



Semantic Lifting of Budget Data

Semantic Data Web Technologies Lab

Mentor:

Fathoni Musyaffa

Group members:

Florian Weile (L)

Tatiana Novikova

Aberham Gebreyohannes

Samuel Y. Ayele

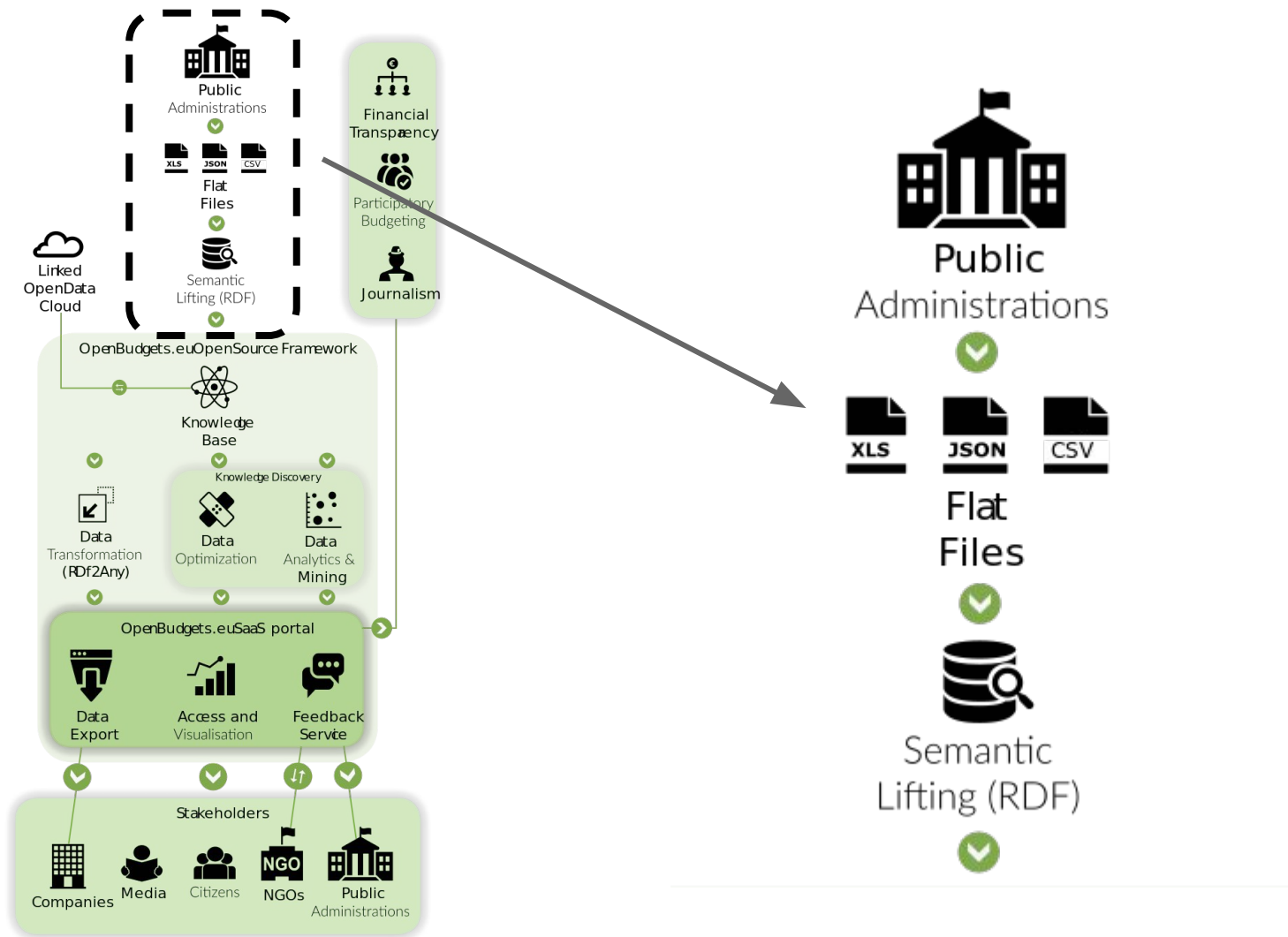
Outline

1. Motivation
2. Use Cases
3. ABuDaT
 - 3.1. System Architecture
 - 3.2. Facts about ABuDaT
 - 3.3. Demo
 - 3.4. Responsibilities
4. Challenges & Limitations
5. Conclusion

1.

