Guidelines for handling variables in repeated contexts

Romain Tailhurat - Making Sense

A collective work

- Why?
  - ‣ Managing repeated measures over time in order to
    - – improve discovery and comparison
    - – Enable better metadata management
    - – Support harmonisation of data

A collective work

- Who?
  - ‣ Domain experts
    - – CDSP
    - – CLOSER
    - – Constances
    - – France Cohortes
    - – Insee
  - ‣ Making Sense as a coordinator

A collective work

- How?
  - ‣ Opportunistic meetings (in-person and remote)
  - ‣ Open documentation (Varese)

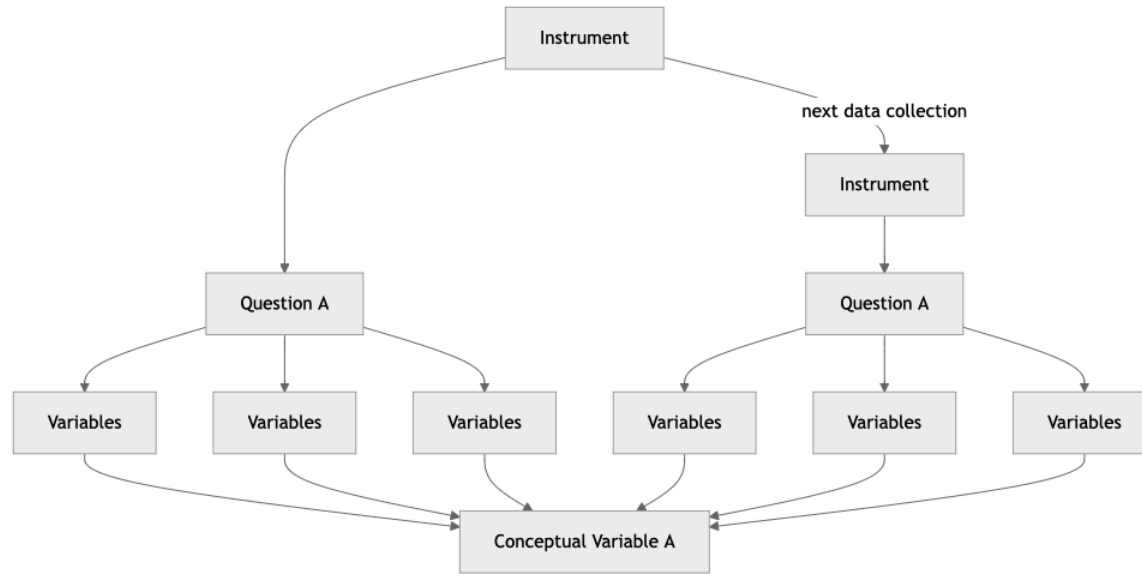# What has been done so far?

- EDDI23 talk
- Identify more use cases
- Describe tools and patterns
- Draft a guideline document

