Use of DDI at Insee

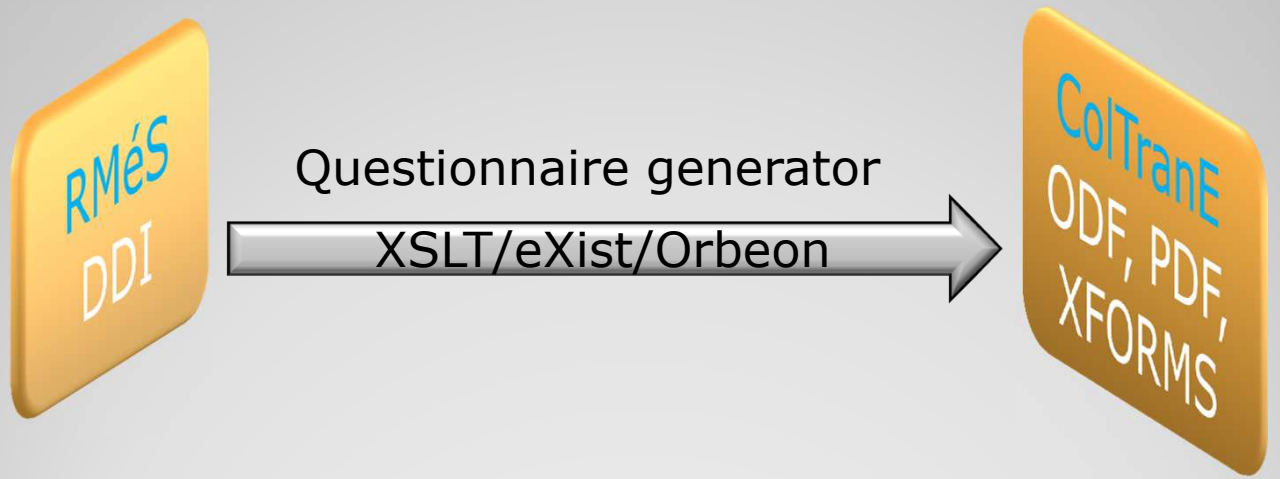
EDDI14 Conference – Guillaume Duffes - Insee

Contents

1. Brief history of DDI at Insee
2. GSIM as a conceptual model
3. Towards a Resource Oriented Architecture (ROA)
4. IT tools
5. Next steps

Brief history of DDI at Insee

- Two major projects involved in the implementation of DDI:
  - ColTranE (transverse data collection for business surveys)
  - RMéS (Statistical Metadata Repository)



