

---

# The process of creating an in-house interface for converting Blaise output into DDI metadata

Alexandre Mairot  
EDDI19- 11th Annual European DDI Users Conference  
Tampere, Finland  
December, 3 - December, 4 2019

---

# ELIPSS : a longitudinal survey instrument

- 70 surveys between 2013 and 2019 (one monthly survey of 25-30 minutes)
- The implementation of the questionnaires is based on BLAISE 4.8 associated with add-on built by the CDSP IT Team to fix the complicated instrument specifications asked by the researchers like a video player, a autocompletion function for a long list of values
- The survey datafiles are enhanced with background and weight variables. The data are documented for the researchers community with the regular tools of the CDSP
  - DDI standard for the metadata management
  - mainly SPSS for the data management

---

# The way of the variables' documentation

It is inspired by the recommendations of the [ICPSR](#) and the practice of the ESS Team (for the missing values)

## Variable ea18\_B5B\_rec: Type de temps partiel (recodage)

### LITERAL QUESTION

Quel est le type de ce temps partiel?

| Values | Categories                         | N    |       |
|--------|------------------------------------|------|-------|
| 1      | (80% et plus)                      | 95   | 43.2% |
| 3      | Entre 50 et 80%                    | 57   | 25.9% |
| 4      | Mi-temps (50%)                     | 35   | 15.9% |
| 5      | Moins d'un mi-temps (moins de 50%) | 33   | 15.0% |
| 6      | Non concerné                       | 2149 |       |
| 7      | Refuse de répondre                 | 1    |       |

### SUMMARY STATISTICS

Valid cases 220  
Missing cases 2150  
This variable is numeric

### UNIVERSE

Si l'interviewé occupe un emploi à temps partiel (c'est-à-dire (B5A==2))

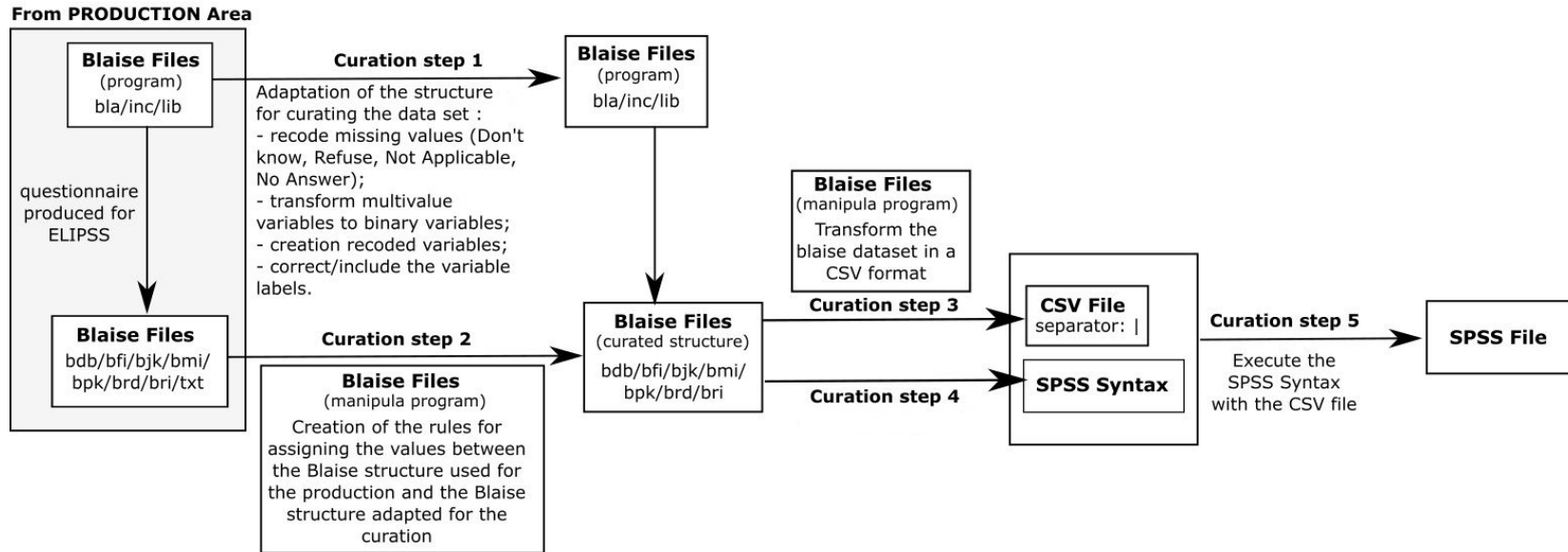
### NOTES

Variable recodée à partir des réponses à la question B5B (Quel est le type de ce temps partiel ? ; universe=(B5A==2)). La modalité [80 %] et [Plus de 80 %] sont regroupées car les effectifs de la modalité [Plus de 80 %] sont faibles.