Use cases

- Original and new uses cases from Constances, Insee and CLOSER
- and new ones from new partners, CDSP, France Cohortes
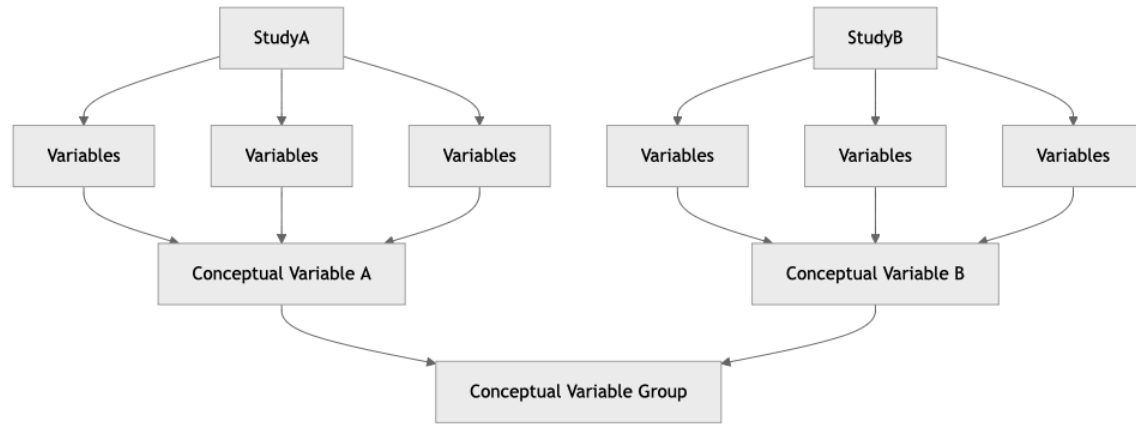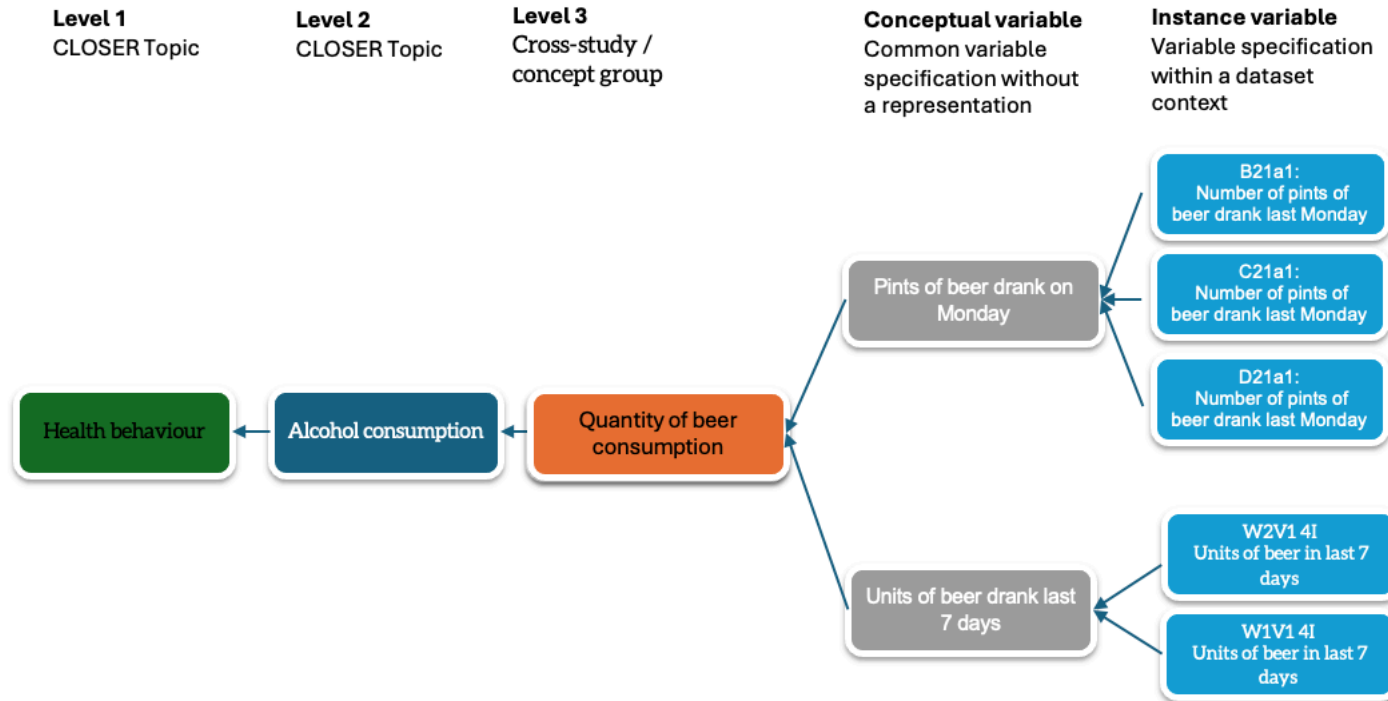
CLOSER

