

Metadata Management Variables and Variable Cascade

Jon Johnson

12 September 2019

CLOSER, UCL Institute of Education



Contents

- Datasets
- Variables and representation types
- Reuse of representations
- Variable cascade
- Use of the variable cascade
- Measurement Unit



Example: Dataset1

| name | height | birthdate | maritalstatus |
|-------|--------|------------|---------------|
| John | 178 | 1998-09-02 | S |
| Gill | 200 | 1934-06-12 | M |
| Alice | 182 | 1922-12-24 | M |



Variable example

| name | height | birthdate | maritalstatus |
|-------|--------|------------|---------------|
| John | 178 | 1998-09-02 | S |
| Gill | 200 | 1934-06-12 | M |
| Alice | 182 | 1922-12-24 | M |



Different variable representation types

| text | numeric | date | code |
|-------------|----------------|-------------|---------------|
| name | height | birthdate | maritalstatus |
| John | 178 | 1998-09-02 | S |
| Gill | 200 | 1934-06-12 | M |
| Alice | 182 | 1922-12-24 | M |

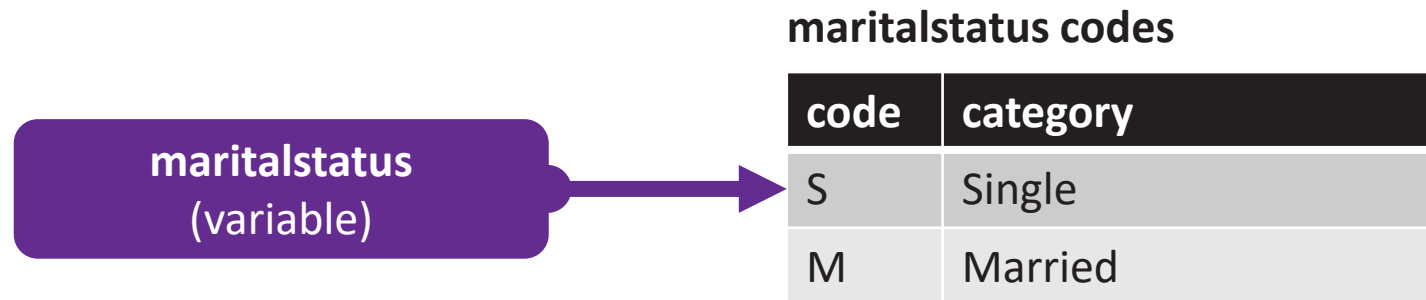


Variable with code representation

| name | height | birthdate | maritalstatus |
|-------|--------|------------|---------------|
| John | 178 | 1998-09-02 | S |
| Gill | 200 | 1934-06-12 | M |
| Alice | 182 | 1922-12-24 | M |



Variable - code representation





Three similar datasets

dataset1

| name | height | birthdate | maritalstatus |
|-------|--------|------------|---------------|
| John | 178 | 1998-09-02 | S |
| Gill | 200 | 1934-06-12 | M |
| Alice | 182 | 1922-12-24 | M |

dataset2

| firstname | personheight | dateofbirth | maritalstatus2010 |
|-----------|--------------|-------------|-------------------|
| Bob | 70 | 1995-09-02 | S |
| Lars | 76 | 1954-06-21 | M |
| Gerald | 66 | 1972-11-23 | S |

dataset3

| firstname | imperialheight | dateofbirth | maritalstatus2018 |
|-----------|----------------|-------------|-------------------|
| Lisa | 69 | 1995-09-02 | S |
| Bart | 75 | 1954-06-21 | M |
| Homer | 68 | 1972-11-23 | D |



Three similar datasets – code representation

dataset1

| name | height | birthdate | maritalstatus |
|-------|--------|------------|---------------|
| John | 178 | 1998-09-02 | S |
| Gill | 200 | 1934-06-12 | M |
| Alice | 182 | 1922-12-24 | M |

dataset2

| firstname | personheight | dateofbirth | maritalstatus2010 |
|-----------|--------------|-------------|-------------------|
| Bob | 70 | 1995-09-02 | S |
| Lars | 76 | 1954-06-21 | M |
| Gerald | 66 | 1972-11-23 | S |