Motivation

OpenBudgets.eu Overview



Motivation

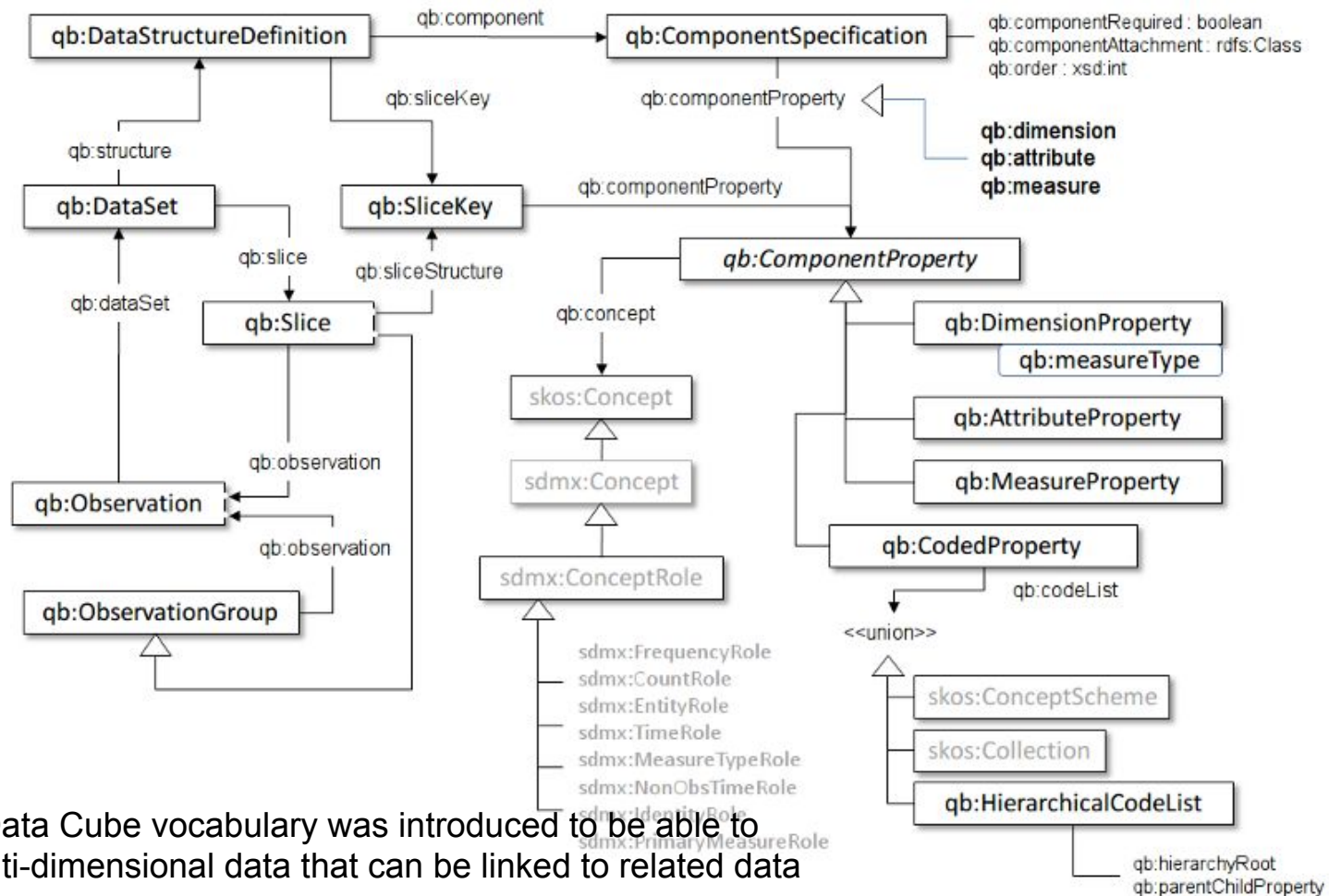
Definition: **Semantic Lifting**

Budget and transaction data published by public entities do not adhere to the schema as developed in OpenBudgets. **Semantic lifting** will transform the raw data into a well defined format.

Motivation (cont.)

- **Governments publish fiscal data**
 - Heterogeneous datasets
 - Various formats
 - Not comparable
- **OpenBudgets.eu**
 - Aims at fiscal transparency
 - Semantic data model based on the [RDF Data Cube](#)
 - **Goal:** Analysis and comparison across government levels, regions, and countries

RDF Data Cube



The RDF Data Cube vocabulary was introduced to be able to publish multi-dimensional data that can be linked to related data sets.

<https://www.w3.org/TR/vocab-data-cube/>

Motivation (cont.)

Goal of the Lab:

Make fiscal data transformation easy for users without knowledge of

- RDF
- SPARQL
- ETL

2.

Use Cases

Fiscal Data Transformation

EJERCICIO	CENTRO GESTOR	FUNCIONAL	ECONOMICA	FINANCIACION	DESCRIPCION	IMPORTE
2012	1010	1111	100000	91002	Retribuciones básicas de Altos Cargos	2414299,58
2012	1010	1111	100001	91002	Otras remuneraciones de Altos Cargos	261643,9
2012	1010	1111	110000	91002	Retribuciones básicas Pers. Eventual de Gabinete	145184,76
2012	1010	1111	110001	91002	Otras remuneraciones de Pers. Eventual de Gabinete	167792,52
2012	1010	1111	120000	91002	Sueldos de Personal Funcionario	1437683,04
