



RIKSBANKENS
JUBILEUMSFOND

STIFTELSEN FÖR HUMANISTISK OCH
SAMHÄLLSVETENSKAPLIG FORSKNING



GÖTEBORGS
UNIVERSITET

SPRÅKBANKENTEXT



Annotation Management Tool: a Requirement for Corpus Construction

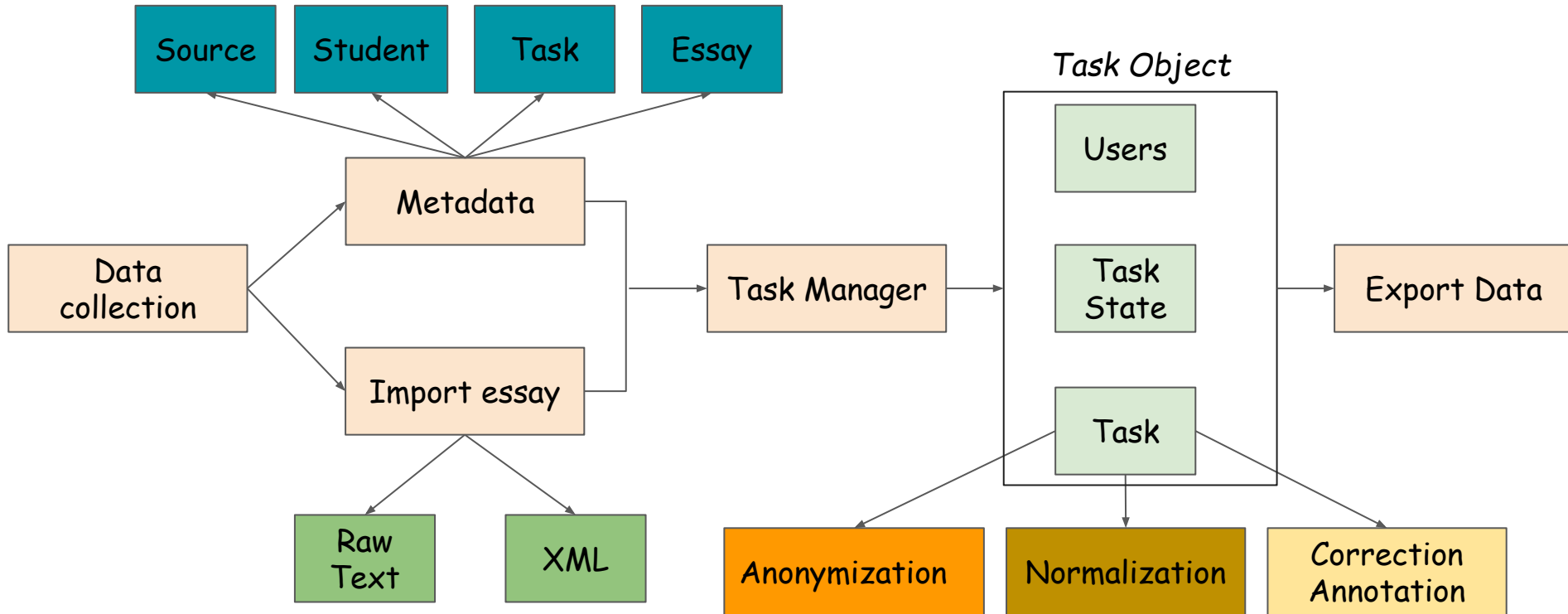
Yousuf Ali Mohammed, Arild Matsson, Elena Volodina