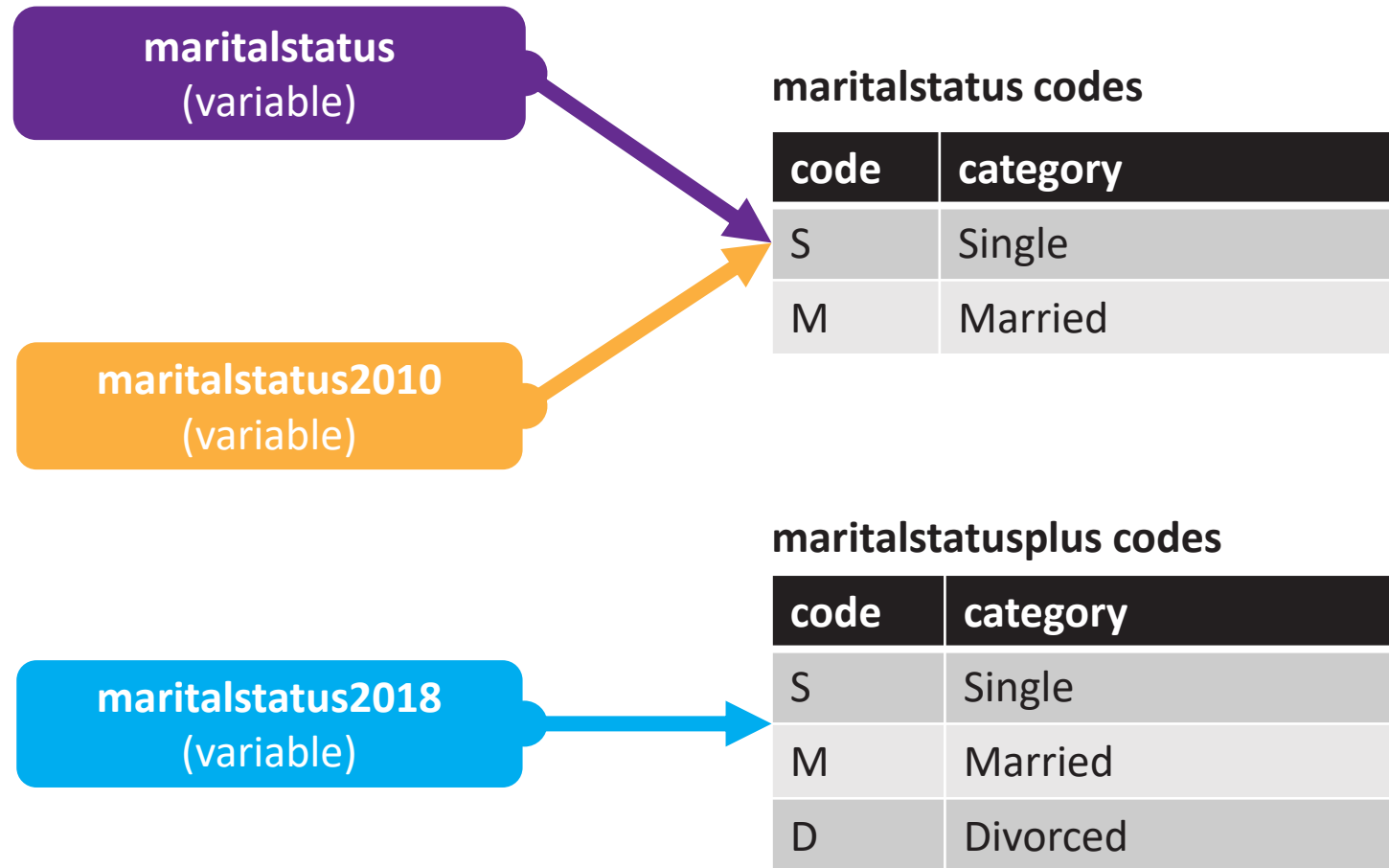
dataset3

| firstname | imperialheight | dateofbirth | maritalstatus2018 |
|-----------|----------------|-------------|-------------------|
| Lisa | 69 | 1995-09-02 | S |
| Bart | 75 | 1954-06-21 | M |
| Homer | 68 | 1972-11-23 | D |