2012	1010	1111	120005	91002	Trienios de Personal Funcionario	567097,34
2012	1010	1111	120006	91002	Pagas extraordinarias de Personal Funcionario	527906,5
2012	1010	1111	121000	91002	Complemento de destino de Personal Funcionario	648462,6
2012	1010	1111	121001	91002	Complemento específico de Personal Funcionario	587647,56
2012	1010	1111	122002	91002	Otras retribuciones en especie	130000
2012	1010	1111	130000	91002	Salario base de Personal Laboral Fijo	96070,8
2012	1010	1111	130001	91002	Antigüedad de Personal Laboral Fijo	37800,6
2012	1010	1111	130002	91002	Pagas extraordinarias de Personal Laboral Fijo	30109,02

@prefix qb: <http://purl.org/linked-data/cube#> .

@prefix aragon-dimension: <http://data.openbudgets.eu/ontology/dsd/aragon-budget-exp-2012/dimension> .

<http://data.openbudgets.eu/ontology/dsd/aragon-budget-exp-2012/observation/1> a qb:Observation ;
aragon-dimension:administrativeClassification

<http://data.openbudgets.eu/resource/codelist/estructura_organica_aragon_2012/01010> ;
aragon-dimension:economicClassification

<http://data.openbudgets.eu/resource/codelist/estructura_economica_aragon_2012/100000> ;
aragon-dimension:functionalClassification

<http://data.openbudgets.eu/resource/codelist/estructura_funcional_aragon_2012/1111> ;
aragon-dimension:fundingClassification

<http://data.openbudgets.eu/resource/codelist/estructura_financiacion_aragon_2012/91002> ;
obeu-measure:amount 2414299.58 ;

qb:dataSet <http://data.openbudgets.eu/resource/datasets/aragon-expenditure-2012> .