- Within studies
  - ‣ Variables are referenced by the same conceptual variable within a single longitudinal study
  - ‣ Definition of a conceptual variable within a study is strongly related to the question item within a questionnaire

```
                          ┌────────────┐
                          │ Instrument │
                          └────────────┘
                        ╱                ╲
                       ╱                  ╲  next data collection
                      ╱                    ↓
                     │                 ┌────────────┐
                     │                 │ Instrument │
                     │                 └────────────┘
                     ↓                        ↓
              ┌────────────┐           ┌────────────┐
              │ Question A │           │ Question A │
              └────────────┘           └────────────┘
```
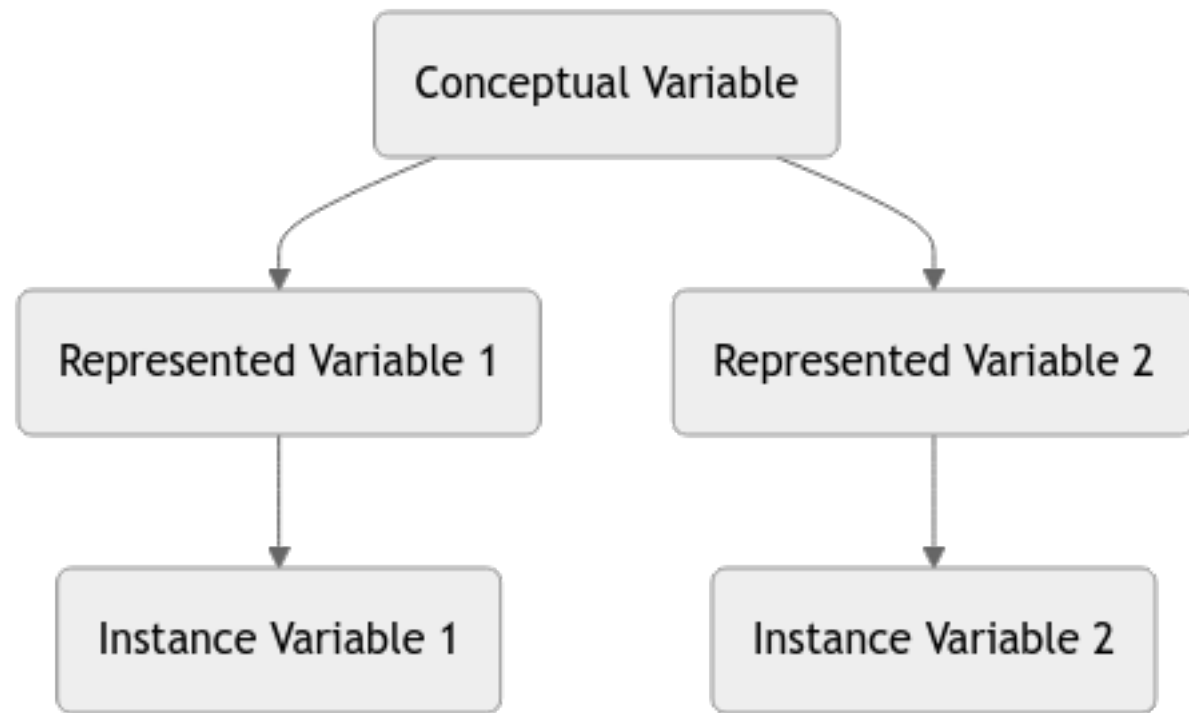
CLOSER

- Across studies
  - ‣ Cross study conceptual variable groups are created which reference study level conceptual variables
  - ‣ This enables a researcher to find similar variables across longitudinal population study (LPS) for potential harmonisation

```
                    StudyA                                      StudyB

    Variables    Variables    Variables        Variables    Variables    Variables

              Conceptual Variable A                    Conceptual Variable B

                          Conceptual Variable Group
```

Level 1
CLOSER Topic

Level 2
CLOSER Topic

Level 3
Cross-study /
concept group

Conceptual variable
Common variable
specification without
a representation

Instance variable
Variable specification
within a dataset
context

Health behaviour

Alcohol consumption

Quantity of beer
consumption

Pints of beer drank on
Monday

Units of beer drank last
7 days

B21a1:
Number of pints of
beer drank last Monday

C21a1:
Number of pints of
beer drank last Monday

D21a1:
Number of pints of
beer drank last Monday

W2V1 4I
Units of beer in last 7
days

W1V1 4I
Units of beer in last 7
days

Constances (wip)

- Documenting changes in response domains

```mermaid
flowchart TD
    A[Conceptual Variable]
    B[Represented Variable 1]
    C[Represented Variable 2]
    D[Instance Variable 1]
    E[Instance Variable 2]
    A --> B
    A --> C
    B --> D
    C --> E
```