Brief history of DDI at Insee

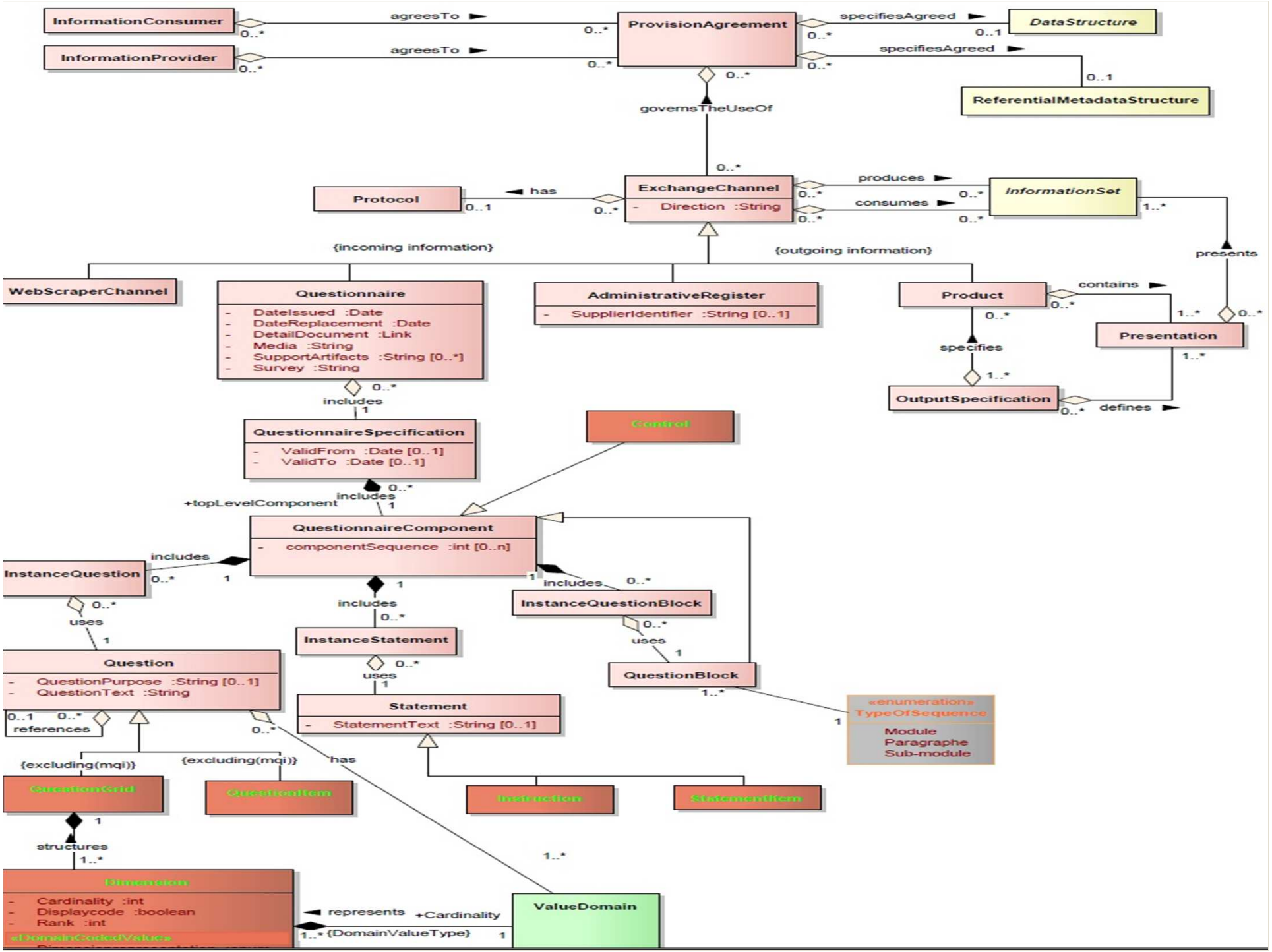
- February 2012: RMéS was a bit lost, too many metadata scattered all over, too many standards, didn't know where to start.
- ColTranE needed to put online its SBS survey questionnaires
- Both projects jumped in. DDI 3.1 was chosen. PDF fillable forms generated.

Brief history of DDI at Insee

- 01/2014: Decision was taken to move to DDI 3.2 to enhance the representation of grids, inter alia.
- First, the questionnaires for the CIS survey, then the Labour Cost Survey.
- XForms was the output format.
- The questionnaire generator is compliant with 3.2.
- Work in progress: Questionnaire for the online Labour Force Survey

GSIM as a conceptual model

- The Generic Statistical Information Model chosen as the reference conceptual information model at Insee.
- GSIM is a reference framework of information objects.