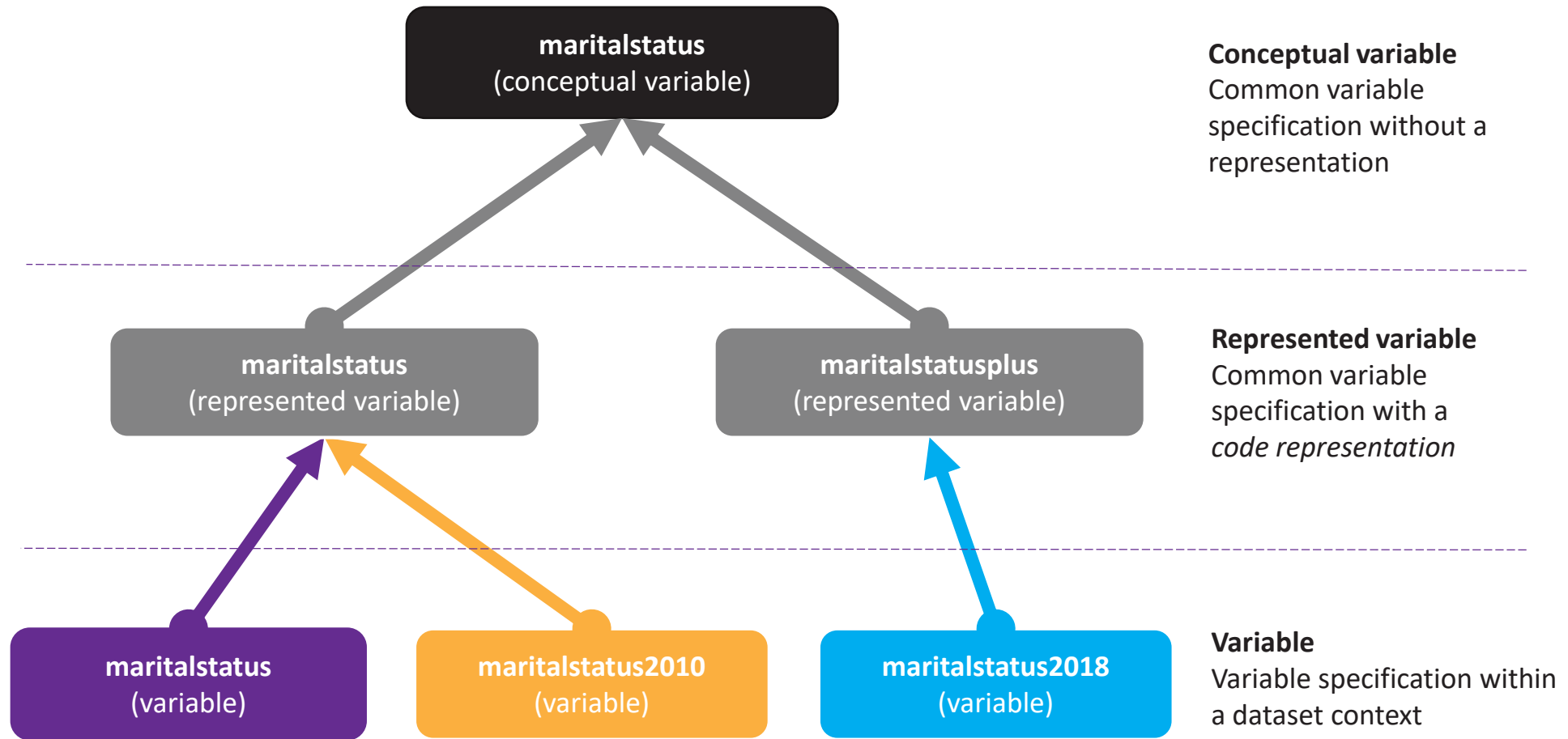


Variables re-using sets of codes





Documenting comparabilities among variables





Three similar datasets - text representation

dataset1

| name | height | birthdate | maritalstatus |
|-------|--------|------------|---------------|
| John | 178 | 1998-09-02 | S |
| Gill | 200 | 1934-06-12 | M |
| Alice | 182 | 1922-12-24 | M |

