



**A state aid vocabulary for DG Competition to publish
transparency information as open, machine-readable
data**



Document Metadata

This study was prepared for the ISA Programme by:

PwC EU Services

Disclaimer:

The views expressed in this report are purely those of the authors and may not, in any circumstances, be interpreted as stating an official position of the European Commission.

The European Commission does not guarantee the accuracy of the information included in this study, nor does it accept any responsibility for any use thereof.

Reference herein to any specific products, specifications, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favouring by the European Commission.

All care has been taken by the author to ensure that s/he has obtained, where necessary, permission to use any parts of manuscripts including illustrations, maps, and graphs, on which intellectual property rights already exist from the titular holder(s) of such rights or from her/his or their legal representative.

Table of Contents

1. INTRODUCTION	1
1.1. CONTEXT	1
1.2. OBJECTIVES	1
1.3. STRUCTURE	2
1.4. GLOSSARY	2
2. BUSINESS CASE	3
2.1. STAKEHOLDERS	3
2.2. BUSINESS NEED	3
2.3. PROPOSED SOLUTION	4
2.4. EXPECTED BENEFITS	4
3. DEFINITION OF THE STATE AID VOCABULARY	1
3.1. STEP 1: CONTEXT AND REQUIREMENTS.....	1
3.2. STEP 2: INFORMATION MODELLING.....	2
3.3. STEP 3: BUSINESS RULES.....	5
3.4. STEP 4: SYNTAX BINDING	7
3.5. STEP 5: SYNTAX DOCUMENTING AND MAPPING	7
4. PROOF OF CONCEPT	9
5. CONCLUSIONS AND NEXT STEPS	11
REFERENCES.....	12
ANNEX I: CONCEPTUAL MAPPING BETWEEN THE CORE VOCABULARIES AND ANNEX III OF THE COMMISSION REGULATION (EC) 651/2014	1
ANNEX II: THE STATE AID TRANSPARENCY RDF VOCABULARY	1
ANNEX III: USE THE STATE AID TRANSPARENCY VOCABULARY TO PUBLISH A MOCK DATA SET AS OPEN, MACHINE-READABLE DATA.....	5

1. INTRODUCTION

This report was prepared in the context of Action 1.1 of the Interoperability Solutions for European Public Administrations (ISA) Programme [1]. It provides an overview of an RDF dissemination vocabulary for State Aid Transparency data.

1.1. Context

The Commission Regulation (EU) 651/2014 [2] requires the MSs to collect and publish on the internet, information related to beneficiaries of aid awards exceeding EUR 500.000. The data to be collected and published is known as “transparency information”.

In this context, DG Competition and the MSs are interested publishing the transparency information in an open, machine-readable format. The transparency information can be viewed as a high-value dataset for both the data providers and data consumers, as it provides details related to beneficiaries of state aid awards, which are of a high general public interest [3].

Some of the MSs have already implemented their own websites for collecting transparency information, while others are currently working on their solutions. DG Competition will have to integrate these applications with their own solution, and publish the entire set of transparency information in an open, machine-readable format.

To enable interoperability on the semantic level, common vocabularies should be used as much as possible. To this end, the ISA Core Vocabularies [4] provide simplified, reusable and extensible data models that capture the fundamental characteristics of concepts, such as persons, organisations and addresses. For example, the Core Business Vocabulary¹ describes a legal entity having a legal name, an activity, an address, etc., and the Core Location Vocabulary² describes the characteristics of a location represented as an address, a geographic name, or a geometry. The Core Vocabularies can be used and extended in order to develop an RDF vocabulary for publishing transparency information as machine-readable open data.

This will allow DG Competition and the MSs not only to comply with the requirements of the Article 9 of the Commission Regulation (EU) 651/2014 [2], but even go a step further toward increasing transparency in the domain of state aid, making the data automatically discoverable for machines.

1.2. Objectives

The objective of the current project is to build an RDF vocabulary for DG Competition that will be used for disseminating transparency information as machine-readable open data.

¹ https://joinup.ec.europa.eu/asset/core_business/description

² http://joinup.ec.europa.eu/asset/core_location/description

This vocabulary is based on the RDF syntaxes of the Core Vocabularies.

1.3. Structure

The remainder of this report is structured as follows. Chapter 2 presents a basic description of the main aspects of the project, including the business needs, its scope and objectives, and expected benefits. Chapter 3 describes the proposed RDF vocabulary developed following the methodology explained in the handbook “e-Government Core Vocabularies handbook” [5]. Chapter 4 presents a proof-of-concept that demonstrates the publication of transparency information as machine-readable open data using the developed RDF vocabulary. Finally, Chapter 5 summarises the work and discusses possible next steps.