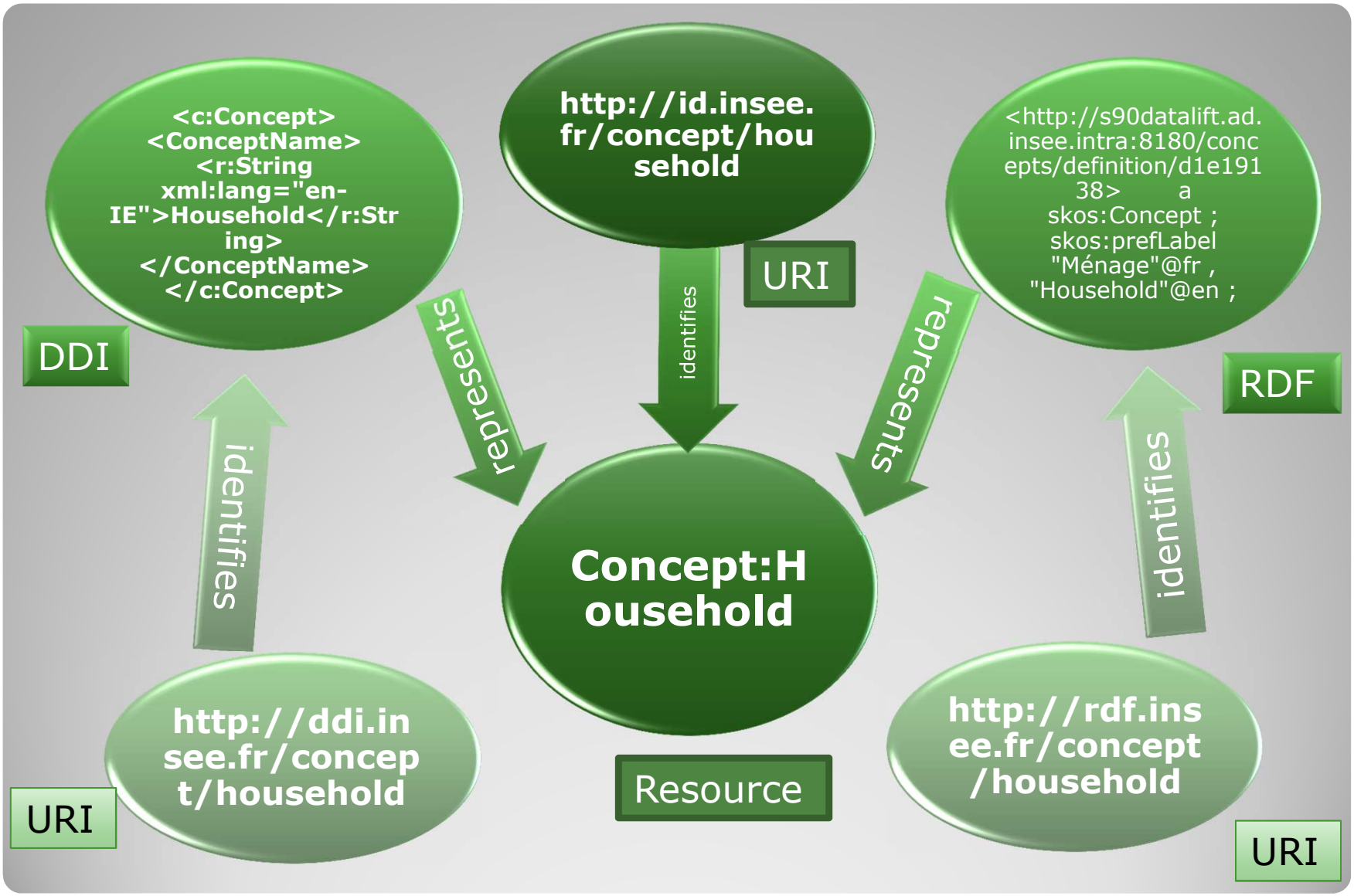

InformationConsumer
InformationProvider
agreesTo
agreesTo
ProvisionAgreement
specifiesAgreed
DataStructure
specifiesAgreed
ReferentialMetadataStructure
governsTheUseOf
Protocol
has
ExchangeChannel
- Direction :String
produces
consumes
InformationSet
presents
{incoming information}
{outgoing information}
WebScraperChannel
Questionnaire
- DateIssued :Date
- DateReplacement :Date
- DetailDocument :Link
- Media :String
- SupportArtifacts :String [0..*]
- Survey :String
AdministrativeRegister
- SupplierIdentifier :String [0..1]
Product
contains
Presentation
specifies
OutputSpecification
defines
includes
QuestionnaireSpecification
- ValidFrom :Date [0..1]
- ValidTo :Date [0..1]
Control
+topLevelComponent
includes
QuestionnaireComponent
- componentSequence :int [0..n]
