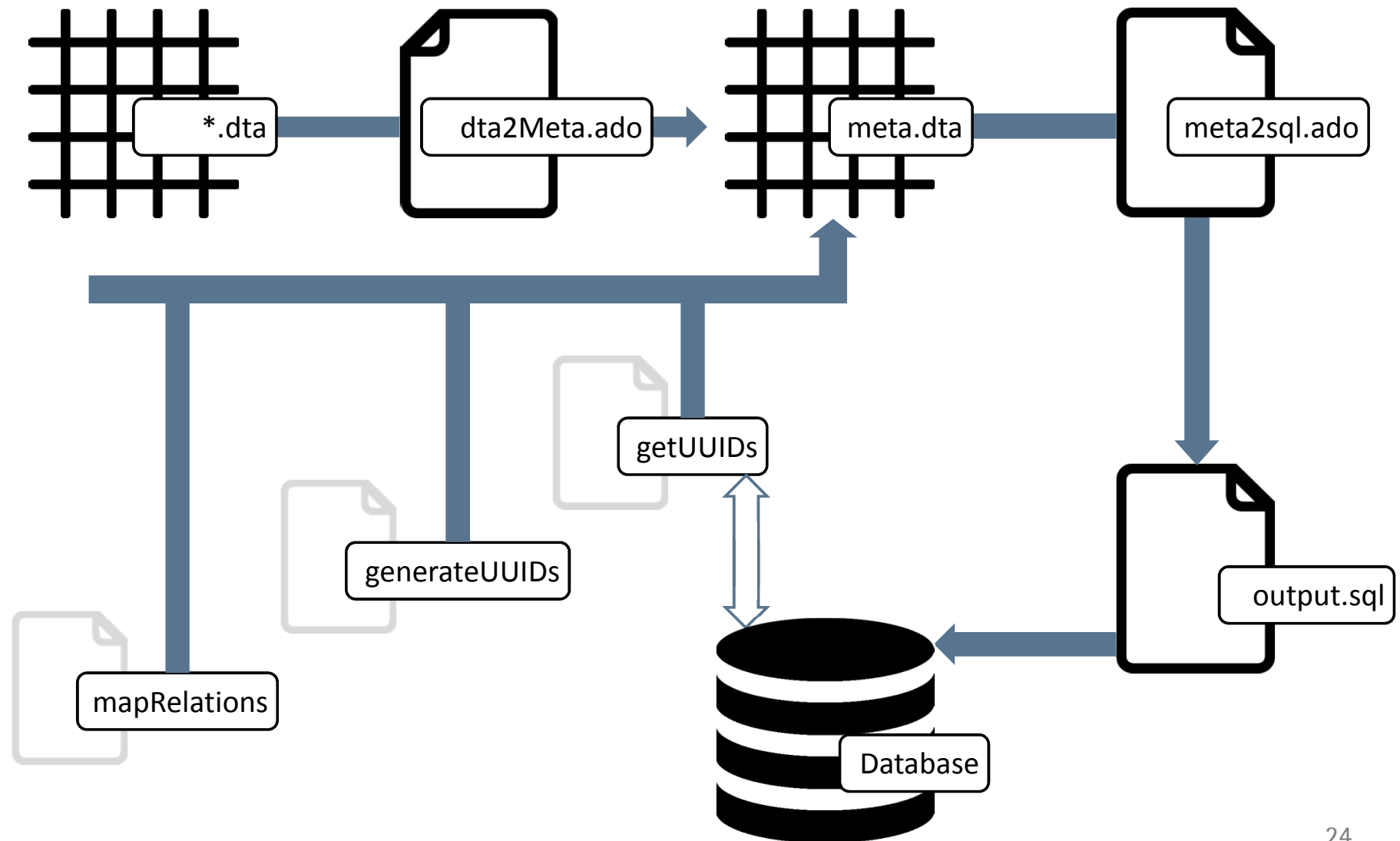
meta2sql.ado

- *‚file'* command is used

- different frame

- *‚forvalues'* for each database-table