Conceptual Variable

Represented Variable 1

Represented Variable 2

Instance Variable 1

Instance Variable 2

Represented Variable 2 —Based on— Represented Variable 1

(Gesis) `IsBasedOn.DifferentResponseSchema`

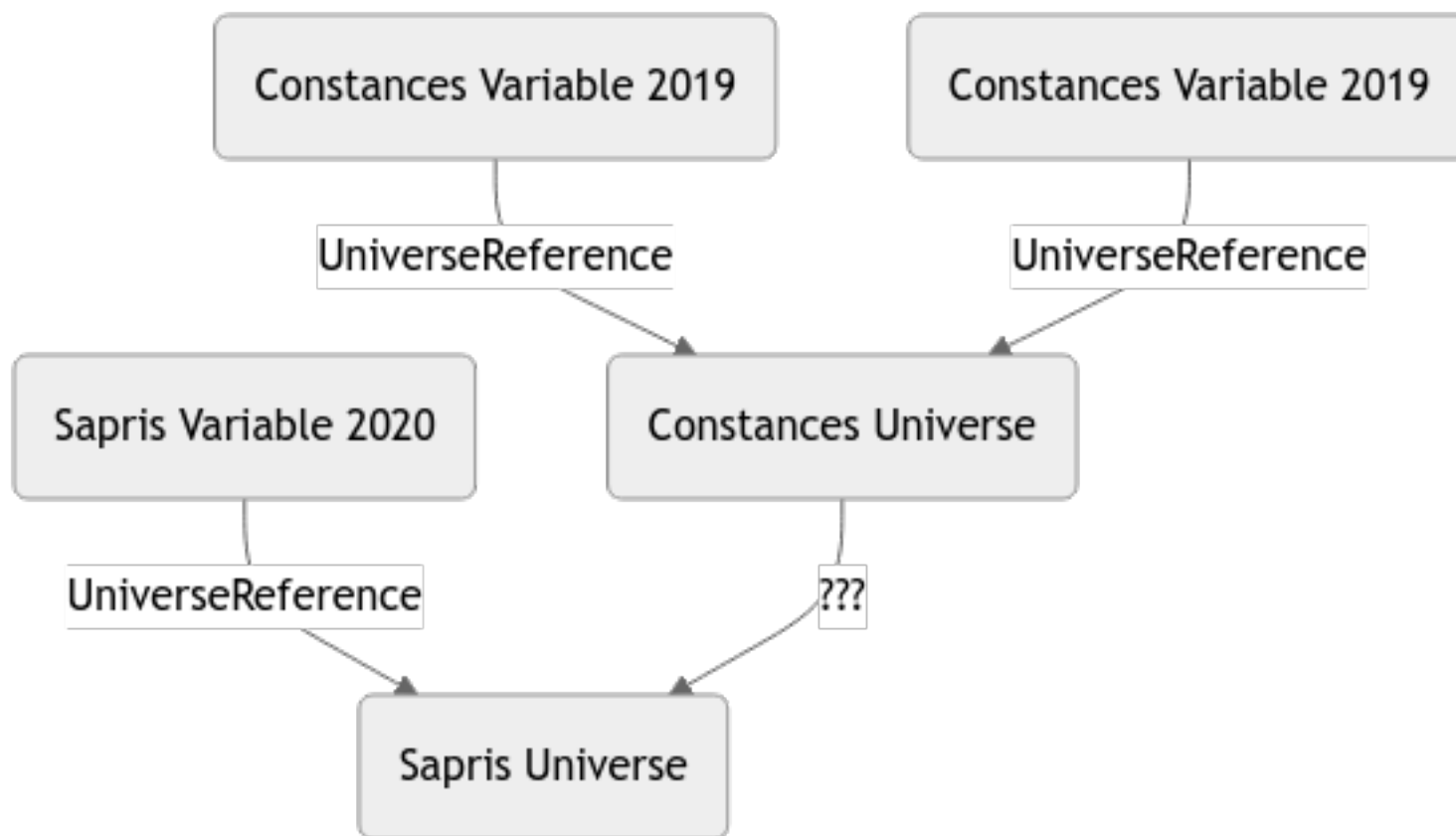# Constances (wip)