dataset2

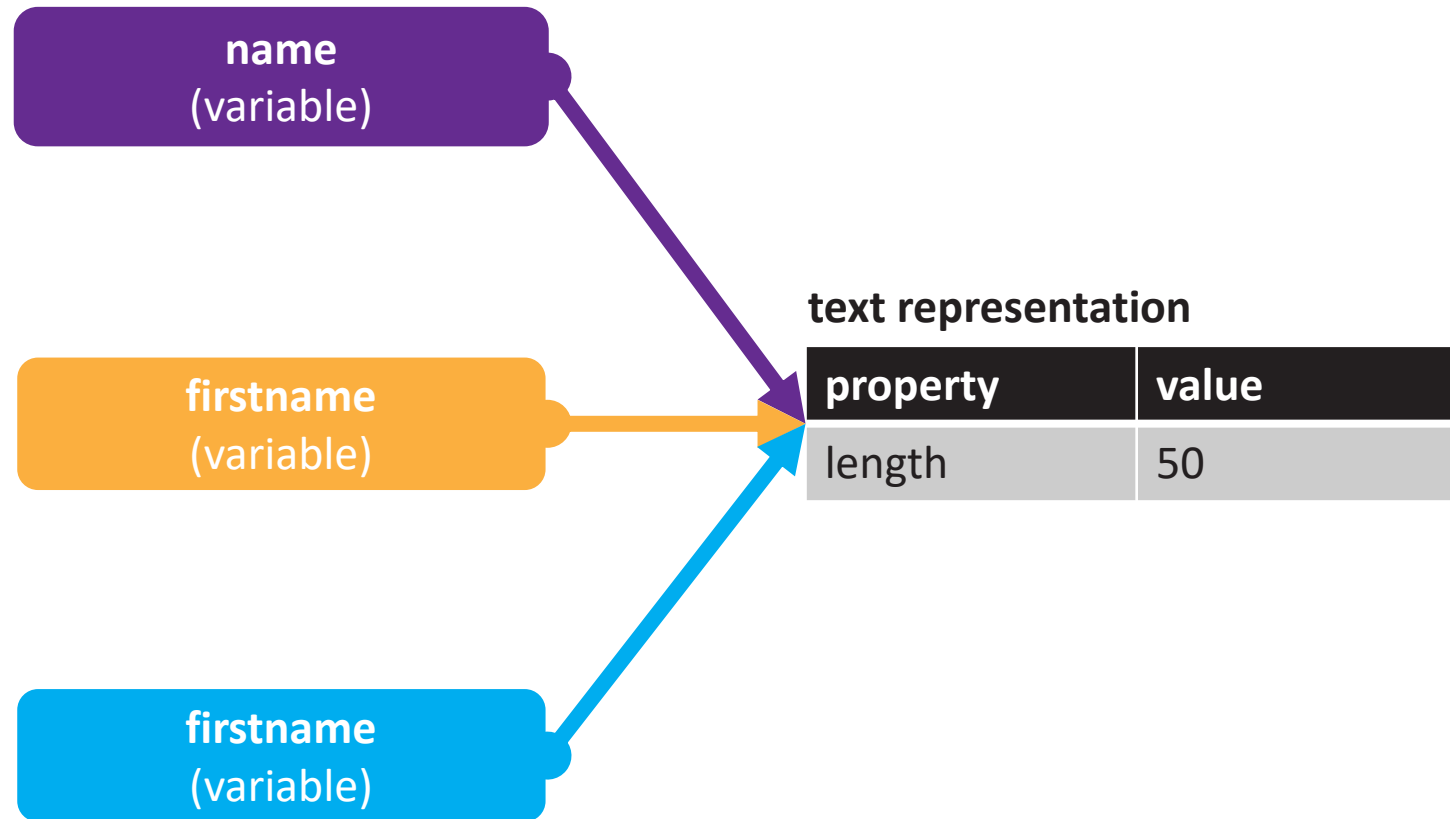
| firstname | personheight | dateofbirth | maritalstatus2010 |
|-----------|--------------|-------------|-------------------|
| Bob | 70 | 1995-09-02 | S |
| Lars | 76 | 1954-06-21 | M |
| Gerald | 66 | 1972-11-23 | S |

dataset3

| firstname | imperialheight | dateofbirth | maritalstatus2018 |
|-----------|----------------|-------------|-------------------|
| Lisa | 69 | 1995-09-02 | S |
| Bart | 75 | 1954-06-21 | M |
| Homer | 68 | 1972-11-23 | D |