InstanceQuestion
includes
InstanceQuestionBlock
uses
InstanceStatement
uses
QuestionBlock
Question
- QuestionPurpose :String [0..1]
- QuestionText :String
references
Statement
- StatementText :String [0..1]
«enumeration»
TypeOfSequence
Module
Paragraphe
Sub-module
{excluding(mqi)}
{excluding(mqi)}
has
QuestionGrid
QuestionItem
Instruction
StatementItem
structures
Dimension
- Cardinality :int
- Displaycode :boolean
- Rank :int
«DomainCodedValue»
represents
+Cardinality
{DomainValueType}
ValueDomain

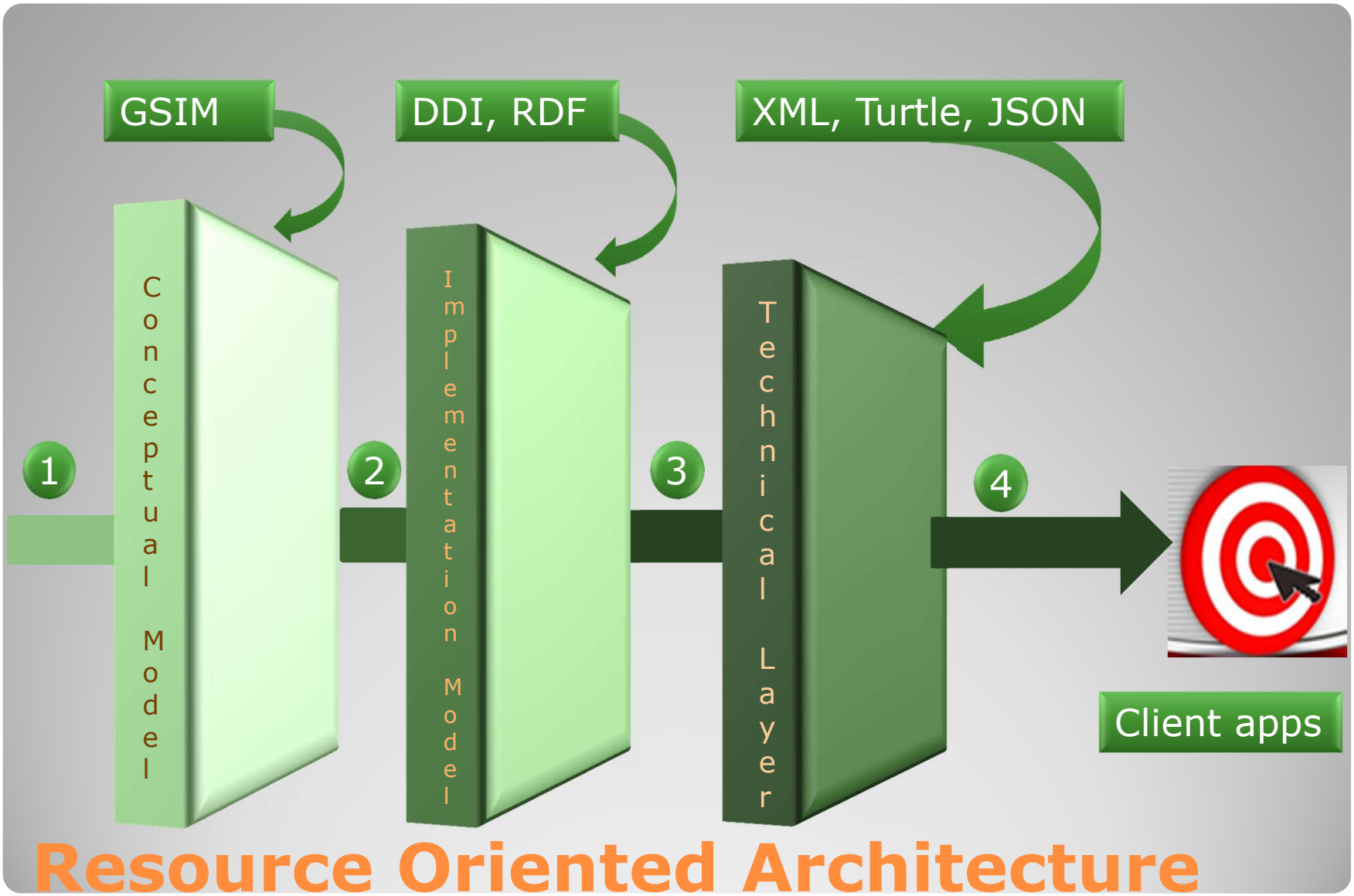
- RMéS is appointed to build a central, authoritative and GSIM-based Metadata Repository.
- Not an easy task to come to an agreement with stakeholders
- The implementation is also discussed with third-parties (e.g major and a minor versions of a concept)