- Documenting change in concepts
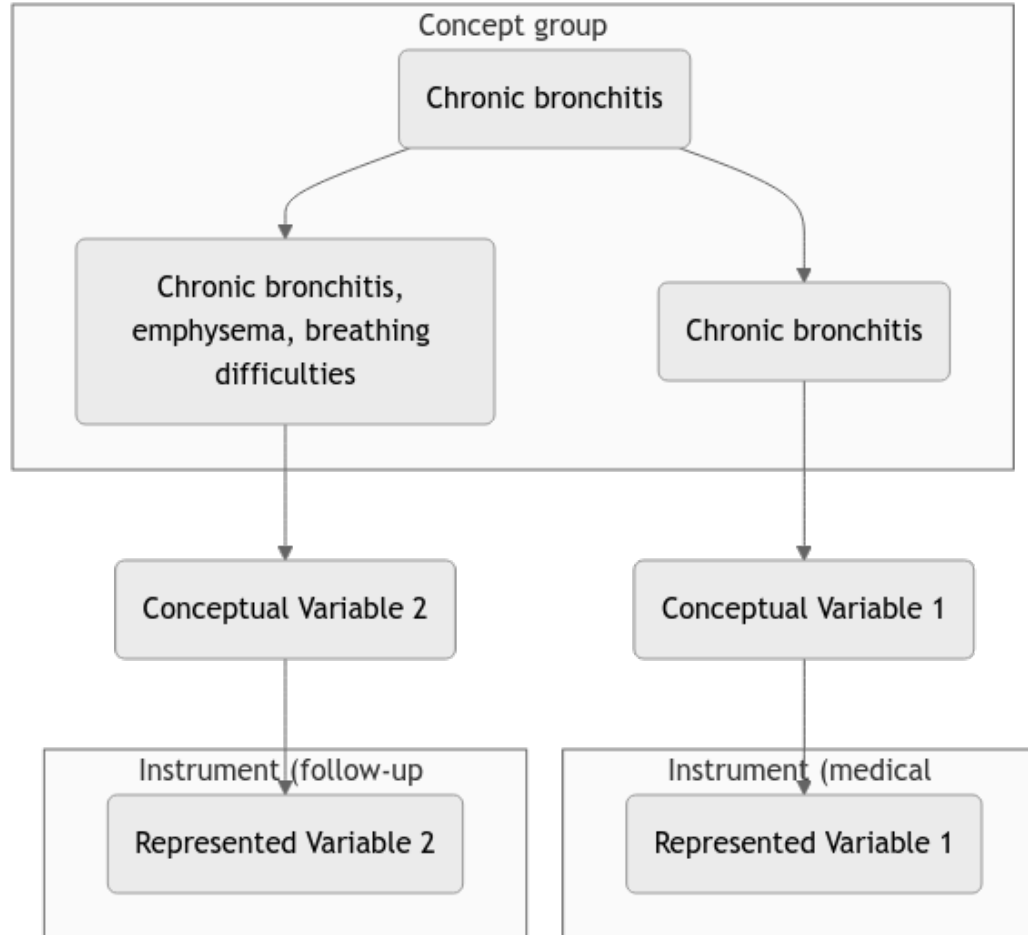
Constances (wip)