1.4. Glossary

The table below defines the most important terms and acronyms used in the remainder of this report.

Table 1: Glossary

Term / Acronym	Description
ISA Programme	The Interoperability Solutions for European Public Administrations Programme of the European Union [1]
LOGD	Linked Open Government Data
MS	Member State
RDF	Resource Description Framework
URI	Uniform Resource Identifier

2. BUSINESS CASE

This section describes the business case for extending the Core Vocabularies [4] for the publication of public data related to state aid. It identifies the business need, stakeholders, proposed solution, and expected benefits.

2.1. Stakeholders

The stakeholders and their roles in the project are summarised in the table below.

Table 2: Stakeholders and Roles

Stakeholder	Role	Description
DG Competition	Main beneficiary and owner of the outcome of the project: the state aid transparency RDF vocabulary.	Directorate General of Competition is willing to provide the final solution which will allow collecting and publishing transparency information as linked data.
MSs	Producer of the data to be published as open, machine-readable data.	The MSs provide their own solutions to collect transparency information, which will be afterward published as linked data by DG Competition's final solution.
EC, ISA Programme [1]	The ISA Programme has supported the project, reviewed the work carried out, and co-financed it as a demonstration of the use of Core Vocabularies.	The Interoperability Solutions for European Public Administrations (ISA) Programme of the European Commission [1] supports and facilitates efficient and effective cross-border electronic collaboration between European public administrations. The programme aims at enabling the delivery of electronic public services and the availability, interoperability, re-use and sharing of common solutions.
PwC	Solution supplier.	PwC builds a state aid vocabulary for the purpose of enabling interoperability at the semantic between all the applications involved collecting and publishing transparency information.

2.2. Business need

DG Competition needs to provide the means through which the MSs will comply with Article 9 of the Commission Regulation (EC) 651/2014 [2], which states that the MSs shall ensure the publication on a comprehensive state aid website of information about their aid measures and on individual aid awards exceeding EUR 500.000.

To comply with the regulation, the MSs will collect, verify and submit the transparency information to DG Competition. DG Competition will then, publish the transparency information as open, machine-readable data.

The Annex III of the Commission Regulation (EC) 651/2014 [2] provides a description of the transparency information, which consists of data about the MSs' implemented aid measures, such as objective of the aid, date of granting, value expressed as

amount in national currency, etc., and data about the beneficiaries of the aid awards, such as the name of the beneficiaries, type of the enterprise, sector of activity, etc. This information, which is of a high public value, is currently not available on the Web as reusable open data.

Communities such as openspending.org [7] and publicspending.net [8] who are interested in public financial transactions, could benefit from reusing this data and possibly connect with other related data in the context of budget monitoring/watchdog, visualisation and data journalism scenarios. These communities position themselves as “the resource for the many individuals and groups who wish to discuss and investigate public financial information, including journalists, academics, campaigners, and more”.

2.3. Proposed solution

The solution presented in this report comprises the development of RDF vocabulary for State Aid Transparency.

The following steps were performed:

- Analysis of the Annex III of the Commission Regulation (EC) 651/2014 [2] and map its properties with those of the Core Vocabulary.
- Application of the methodology outlined in “Handbook for using the Core Vocabularies” [5] to create a new vocabulary for the State Aid Transparency domain which reuses the Core Vocabularies. The methodology follows the steps depicted in Figure 1.
- Development of a proof-of-concept demonstrating the publication of transparency information as machine-readable open data using the developed RDF vocabulary.

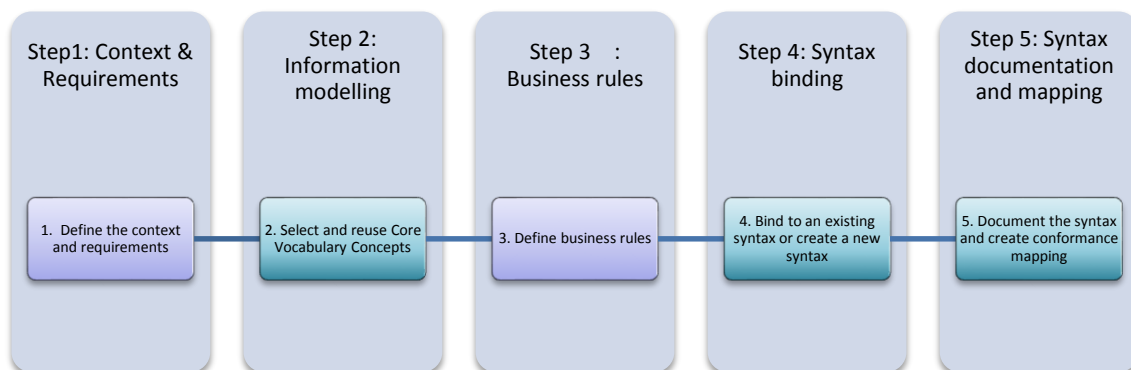


Figure 1 – Methodology on using the Core Vocabularies

2.4. Expected benefits

We anticipate the following benefits as a result of publishing transparency information as open, machine-readable data:

- Citizens, i.e. the tax payers, will discover easier information about public expenditures of their governments. This contributes directly to the core value

of government transparency.