GSIM as a conceptual model

Resource Oriented Architecture

- A Resource Oriented Architecture is based on three main features:
  - Uniform Resource Identifier (URI). Standard Structure to be used at Insee: http://id.insee.fr/geo/commune/75056
  - ReST (Representational State Transfer) Web Services.
  - Informational Resources (or representations) and non-informational Resources (or simply resource)

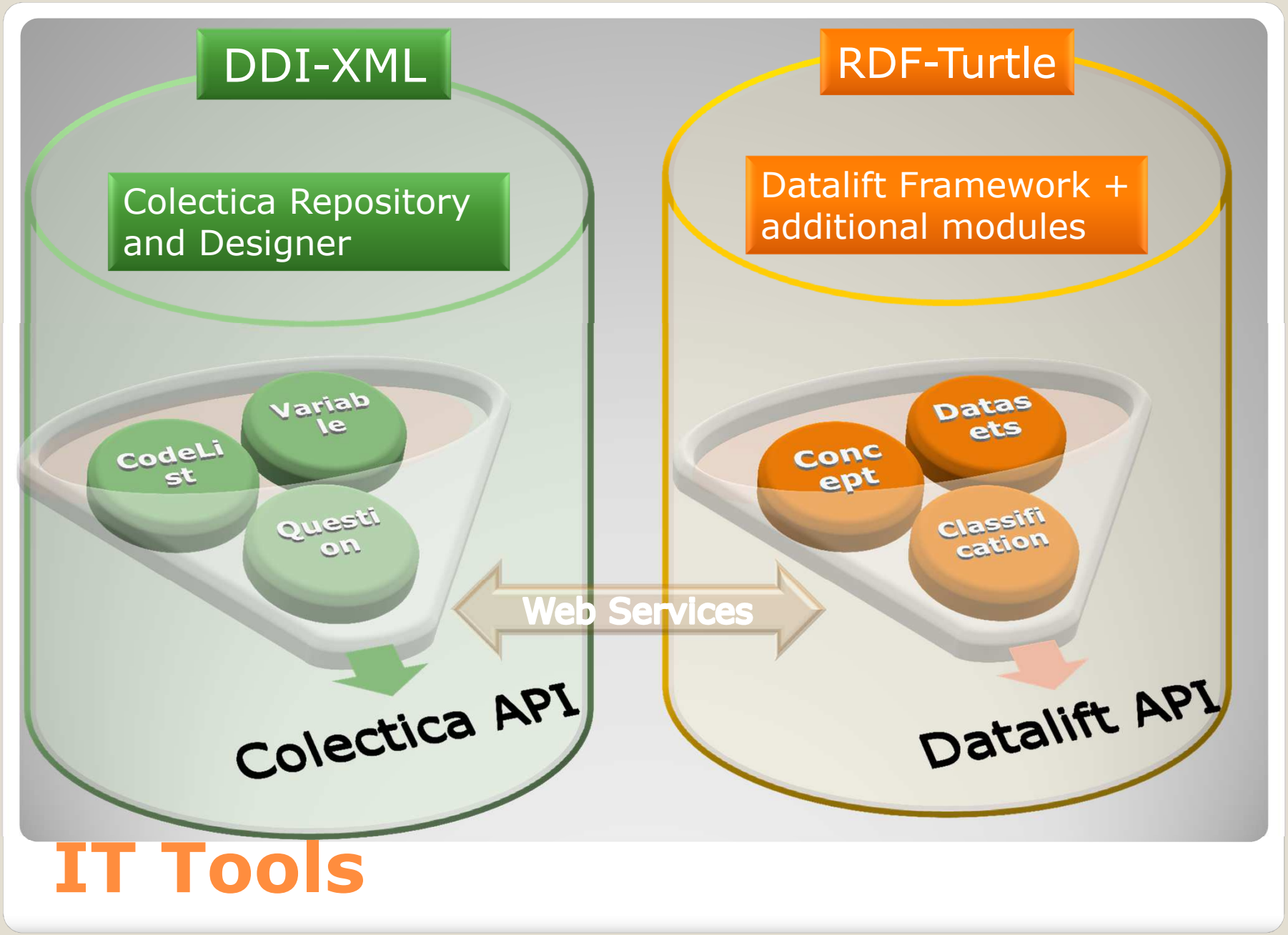

<c:Concept>
<ConceptName>
<r:String
xml:lang="en-
IE">Household</r:Str
ing>
</ConceptName>
</c:Concept>
DDI
http://id.insee.
fr/concept/hou
sehold
URI
<http://s90datalift.ad.
insee.intra:8180/conc
epts/definition/d1e191
38> a
skos:Concept ;
skos:prefLabel
"Ménage"@fr ,
"Household"@en ;
RDF
represents
identifies
represents
identifies
identifies
Concept:H
ousehold
Resource
http://ddi.in
see.fr/concep
t/household
URI
http://rdf.ins
ee.fr/concept
/household
URI



Resource Oriented Architecture

- 1 UML. Expand if necessary for addressing business requirements. A URI per resource.
- 2 DDI profile that can relate to other standards (SDMX, RDF). A URI per representation of a resource.
- 3 URI Dereferencing of resources via HTTP methods
- 4 XML or JSON returned by the HTTP based RESTful APIs

Resource Oriented Architecture


DDI-XML
Colectica Repository and Designer
CodeList
Variable
Question
Colectica API
RDF-Turtle
Datalift Framework + additional modules
Concept
Datasets
Classification
Datalift API
Web Services

IT Tools

- GUI for the management of objects:
  - DDI: Colectica Designer is going to be tested
  - RDF: VocBench (FAO), VITRO/VIVO (Cornell).
- Communication channels between tools still to be tackled:
  - Which syntax and rules for the Web Services?
  - Can the RDF triplestore support distributed queries?
  - Security requirements and restrictions, etc.

IT Tools

The future

- Insee pins high hopes on **DDI 4**:
  - DDI 4 and its **RDF serialisation** could be combined easily with existing RDF metadata
  - Thus, a full RDF-based architecture
  - DDI 4 would be the core implementation model completed by other RDF vocabularies (skos, xkos, dcterms, foaf, etc.).
- DDI 4 gives rise to some expectations:
  - To be aligned as much as possible with GSIM
  - **Fully** backward-compatible with 3.2
  - Commercial DDI-based platforms integrate RDF schema-based repository and querying facility?

Questions?