Codelist Transformation

EJERCICIO	TYPE	CAPITULO	ARTICULO	CONCEPTO	SUBCONC	DESCRIPCION CORTA	DESCRIPCION LARGA
2012	G	1	10	100	100000	Retr.básicas de Alt.Cargo	Retribuciones básicas de Altos Cargos
2012	G	1	10	100	100001	Otras remun. de Alt.Cargo	Otras remuneraciones de Altos Cargos
2012	G	1	11	110	110000	Retr.básicas de P.Ev.Gab.	Retribuciones básicas de Personal Eventual de Gabinete
2012	G	1	11	110	110001	Otras remun. de P.Ev.Gab.	Otras remuneraciones de Personal Eventual de Gabinete
2012	G	1	12	120	120000	Sueldos de Funcionarios	Sueldos de Personal Funcionario
2012	G	1	12	120	120005	Trienios de Funcionarios	Trienios de Personal Funcionario
2012	G	1	12	120	120006	Pagas extras Funcionarios	Pagas extraordinarias de Personal Funcionario
2012	G	1	12	121	121000	C. destino Funcionarios	Complemento de destino de Personal Funcionario
2012	G	1	12	121	121001	C. especific. Funcionarios	Complemento específico de Personal Funcionario
2012	G	1	12	121	121008	C. progr. esp. Justicia	Complementos para programas concretos de actuación en el ámbito de Justicia
2012	G	1	12	121	121009	Otras retr. compl. Func.	Otras retribuciones complementarias de Personal Funcionario

@prefix skos-core: <<http://www.w3.org/2004/02/skos/core#>>

<http://data.openbudgets.eu/resource/codelist/estructura_economica_aragon_2012/100000> a
skos-core:Concept ;

skos-core:prefLabel "Retribuciones básicas de Altos Cargos"@es ;

skos-core:altLabel "Retr.básicas de Alt.Cargo"@es ;

skos-core:notation "100000" ;

skos-core:inScheme

<http://data.openbudgets.eu/resource/codelist/estructura_economica_g_aragon_2012> .

[...]

Data Cube Validation


21 Integrity constraints





W3C Recommendation, January 16, 2014


Examples:

- Every **qb:Observation** has exactly one associated **qb:DataSet**.
- Every **qb:DataSet** has exactly one associated **qb:DataStructureDefinition**.
- Every **qb:Observation** has a value for each declared attribute that is marked as required.





Triple Store Upload

 Apache Jena Fuseki

  dataset  manage datasets  help

Server status: 

Dataset: /abudat_aragon_exp_2016 ▾

 query  upload files  edit  info

Available services

File Upload: http://localhost:3030/abudat_aragon_exp_2016/upload

Graph Store Protocol: http://localhost:3030/abudat_aragon_exp_2016/data

Graph Store Protocol (Read): http://localhost:3030/abudat_aragon_exp_2016/get

HTTP Quads: http://localhost:3030/abudat_aragon_exp_2016/

SPARQL Query: http://localhost:3030/abudat_aragon_exp_2016/query


SPARQL Query: http://localhost:3030/abudat_aragon_exp_2016/sparql

SPARQL Update: http://localhost:3030/abudat_aragon_exp_2016/update

Statistics

Name	Overall	Overall good	Overall bad	SPARQL Query	SPARQL Query	File Upload	Graph Store Protocol	Graph Store Protocol (Read)	SPARQL Update	HTTP Quads
/abudat_aragon_exp_2016	8	7	0	8 (0 bad)	0	0	0	0	0	0