- Journalists will benefit from a better understanding of how governments spend money as open, machine-readable data will allow for automatic processing and easier visualisation. It will thus be easier and more effective to write articles based on actual, factual data.
- Publishing state aid high-value data sets as open data, will support the mission of civic organisations such as [openspending.org](https://www.openspending.org/) who positions itself as a resource for the many individuals and groups who wish to discuss and investigate public financial information [7].
- Increased state aid transparency will support the mission of research initiatives such as [publicspending.net](https://www.publicspending.net/) who invest into demonstrating the usefulness of economic Linked Open Data in promoting transparency and innovation [8].
- MSs who will submit their transparency data to DG Competition, will not bear the costs of publication, as DG Competition will handle this process.
- By increasing transparency in the state aid domain, DG Competition will have the opportunity to focus more on the state aid cases that have a higher distortive potential, while allowing the MSs to implement state aid cases with less distortive potential. For DG Competition, this shift of focus will create the opportunity to optimize internal operational costs. For the MSs, this will mean a faster implementation of certain state aid measures.
- Once transparency data is published as open, machine-readable data, DG Competition will have the opportunity to publish as well the state aid expenditure data as open, machine-readable data, which will help for example journalists to do gap analysis between state aid commitments and state aid expenditures of their governments. The results of such analysis could help governments improve either their internal workflows to report state aid expenditures, or better target their state aids. In either case, the outcome will be an increased transparency of the state aid sector, and eventually better targeted state aid.

3. DEFINITION OF THE STATE AID VOCABULARY

3.1. Step 1: Context and Requirements

The first step is to understand the requirements for creating a new vocabulary. Once a MS grants an aid award to a certain beneficiary, the MS must submit the transparency information to DG Competition. DG Competition will then publish the information as open, machine-readable data. The diagram below depicts the steps to publish transparency information.

During this process, some of the MSs will submit transparency information from their own systems via APIs, while others will do it via UIs.

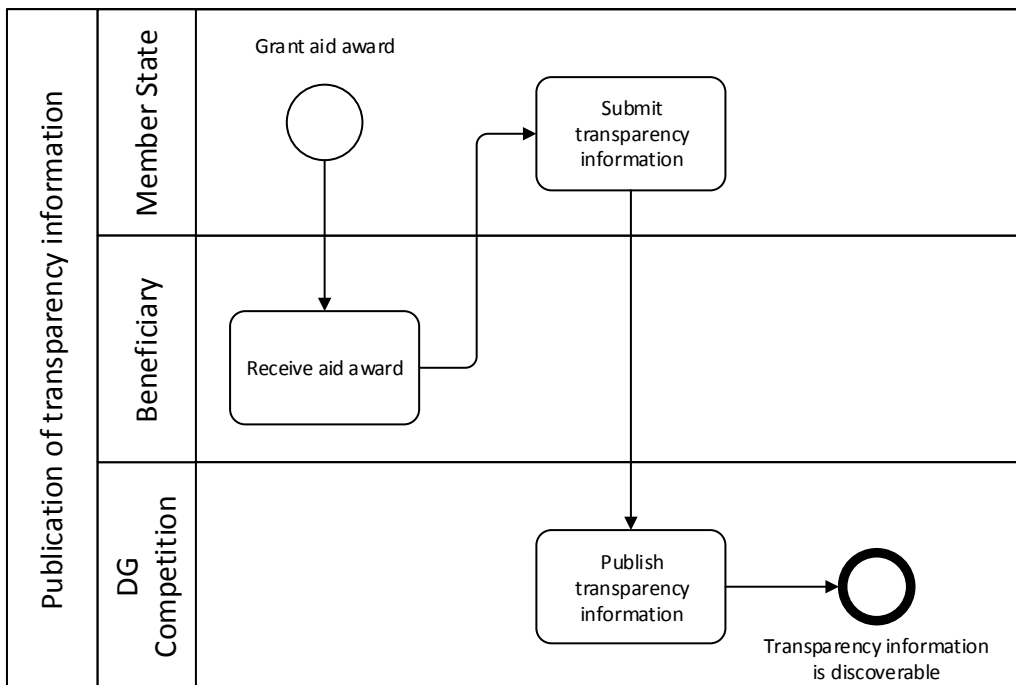


Figure 2 – The process for publishing transparency information

The table below describes the goals of the State Aid Transparency project.