SweLL Infrastructure

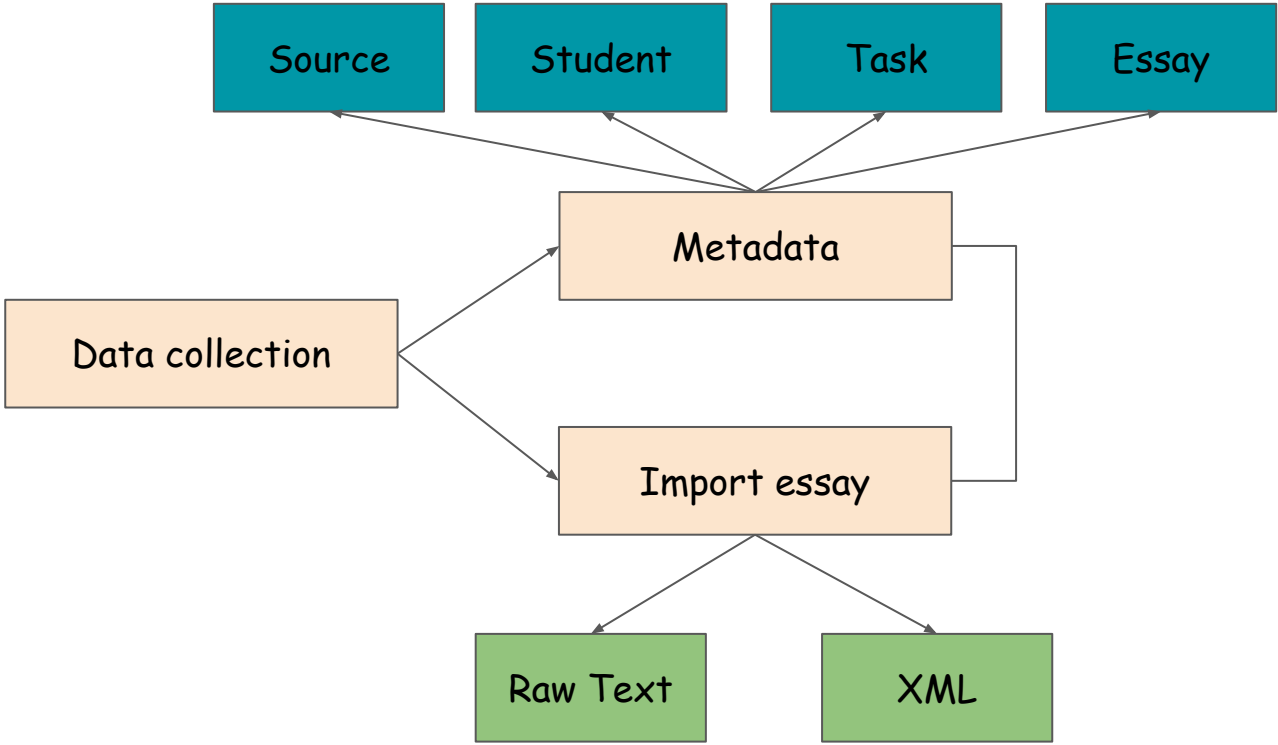
Swedish Learner Language

<https://spraakbanken.gu.se/en/projects/swell>

Data collection and Data management



Data collection



Metadata

Source