Dataset size

Note this may be slow and impose a significant load on large datasets: 

graph name: default graph

triples: 39045

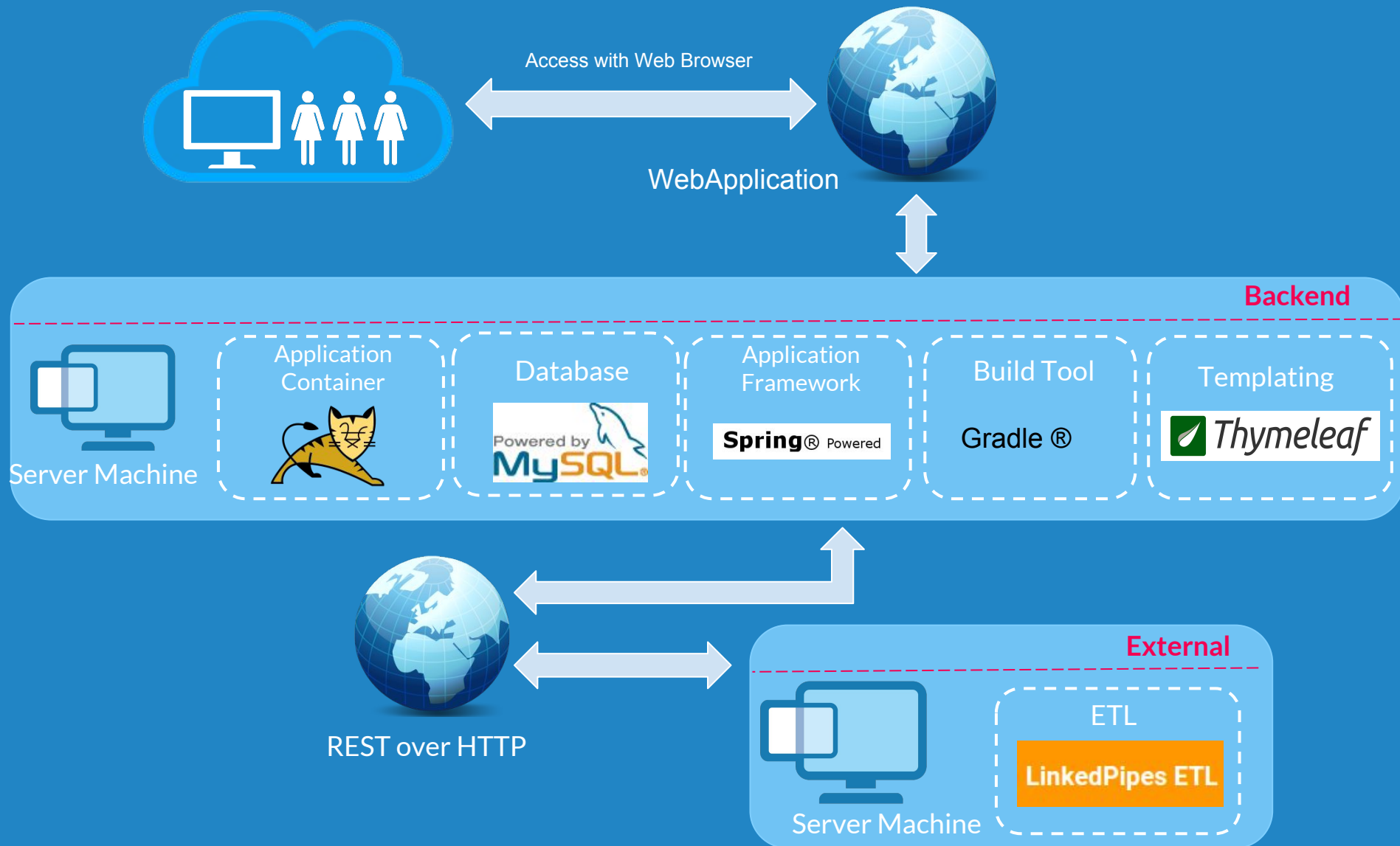
Ongoing operations

TBD. Will list any long-lasting operations that are ongoing or recently completed, e.g. backups.