Table 3 – Goals of the State Aid Transparency project

Goal ID	Goal Name	Goal Description
G1	Open up state aid, high-value data sets	Publish transparency information in reusable formats, using existing standard.
G2	Openness and machine-readability	Publish transparency data as open, machine-readable data, for reuse.

The table below lists the information requirements identified within the scope of the State Aid Transparency project. The identified requirements define the specific data items that must be included as part of the solution, in order to fulfil the goals of the project.

Table 4 – General and information requirements

Requirement identifier	Requirement name	Requirement statement	Rationale
R1	State aid award	Identify the state aid award	The state aid award is the main entity, which describes who receives the award, when, and what is its value.
R2	Beneficiary	Identify the beneficiary	The beneficiary of an aid award must be identified for further reference.
R3	Location	Location of the beneficiary	A beneficiary can have more than one location, but only one will be related to transparency information published for one state aid award.
R4	Aid measure	Identify the aid measure	The aid measure will be a reference to a full description of the aid measure, which will provide a link to other external systems managing information related to state aid.

3.2. Step 2: Information modelling

The conceptual data model for State Aid Transparency information must be based on the high-level information requirements described in the previous section.

We present in this section how to create this conceptual data model, by reusing conceptual data elements from the ISA Core Vocabularies [4].

The following steps are taken to create the conceptual State Aid Transparency data model:

- **Identify common concepts** between the new data model and concepts defined in the Core Vocabularies. The table below lists matching concepts between the State Aid Transparency data model and the Core Vocabularies.

Table 5 – Matching concepts with the Core Vocabularies

Requirement Identifier	Requirement name	Requirement statement	Core Vocabulary identifier
R2	Beneficiary	The identification of the beneficiary.	Legal Entity
R3	Location	Region where the beneficiary is located	Address

- **Align concepts and classes.**

- **Match names**

- “Beneficiary” is an exact match of the “Legal Entity” concept, as the beneficiary of an aid measure will always be a legally registered entity in a MS.
- “Location” is a close match of the “Address” concept. The location of a beneficiary will be expressed as a NUTS³ level 2 code.

- **Align the data model classes**

Properties of the Core Vocabularies are reused by restriction. Therefore, we identify properties of the classes from the Core Vocabularies to be used in the new data model. The properties of the Core Vocabularies that will not be used by the State Aid Transparency model, will be restricted.

The table below shows the classes and properties in the Core Vocabularies that were identified as matching elements from the State Aid Transparency domain. These are the LegalEntity and the Address classes. The white rows represent properties of classes, and the blue rows represent relationships with other classes. The column “Action” tells whether one property of a class from the Core Vocabulary matches a property from the State Aid Transparency, and it should be kept, or otherwise it should be restricted.

Table 6 – Alignment of the State Aid Transparency data model with the Core Vocabularies

Requirement property	Core Vocabulary property	Action
Beneficiary	Legal Entity	
Beneficiary national ID	LegalEntityLegalIdentifier	Keep
Beneficiary VAT	LegalEntityIdentifier	Keep
Beneficiary Name	LegalEntityLegalName	Keep
	LegalEntityAlternativeName	Restrict
Beneficiary Type	LegalEntityCompanyType	Keep
	LegalEntityCompanyStatus	Restrict
Sector	LegalEntityCompanyActivity	Keep
	LegalEntityRegisteredAddress	Restrict
	LegalEntityAddress	Keep
	LegalEntityLocation	Restrict
Location	Address	
	AddressFullAddress	Restrict
	AddressPOBox	Restrict

³ http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/introduction

	AddressThoroughfare	Restrict
	AddressLocatorDesignator	Restrict
	AddressLocatorName	Restrict
	AddressAddressArea	Restrict
	AddressPostName	Restrict
Location	AddressAdminUnitL2	Keep

The table below displays a final mapping between the information requirements of the Annex III of the Commission Regulation (EC) 651/2014 [2], and the corresponding properties of the Core Vocabularies.

