

Experimenting with DDI-L at the French Center of Socio-Political Data

*EDDI15 – 7th Annual European DDI User Conference
December, 2 2015, Copenhagen*

Alina Danciu, Simon Le Corgne, Alexandre Mairot

Institution

- The Center of Socio-Political Data (CDSP)
 - ✓ The Réseau Quetelet partner, the French portal for data in humanities and social sciences

- The CDSP started disseminating surveys in 2005
 - ✓ About 220 data files are documented and disseminated

Summary

- Definition of a DDI-L profile
 - ✓ Our DDI-C template
 - ✓ Mapping a DDI-C template into a DDI-L template
 - ✓ XSLT to create a DDI-L profile

- A first version of a metadata system
 - ✓ Tools
 - ✓ Logical data model
 - ✓ Programs, interfaces and procedures

Our first steps towards a new metadata system

- The datasets are disseminated in **DDI-C** on a Nesstar server;
- Today, we are building a **DDI-L compliant metadata system**:
 - ✓ DDI-L profile defined
 - ✓ First version of our metadata system built
 - ✓ First programs and procedures written

DDI-L Profile: our DDI-C template

```
codebook
+ docDscr
  + citation
    + titlStmt
+ stdyDscr
  + citation
    + titlStmt
    + rspStmt
    + prodStmt
    + distStmt
    + serStmt
    + verStmt
    + biblCit
  + stdyInfo
    + subject
    + abstract
    + sumDscr
+ method
  + dataColl
  + notes
+ dataAccs
  + setAvail
  + useStmt
+ othrStdyMat
  + relMat
  + relStdy
  + relPubl
  + othRefs
```

```
+ fileDscr
  + fileTxt
    + fileName
    + fileCont
    + dimensns
    + fileType
    + filePlace
    + dataChck
    + ProcStat
    + dataMsng
    + verStmt
+ dataDscr
  + varGrp
    + labl
    + txt
    + defntn
    + universe
    + notes
  + var
    + location
    + labl
    + qustn
    + valrng
    + invalrng
    + universe
    + TotlResp
    + sumStat
    + txt
    + catgry
    + concept
    + varFormat
    + notes
```

Mapping DDI-C into DDI-L: DDI-L template

- Mapping of DDI-C elements into DDI-L using the documentation on the DDI Alliance website
 - ✓ <http://www.ddi-alliance.org/Specification/DDI-Lifecycle/3.0/#3.0convert>
- Definition of a DDI-L template

XSLT to create a DDI-L profile

<http://www.ddialliance.org/tool/ddiprofilexslt>

ddi:DDIInstance	+	r:AnalysisUnit	+	p:PhysicalDataProduct
+ r:URN	+	r:KindOfData	+	+ r:URN
+ r:Citation	+	r:OtherMaterial	+	+ p:PhysicalStructureScheme
+ r:Title	+	+ r:URN	+	+ p:RecordLayoutScheme
+ r:SubTitle	+	+ r:TypeOfMaterial	+	pi:PhysicalInstance
+ r:AlternateTitle	+	+ r:Citation	+	+ r:URN
+ s:StudyUnit	+	+ r:ExternalURLReference	+	+ r:VersionRationale
+ r:URN	+	c:ConceptualComponent	+	+ r:Note
+ r:VersionRationale	+	+ r:URN	+	+ r:Citation
+ r:Note	+	+ c:UniverseScheme	+	+ pi:DefaultMissingValuesReference
+ r:Abstract	+	+ c:Concept	+	+ pi:GrossFileStructure
+ r:UniverseReference	+	d:DataCollection	+	+ pi:StatisticalSummary
+ r:SeriesStatement	+	+ r:URN	+	a:Archive
+ r:SeriesRepositoryLocation	+	+ d:Methodology	+	+ r:URN
+ r:SeriesName	+	+ d:CollectionEvent	+	+ a:ArchiveModuleName
+ r:SeriesDescription	+	+ d:InstrumentScheme	+	+ a:ArchiveSpecific
+ r:FundingInformation	+	+ d:ProcessingEventScheme	+	+ a:OrganizationScheme
+ r:AgencyOrganizationReference	+	+ d:QuestionScheme		
+ r:GrantNumber	+	+ d:ControlConstructScheme		
+ r:Description	+	+ d:InterviewerInstructionScheme		
+ r:Coverage	+	l:LogicalProduct		
+ r:TopicalCoverage	+	+ r:URN		
+ r:SpatialCoverage	+	+ l:DataRelationship		
+ r:TemporalCoverage	+	+ l:CategoryScheme		
	+	+ l:CodeListScheme		
	+	+ r:ManagedRepresentationScheme		
	+	+ l:VariableScheme		



Summary

- Definition of a DDI-L profile
 - ✓ Our DDI-C template
 - ✓ Mapping a DDI-C template into a DDI-L template
 - ✓ XSLT to create a DDI-L profile

- A first version of a metadata system
 - ✓ Tools
 - ✓ Logical data model
 - ✓ Programs, interfaces and procedures