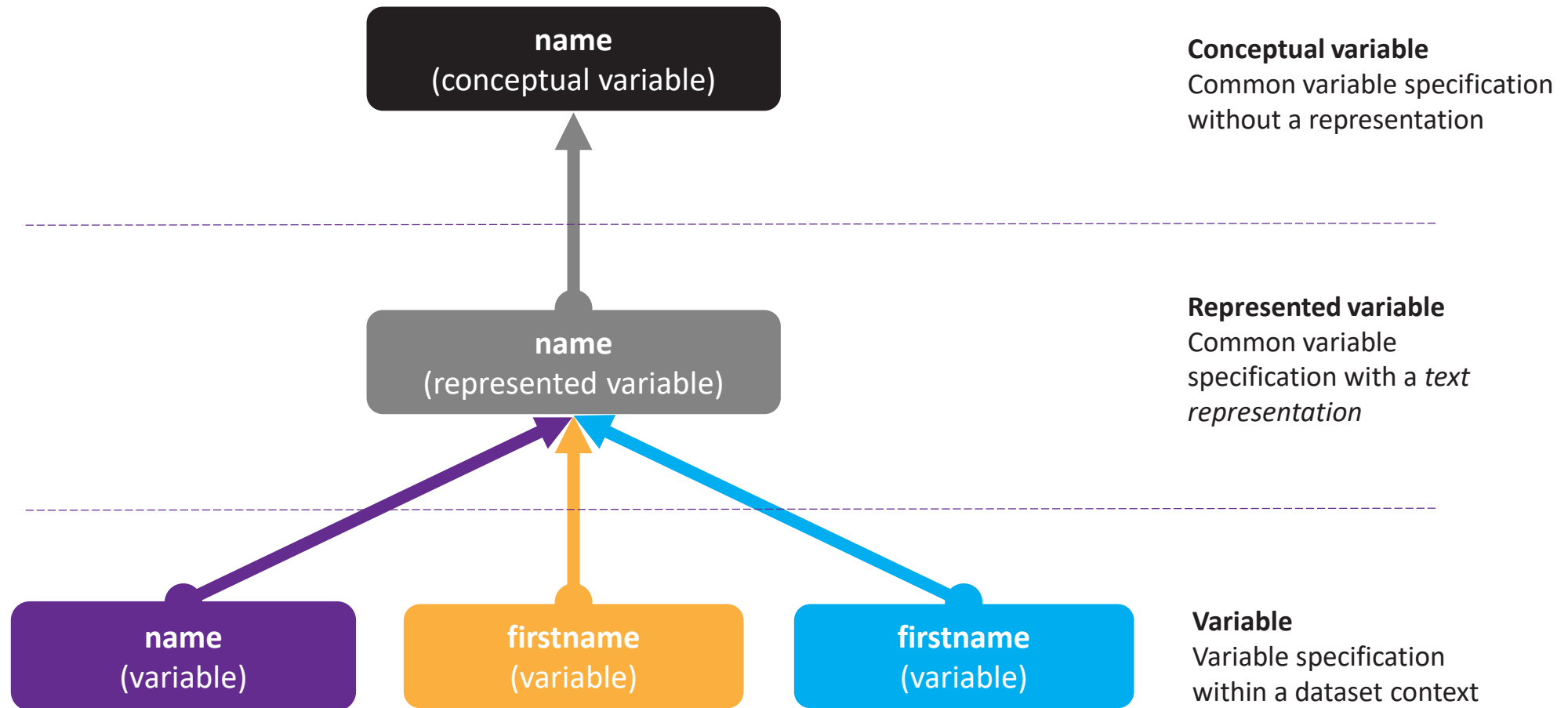


Variables with the same text representation type





Documenting comparabilities among variables





Three similar datasets - numeric representation

dataset1

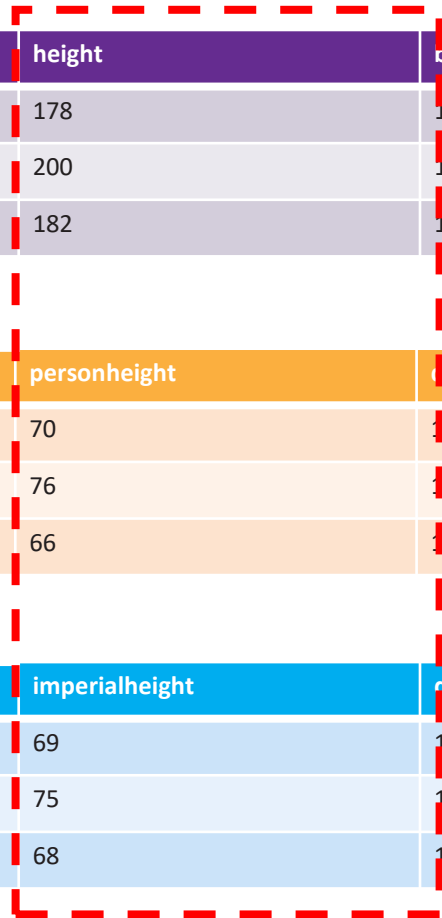
| name | height | birthdate | maritalstatus |
|-------|--------|------------|---------------|
| John | 178 | 1998-09-02 | S |
| Gill | 200 | 1934-06-12 | M |
| Alice | 182 | 1922-12-24 | M |

dataset2

| firstname | personheight | dateofbirth | maritalstatus2010 |
|-----------|--------------|-------------|-------------------|
| Bob | 70 | 1995-09-02 | S |
| Lars | 76 | 1954-06-21 | M |
| Gerald | 66 | 1972-11-23 | S |