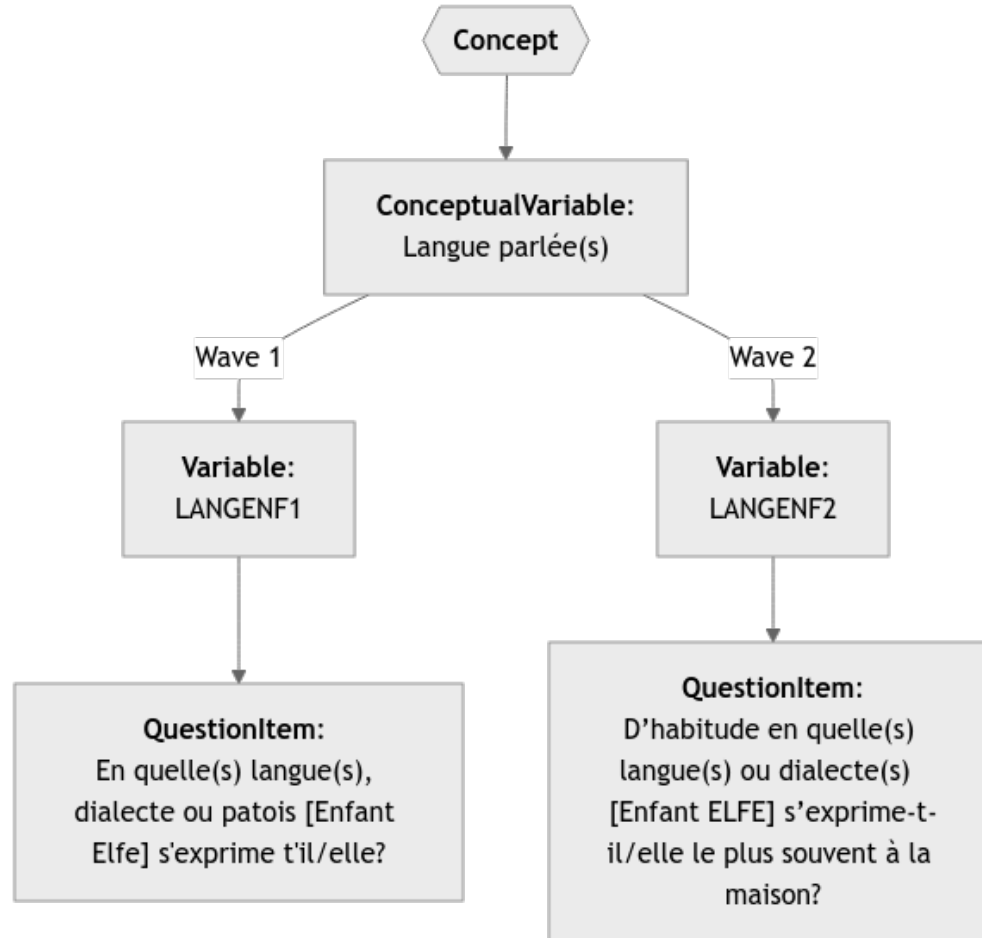- use of SKOS to provide qualified relations between concepts?
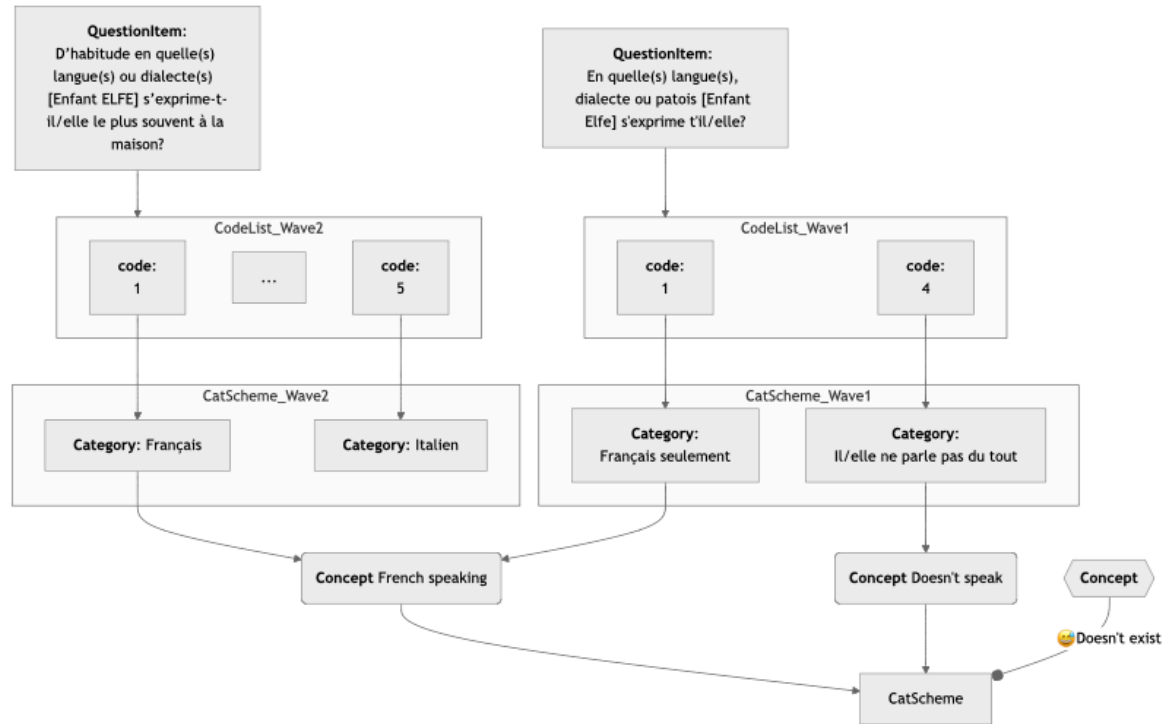
Constances (wip)