3. ABuDaT

Administrative Budget Data Transformer

System architecture:



Facts about ABuDaT:

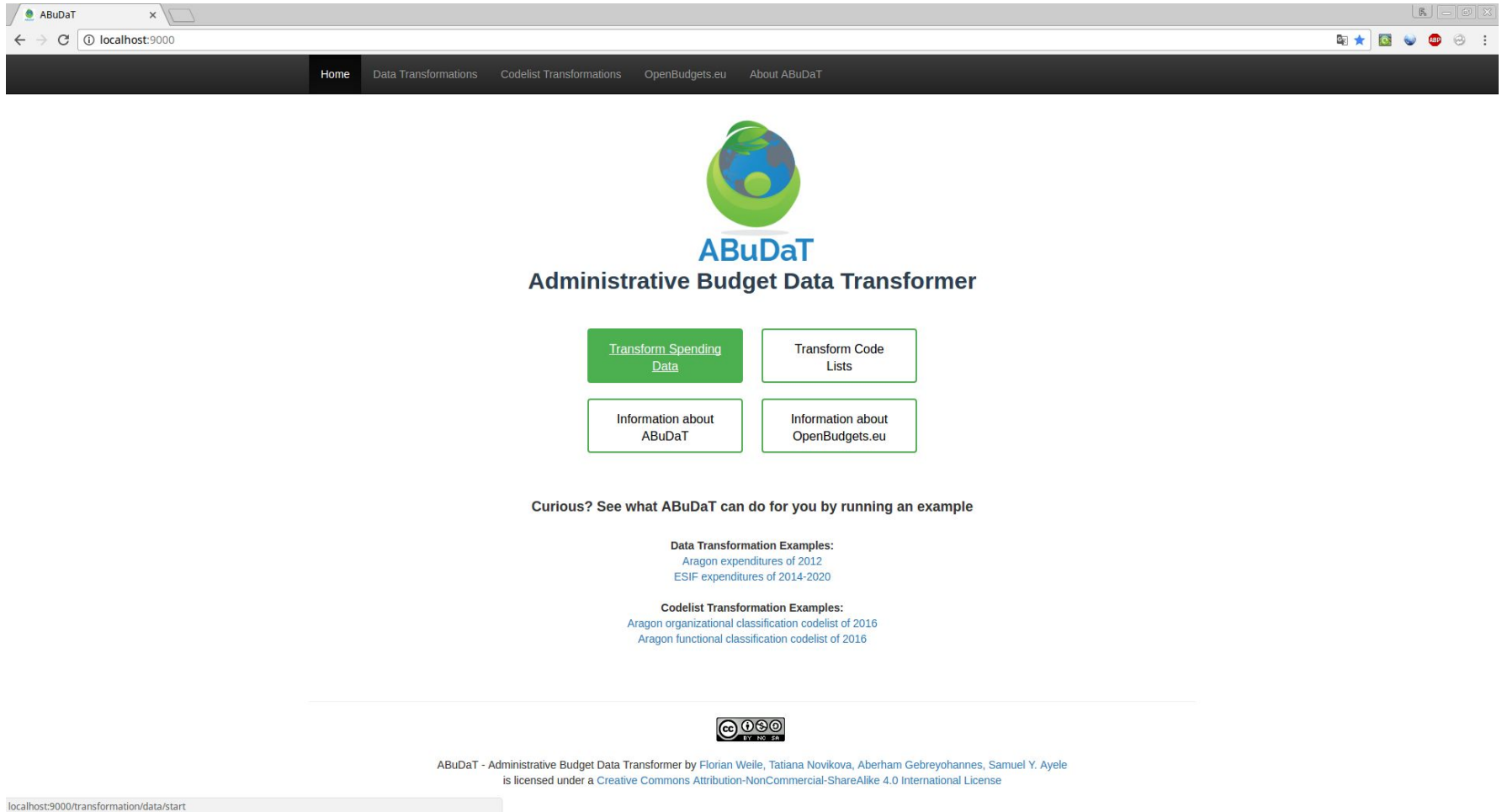
- ~ 6.5k Lines of **Production Code**;
- ~ 3.5k Lines of **Test Code**;
- ~ 4.5k Lines of **HTML**;
- Frontend tested with **Selenium**;
- Unit and Integration Tests.
 - 92% Test Coverage (LOC)