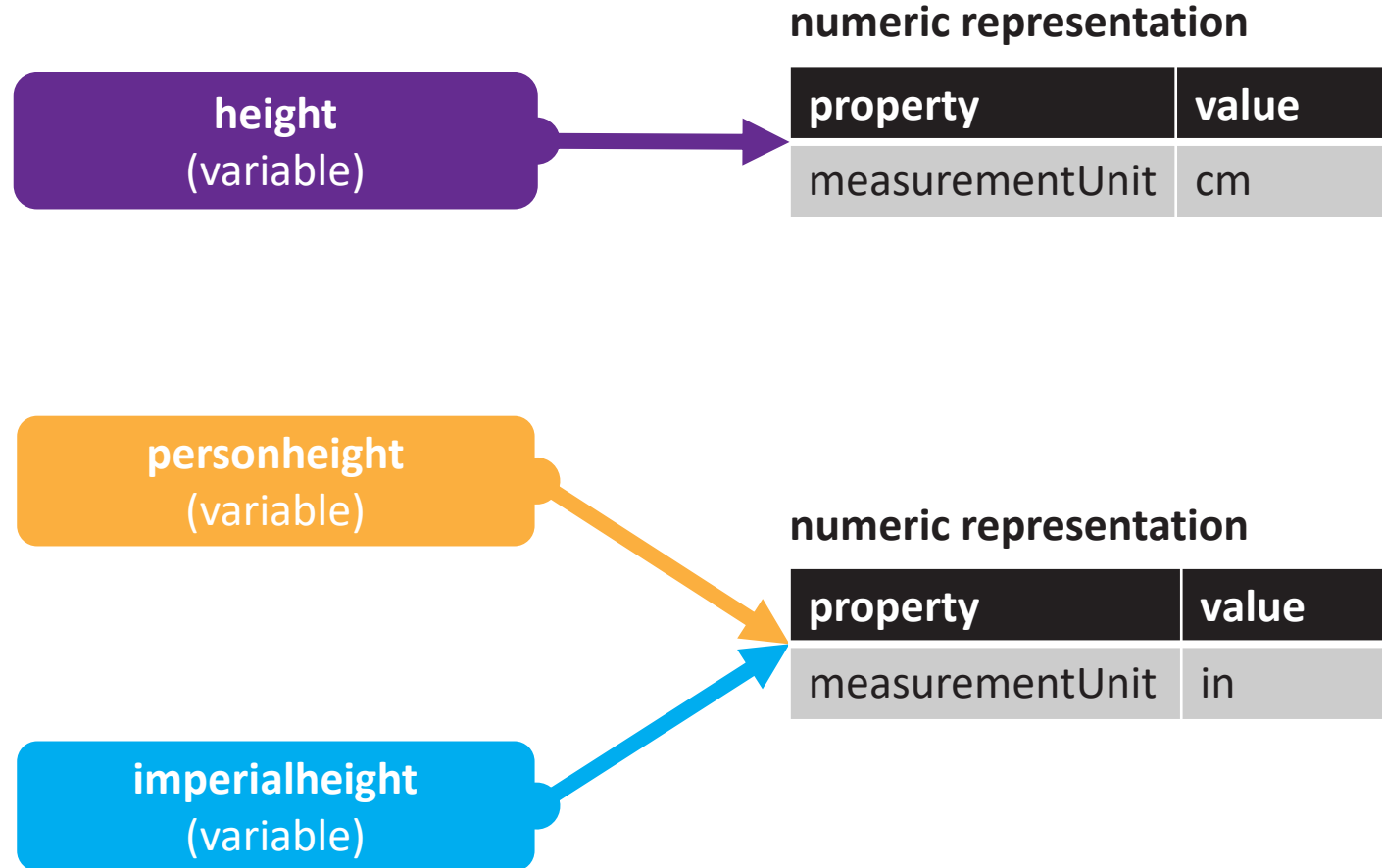
dataset3

| firstname | imperialheight | dateofbirth | maritalstatus2018 |
|-----------|----------------|-------------|-------------------|
| Lisa | 69 | 1995-09-02 | S |
| Bart | 75 | 1954-06-21 | M |
| Homer | 68 | 1972-11-23 | D |