- Source ID
- Name
- Description

Student

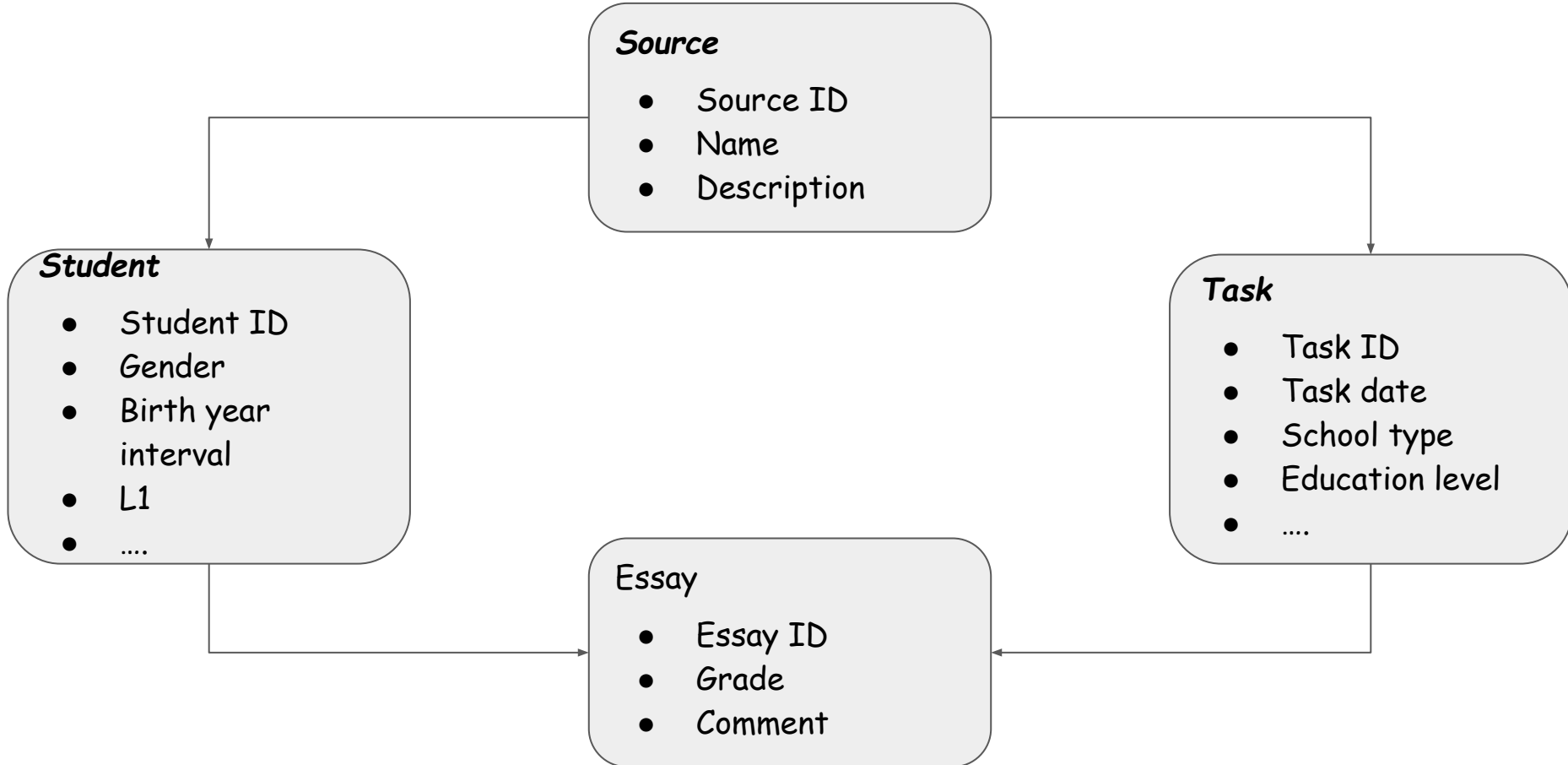
- Student ID
- Gender
- Birth year interval
- L1
- ...

Task

- Task ID
- Task date
- School type
- Education level
- ...