ABuDaT

ABuDaT

Element	Missed Instructions	Cov.	Missed Branches	Cov.
de.uni.bonn.iai.eis.etl.linkedpipes		89%		69%
de.uni.bonn.iai.eis.web		86%		65%
de.uni.bonn.iai.eis.web.controller		87%		42%
de.uni.bonn.iai.eis.web.model		94%		67%
de.uni.bonn.iai.eis		90%		66%
de.uni.bonn.iai.eis.rdf		84%		n/a
de.uni.bonn.iai.eis.web.controller.example		61%		n/a
de.uni.bonn.iai.eis.rdf.obeu		92%		n/a
de.uni.bonn.iai.eis.etl.component		99%		100%
de.uni.bonn.iai.eis.web.model.example		100%		n/a
de.uni.bonn.iai.eis.etl		100%		100%
de.uni.bonn.iai.eis.web.model.mapping		100%		75%
de.uni.bonn.iai.eis.rdf.obeu.classification		100%		n/a
de.uni.bonn.iai.eis.rdf.obeu.measure		100%		n/a
Total	825 of 9.889	92%	124 of 360	66%

ABuDaT Demo



The screenshot shows a web browser window with the address bar set to `localhost:9000`. The application has a dark navigation bar with the following links: [Home](#), [Data Transformations](#), [Codelist Transformations](#), [OpenBudgets.eu](#), and [About ABuDaT](#).

The main content area features the ABuDaT logo, which consists of a stylized globe with green and blue segments. Below the logo, the text "ABuDaT" is displayed in a large, bold, blue font, followed by "Administrative Budget Data Transformer" in a smaller, bold, black font.

There are four buttons arranged in a 2x2 grid:

- [Transform Spending Data](#) (green button)
- [Transform Code Lists](#) (white button with green border)
- [Information about ABuDaT](#) (white button with green border)
- [Information about OpenBudgets.eu](#) (white button with green border)

Below the buttons, there is a section titled "Curious? See what ABuDaT can do for you by running an example". This section contains two sub-sections:

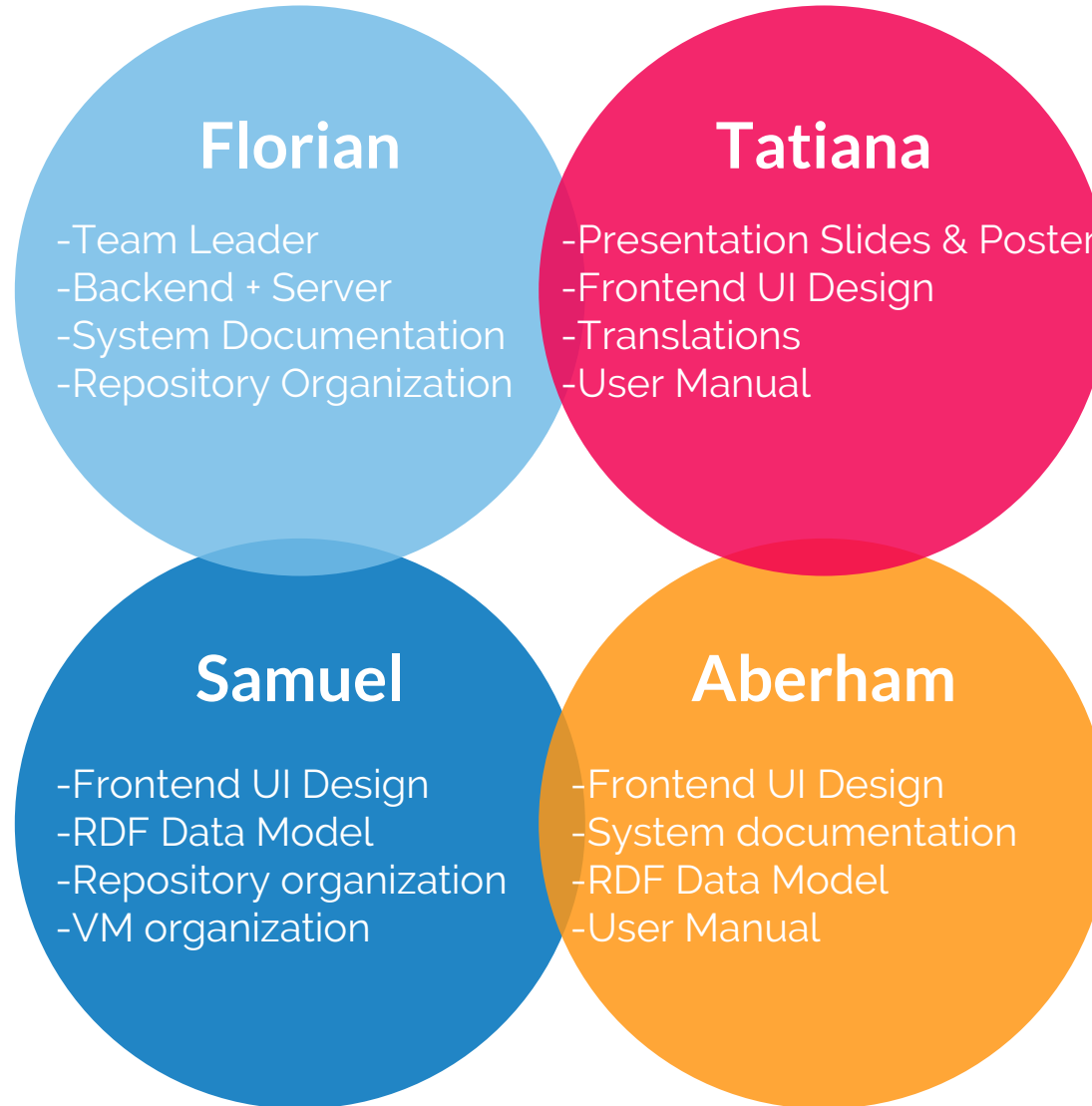
- Data Transformation Examples:**
 - [Aragon expenditures of 2012](#)
 - [ESIF expenditures of 2014-2020](#)
- Codelist Transformation Examples:**
 - [Aragon organizational classification codelist of 2016](#)
 - [Aragon functional classification codelist of 2016](#)

At the bottom of the page, there is a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License logo (CC BY NC SA) and the following text:

ABuDaT - Administrative Budget Data Transformer by Florian Weile, Tatiana Novikova, Aberham Gebreyohannes, Samuel Y. Ayelle is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](#)

The browser's address bar at the bottom shows the URL `localhost:9000/transformation/data/start`.

Team members' responsibilities:



4. Challenges & Limitations

Challenges:

- **Heterogeneous** datasets
 - Various file formats
- SPARQL based data cube validation very **time consuming**
- LinkedPipes ETL REST-API is **not convenient**.

Limitations:

- **Input Format limitations**
 - CSV and XLS supported
 - Users might prepare the data
- **No data upload support**
 - Known bug regarding upload file size
- **LinkedPipes ETL on the same machine.**
 - Security concerns
 - LinkedPipes ETL limitations

Limitations (cont.)

- Duplicate dimensions
 - No aggregation of amounts for observations with duplicate dimensions
- Validation
 - Data Cube validation only
 - No OpenBudgets.eu validation

5. Conclusion

Conclusion

ABuDaT makes fiscal data transformation
EASY.

No technical background knowledge required.

Given that the input data format meets the requirements.

Thank you!
(Any questions?:)