Formule :

```
if (B5B==2) then B5B_rec=1
else B5B_rec=b5b
endif
```

# The Blaise data (curation) processing



All the curation processing is released on GitHub : <https://github.com/alexandremainrot/dime-shs-elipss>

---

# The variable declaration in **Blaise**



- 2 languages :
  - one for the questionnaire (literal question)
  - one for the documentation (variable label)
- as far as possible use library for the variable answer items and declare all the item values

## FIELDS

b5b "Quel est le type de ce temps partiel?"  
"Type de temps partiel" : qi\_b5b, RF

## RULES

IF b5a = part THEN b5b ENDIF

## TYPE

qi\_b5b = (p8p (1) "Plus de 80 %" "Plus de 80 %",  
qvp (2) "80 % (quatre cinquièmes)" "80 % (quatre cinquièmes)",  
cqp (3) "Entre 50 et 80 %" "Entre 50 et 80 %",  
mit (4) "50 % (mi-temps)" "50 % (mi-temps)",  
less (5) "Moins de 50 % (moins d'un mi-temps)" "Moins de 50 % (moins d'un mi-temps)",  
nc (6) "" "Non concerné",  
rr (7) "" "Refuse de répondre")

---

# Curation processing as the source of metadata

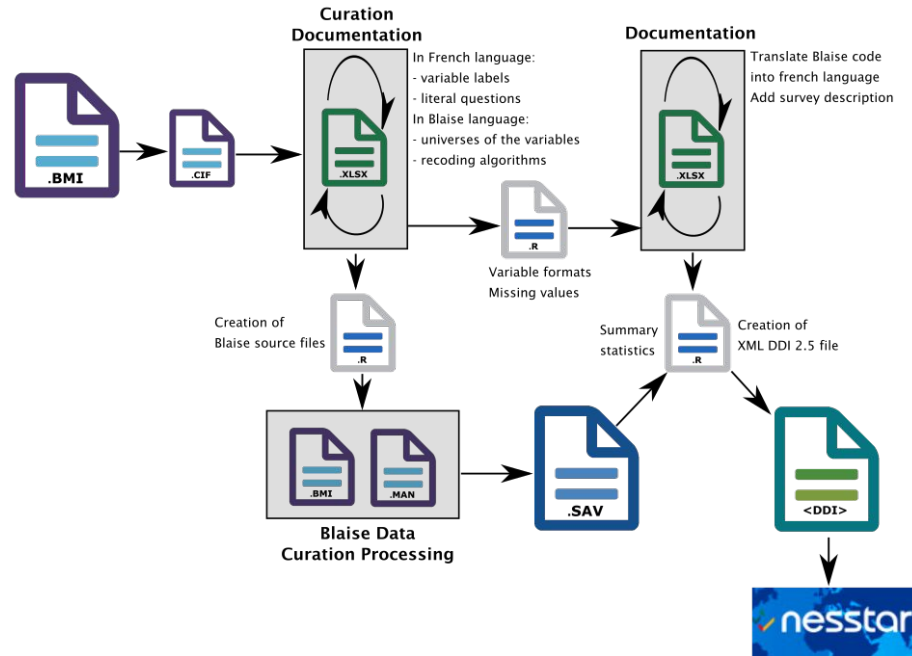
The curation process-chain is based on in-house Blaise Cameleon scripts for generating CSV files of metadata from the BMI survey files :

- One for all the metadata linked with the questions : the name of the variable, the literal question associated, the label of the variable, the format of the data, the options of missing values (DK, RF, EMPTY)
- One for all the questions items included the variable name, the item values, the items labels

The files are imported in a XLS file template to give a file associated to each survey. To complete the file, we read and copy the RULES declared/included in the BMI survey files for each variable.

---

# From the curation to the documentation



---

# An example: the weekly work time (1)

Blaise source

## FIELDS

b5b "Quel est le type de ce temps partiel?"

"Type de temps partiel" :

qi\_b5b, RF

## RULES

IF b5a = part THEN b5b

ENDIF

| XLSX Curation - Documentation |                                       |
|-------------------------------|---------------------------------------|
| variableSourcePath            | b5b                                   |
| variableLabel                 | Type de temps partiel                 |
| literalQuestion               | Quel est le type de ce temps partiel? |
| dataType                      | 1                                     |
| dataItems                     | qi_b5b                                |
| isRf                          | 1                                     |
| universe                      | b5a = part                            |

| XLSX Documentation     |  |
|------------------------|--|
| variableName           | b5b  |
| variableLabel          | Type de temps partiel                            |
| literalQuestion        | Quel est le type de ce temps partiel?            |
| universeLabel          | Si l'interviewé occupe un emploi à temps partiel |
| universeGenerationCode | (B5A==2)   |
| minusInvalrng          | 6  |



---

## An example: the weekly work time (2)

```
<codeBook ID="ddi.b5b" xml-lang="fr" version="2.5" xsi:schemaLocation="ddi:codebook:2_5
https://www.ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/codebook.xsd"
xmlns="ddi:codebook:2_5" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <var ID="V23" name="b5b" wgt="not-wgt" intrvl="discrete">
    <labl>Type de temps partiel</labl>
    <qstn>
      <qstnLit>Quel est le type de ce temps partiel?</qstnLit>
    </qstn>
    <valrng>
      <range UNITS="INT" maxExclusive="6"/>
    </valrng>
    <invalrng>
      <range UNITS="INT" min="6"/>
    </invalrng>
    <universe>Si l'interviewé occupe un emploi à temps partiel (c'est-à-dire (B5A==2))</universe>
  </var>
</codeBook>
```

---

# To move forward

- Document of the importation and treatment R functions for Blaise sources
- Review the R procedures for generating the DDI files => shifting of XML library to xml2
- Refresh the structure of the survey description in the excel files to improve reuse potential
- Release a final version on GitHub in the first semester of 2020  
(GitHub repository : <https://github.com/alexandremairot/xml-ddi-management>)

---

# Questions



[alexandre.mairot@sciencespo.fr](mailto:alexandre.mairot@sciencespo.fr)



@alexandremairot  
@cdsp\_ScPo



<https://github.com/alexandremairot>