Essay

- Essay ID
- Grade
- Comment



Data Import

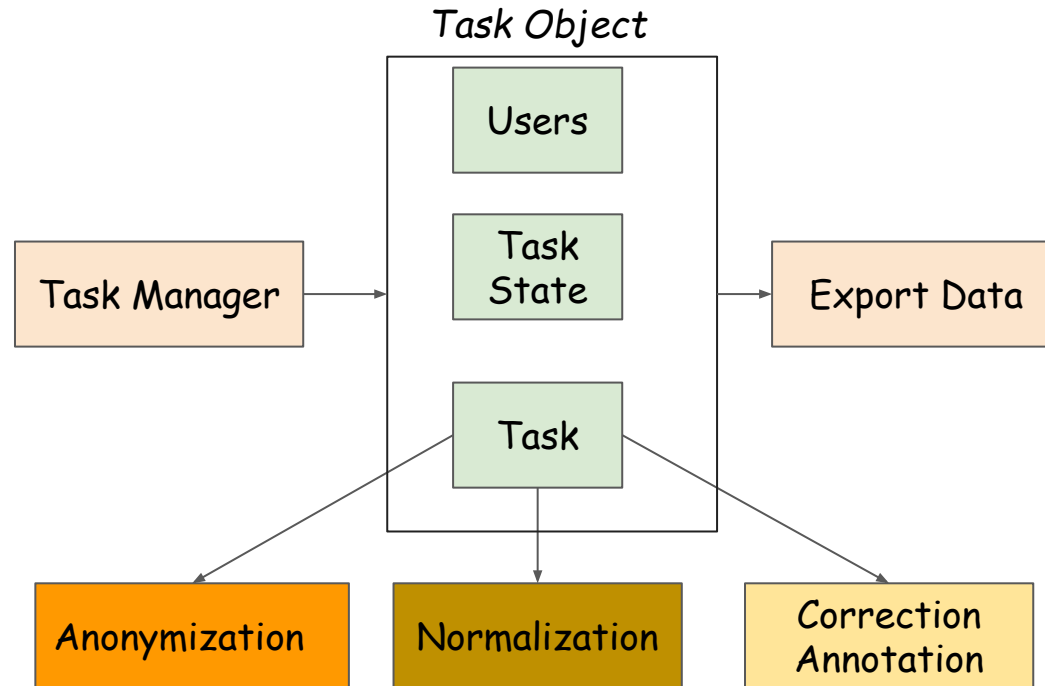
Raw text

Inför gratis körkortsutbildning. Körkort är jätteviktigt för alla människor. Min åsikt är: Inför gratis körkortsutbildning så att alla människor kan lära sig att köra bil. För det första behöver man körkort när man söker jobb därför att många företag frågar "har du körkort?", och då kan man säga "JA, jag har körkort." För det andra tycker jag att körkort gör det lättare att åka någonstans, t.ex. för att handla eller köpa något man behöver. Jag förstår att köra bil inte är bra för miljön, men man behöver inte köra bil varje dag. Man kör bara bil när man vill eller åker långt bort. För det tredje sparar man pengar därför att man betalar mycket pengar när man går och lär sig i körskolan. Mitt förslag är att kommunen ska göra något. Inför gratis körkortsutbildning nu, för alla som är intresserade av att lära sig att köra bil! Viktigt för livet.

XML

```
<?xml version="1.0" encoding="utf-8" ?>
<swell>
  <meta>
    <student id="A1">
      <language situation="Modersmål">Norska</language>
    </student>
    <task id="AT1" grading_scale="CEFR" />
  </meta>
  <essays>
    <essay student="A1" task="AT1" grade="B2">
      <content type="svala">
        {
          "source": [
            {"id": "s0", "text": "Hello "},
            {"id": "s1", "text": "world "}
          ],
          "target": [
            {"id": "t0", "text": "Hello "},
            {"id": "t1", "text": "world "}
          ],
          "edges": {
            "e-s0-t0": {
              "id": "e-s0-t0",
              "ids": ["s0", "t0"],
              "labels": [],
              "manual": false
            },
            "e-s1-t1": {
              "id": "e-s1-t1",
              "ids": ["s1", "t1"],
              "labels": [],
              "manual": false
            }
          }
        }
      </content>
    </essay>
  </essays>
</swell>
```

Data management



Export dataset

Task type

- Anonymized
- Normalized
- Correction anno

School

- A
- B
- C

L1

- English
- Spanish
- Arabic

Data type

- Source
- Target
- Both

Status

- Complete
- All

File format

- XML
- JSON
- Text



RIKSBANKENS
JUBILEUMSFOND

STIFTELSEN FÖR HUMANISTISK OCH
SAMHÄLLSVETENSKAPLIG FORSKNING



GÖTEBORGS
UNIVERSITET

SPRÅKBANKENTEXT



Thank you!
Comments? Questions?