The first version of the metadata system

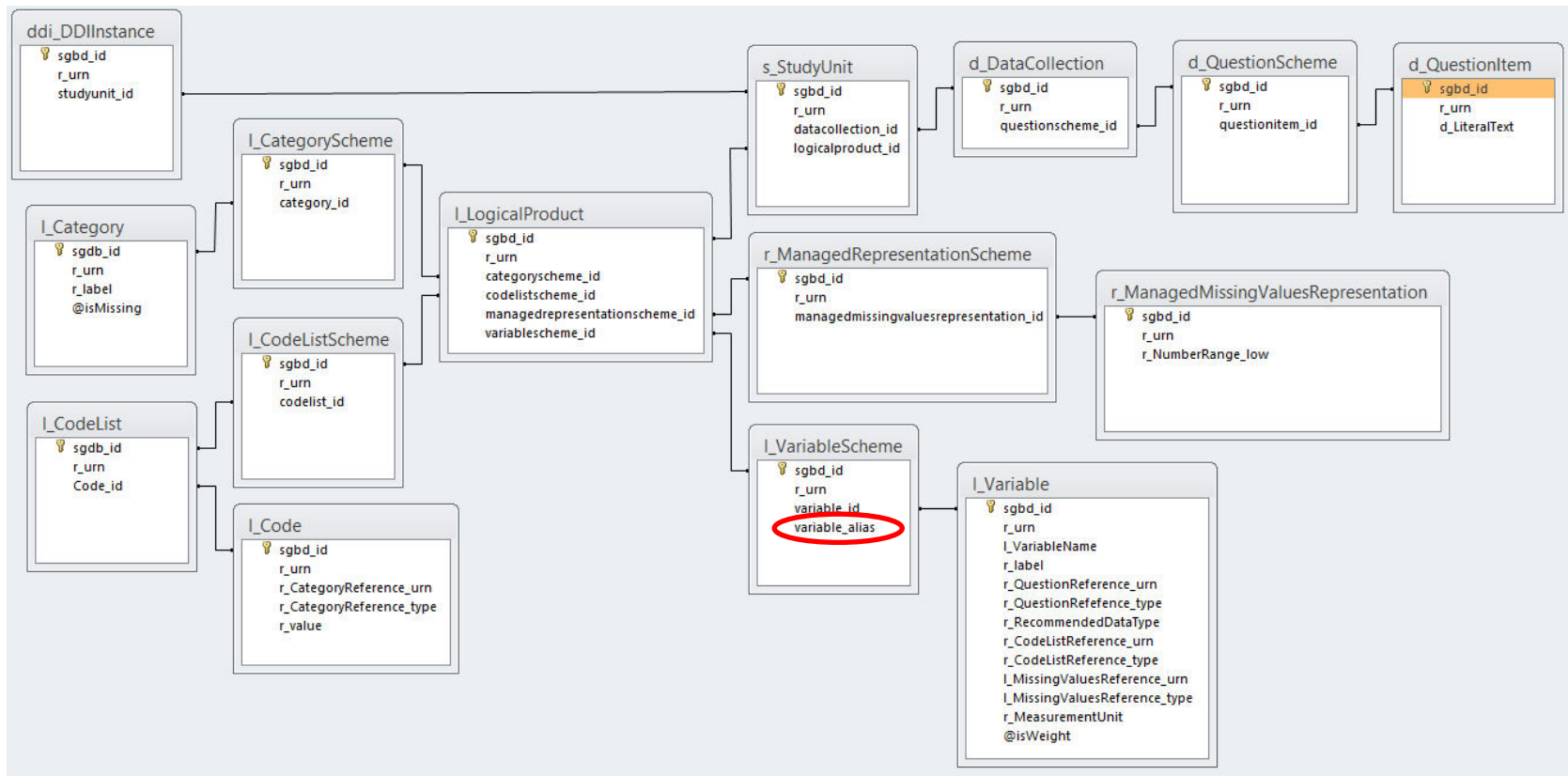
- Data as starting point: a nine-wave political study
 - ✓ Each wave has specific variables and several common variables with the other waves
 - ✓ Background variables collected during an annual study
 - ✓ Some variables based on the French national population census of 2011
 - ✓ Weighting variables
- Using the DDI-L profile to define the structure of the metadata system

Tools

- At the beginning, only 2 software:
 - ✓ MS Excel used as a documentation tool
 - ✓ Notepad++ as XML editor/validator
 - ➔ ***only DDI-C XML files generated***

- Today, new tools have been added:
 - ✓ MS Access: database structure + user interfaces
 - ✓ R: data management + computing + XML generator
 - ➔ ***DDI-C and DDI-L XML files generated***

Logical Data Model (a view)



Programs, interfaces and procedures

- **Scripts/Programs:**
 - ✓ Import Excel files into the metadata system (OK)
 - ✓ Import DDI-C files into the metadata system (in progress)
 - ✓ ...

- **Interfaces**
 - ✓ User interfaces for variables' documentation (in progress)
 - ✓ ...

- **Procedures**
 - ✓ Handling metadata duplication (in progress)
 - ✓ ...

An example: How interested in politics

- **Variable name:** INTPOL
- **Variable label:** How interested in politics
- **Literal question:** How interested would you say you are in politics ? are you?
- **Variable categories**
 - ✓ 1- Very interested
 - ✓ 2- Quite interested
 - ✓ 3- Hardly interested
 - ✓ 4- Not at all interested
 - ✓ 7- Refusal (missing value)
 - ✓ 8- Don't know (missing value)



Road map

- **In the near future:** Finalize the creation of the programs, the interfaces and the procedures
- **In the middle future:** Migrate the database and the interfaces on a web solution

<http://cdsp.sciences-po.fr>

<http://quanti.dime-shs.sciences-po.fr/>

alina.danciu@sciencespo.fr

simon.lecorgne@sciencespo.fr

alexandre.mairot@sciencespo.fr