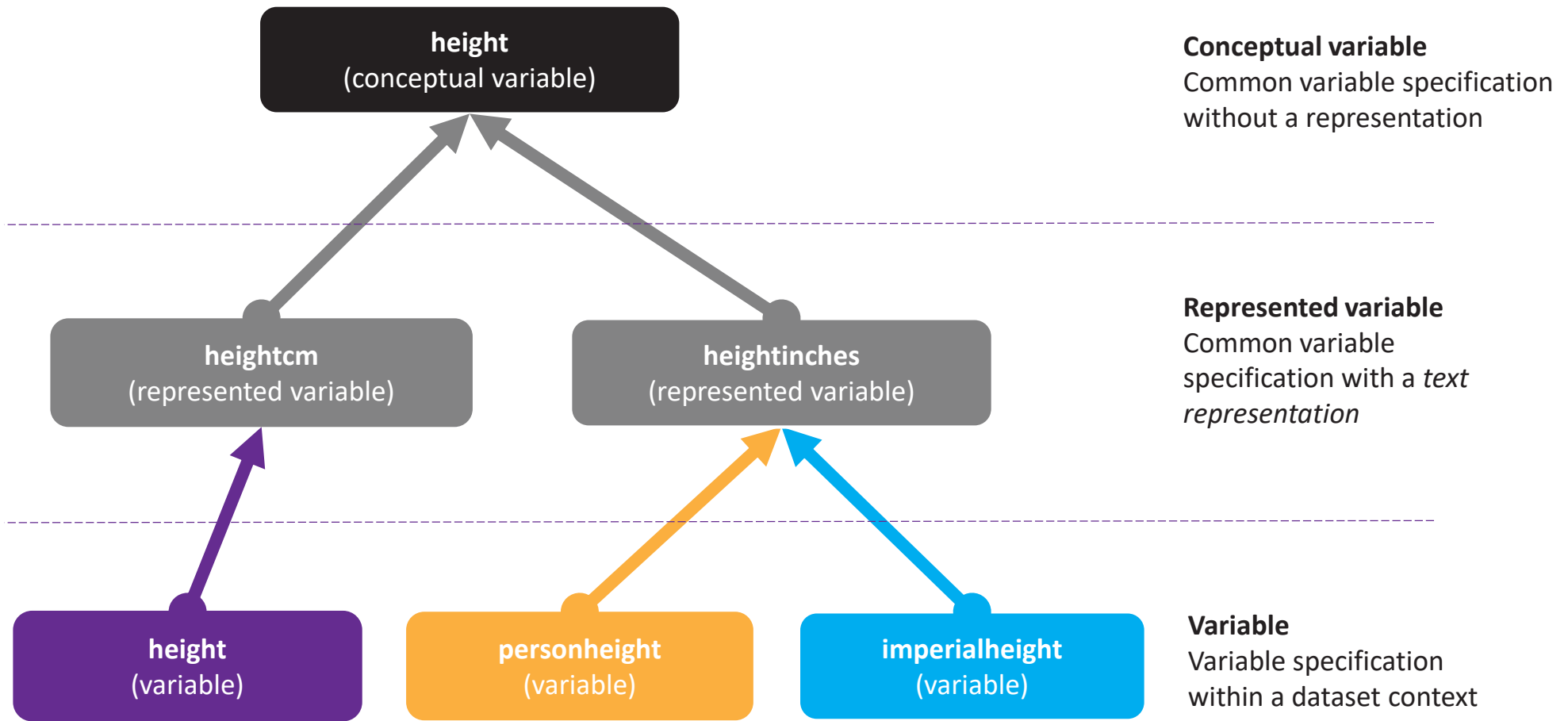


Variables with different unit of measure





Documenting comparabilities among variables





Benefits of the variable cascade structure

- Specify comparability across variables allows comparison across datasets
- Structure for facilitating harmonization across datasets / studies
- Variable names and representations can change
- Documenting changes over time
- Planning for future data collection to ensure comparability