- Documenting change in universe ❌

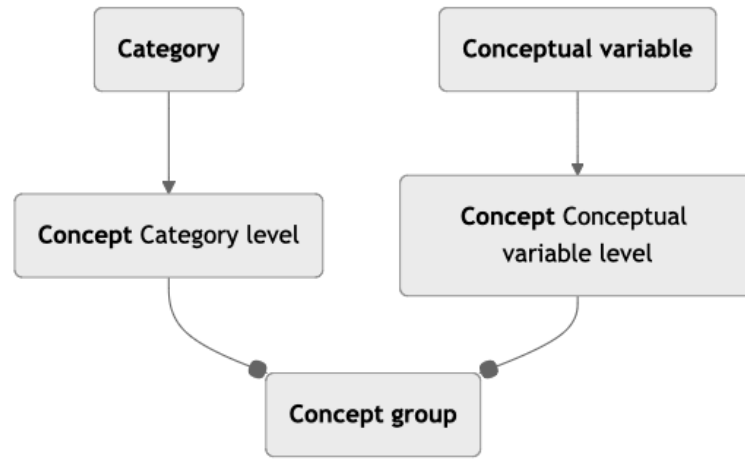# France Cohortes (wip)

- Changes in categories

QuestionItem:
D'habitude en quelle(s) langue(s) ou dialecte(s) [Enfant ELFE] s'exprime-t-il/elle le plus souvent à la maison?

QuestionItem:
En quelle(s) langue(s), dialecte ou patois [Enfant Elfe] s'exprime t'il/elle?

CodeList_Wave2

code: 1

...

code: 5

CodeList_Wave1

code: 1

code: 4

CatScheme_Wave2

Category: Français

Category: Italien

CatScheme_Wave1

Category: Français seulement

Category: Il/elle ne parle pas du tout

Concept French speaking

Concept Doesn't speak

Concept

😊Doesn't exist

CatScheme

```
┌─────────────┐          ┌──────────────────┐          ┌──────────────────────┐
│  Category   │          │ Conceptual variable│         │ How to handle granularity? │
└─────────────┘          └──────────────────┘          │          🤨           │
       │                          │                     └──────────────────────┘
       ▼                          ▼
┌──────────────────┐     ┌──────────────────────┐
│ Concept Category │     │ Concept Conceptual   │
│      level       │     │   variable level     │
└──────────────────┘     └──────────────────────┘
           │                      │
           └──────────●   ●───────┘
                ┌──────────────┐
                │ Concept group│
                └──────────────┘
```

Patterns

- Grouping at the concept / variable level
- Grouping at the study level
- Direct qualified relations between variables
- Indirect relations using the variable cascade (harmonisation at the concept level)
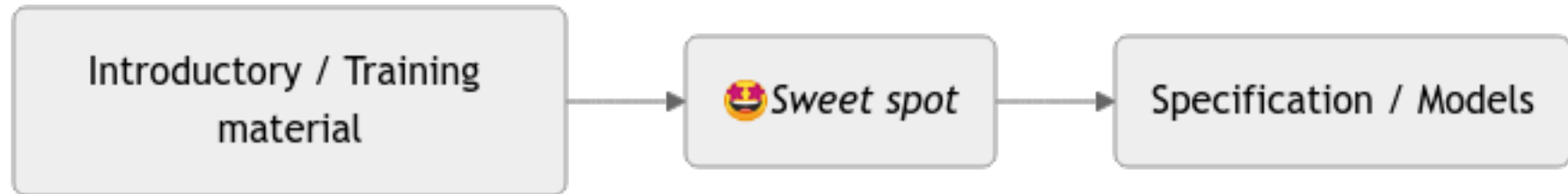- Linking universes using subuniverses or BasedOn attribute (proposed extension to Gesis vocab)

Pending issues

- Relations between universes?
- Extension of the BasedOn links between variables
- Provide clues on which patterns to use?
  - ‣ cf. the Whys? Discovery/comparison, management, harmonisation

Next steps

- Finish and publish the guidelines document
  - ▸ Publish on ddialliance.org?
- Work on issues
- Recruit!

Sweet spot

Introductory / Training material → 🤩 *Sweet spot* → Specification / Models

Thank you!