England & Wales Census – Tenure *

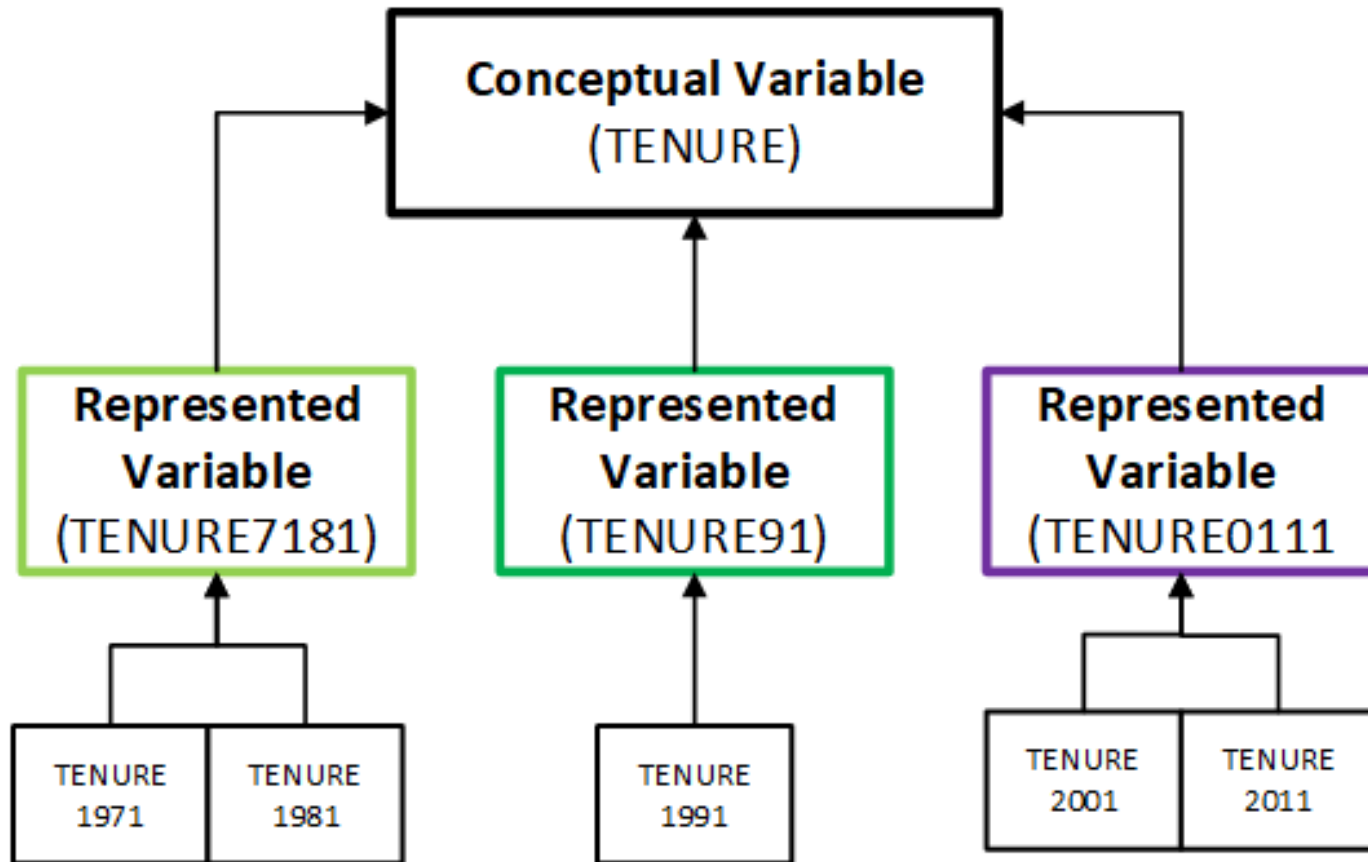
| Categories | TENURE | | | | |
|--|--------|------|------|------|------|
| | 1971 | 1981 | 1991 | 2001 | 2011 |
| Owner occupied freehold | 1 | 1 | | | |
| Owner occupied leasehold | 2 | 2 | | | |
| Renting from local authority | 3 | 3 | 4 | 3 | 3 |
| Renting with a job, shop, farm or other business | 4 | 4 | 3 | | |
| Renting from a housing association or charitable trust | 5 | 5 | 6 | | |
| Other rented, furnished | 6 | 6 | | | |
| Other rented, unfurnished | 7 | 7 | | | |
| Owner-occupier: mortgage | | | 1 | 1 | 1 |
| Owner-occupier: outright | | | 2 | 0 | 0 |
| New Town Development Corporation or HAT | | | 5 | | |
| Private landlord - furnished | | | 7 | | |
| Private landlord - unfurnished | | | 8 | | |
| Shared ownership (part owned and part rented) | | | | 2 | 2 |
| Social rented, Other social rented | | | | 4 | 4 |
| Private rented, Private landlord or letting agency | | | | 5 | 5 |
| Private rented, Employer of a household member | | | | 6 | 6 |
| Private rented, Relative or friend of household member | | | | 7 | 7 |
| Private rented, Other | | | | 8 | 8 |
| Lives rent free | | | | 9 | 9 |

* the conditions under which land or buildings are held or occupied





England & Wales Census – Variable Cascade





Tenure in Longitudinal studies*

| Study | TENURE7181 | TENURE91 | TENURE0111 |
|------------------------------|------------|----------|------------|
| ALSPAC | | X | X |
| BCS70 | X | X | X |
| Census Longitudinal Study | X | X | X |
| NCDS | X | X | X |
| NSHD | X | X | X |
| Millennium Cohort Study | | | X |
| Southampton Women's Study | | | X |
| Understanding Society / BHPS | | X | X |

X – single variable **X** – multiple variables

* simplified