Table 7 – Information requirements aligned with the Core Vocabularies

IRReqID	Type	Business Term	Definition	CoreVocID
	Classes	State Aid Award	Annex III of the General Block Exemption Regulation 651/2014	
IR-001	Prop	beneficiaryNationalID	The national identifier of the beneficiary (if available).	LegalEntityLegalIdentifier
IR-002	Prop	beneficiaryVAT	The EU VAT number of the beneficiary (if available).	LegalEntityIdentifier
IR-003	Prop	beneficiaryName	Name of the beneficiary	LegalEntityLegalName
IR-004	Prop	beneficiaryType	The beneficiary can be identified by one of the two values: SME, or large enterprise.	LegalEntityCompanyType
IR-005	Prop	Sector	The economic sector in which the beneficiary is active. It is expressed as a NACE code.	LegalEntityCompanyActivity
IR-006	Prop	location/region	The location of the beneficiary expressed as a NUTS 2 code: http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_CLS_DLD&StrNom=NUTS_33&StrLanguageCode=EN	AddressAdminUnitL2
IR-007	Prop	aidInstrument	Type of the aid measure. For example: tax reduction, etc.	
IR-008	Prop	objective	Main objective to be achieved by the beneficiary.	
IR-009	Prop	amount	Value of the aid, awarded to the beneficiary by the granting authority.	Value
IR-010	Prop	currency	The corresponding currency of the value of the aid awarded. NAL: http://publications.europa.eu/resource/authority/currency/	Currency
IR-011	Prop	language	The language of all the names in this form (beneficiaryName, grantingAuthority,	Language

			entrustedEntity, financialIntermediary)	
			NAL: http://publications.europa.eu/resource/authority/language/	
IR-012	Prop	dateGranted	Date of the aid award.	
IR-013	Prop	grantingAuthority	Name of the MS's authority that grants the aid	LegalEntityLegalName
IR-014	Prop	entrustedEntity		LegalEntityLegalName
IR-015	Prop	financialIntermediary		LegalEntityLegalName
IR-016	Prop	aidMeasure	Identification of the aid measure. This is the EC's internal identifier, automatically generated by the EC's state aid case management system.	

The data model below, based on the structure of Annex III of the Commission Regulation (EC) 651/2014 [2], depicts the concepts and their relationships that the State Aid Transparency vocabulary will model.

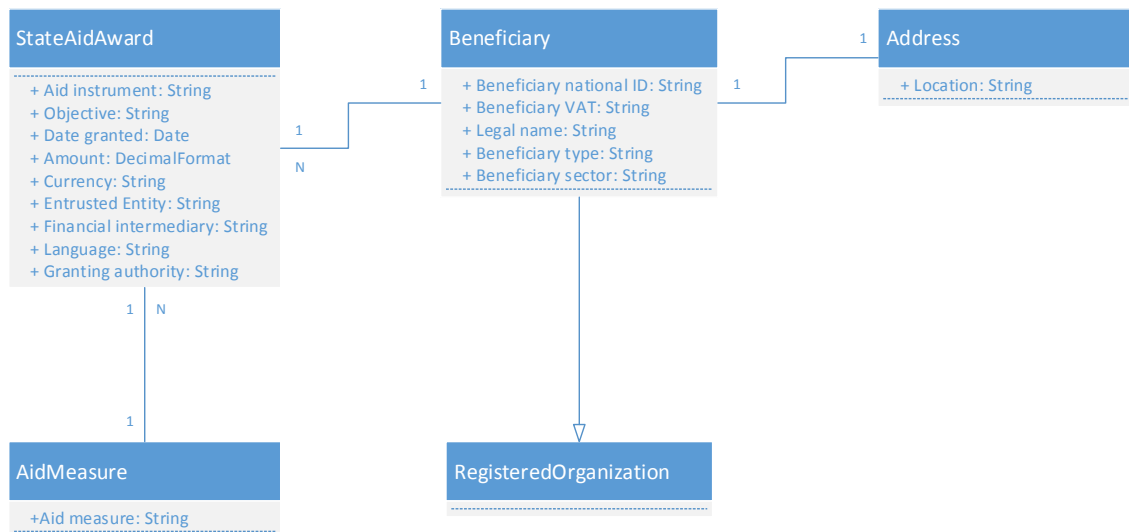


Figure 3 – State Aid Transparency data model

3.3. Step 3: Business rules

At this step, cardinalities of the data elements and new additional business rules are defined.

Table 8 – Information requirements with cardinalities

IReqID	Type	Business Term	Definition	Card	CoreVocID
	Class	State Aid Award	Annex III of the General Block Exemption Regulation 651/2014		

IR-001	Prop	beneficiaryNationalID	The national identifier of the beneficiary (if available).	0..1	LegalEntityLegalIdentifier
IR-002	Prop	beneficiaryVAT	The EU VAT number of the beneficiary (if available).	0..1	LegalEntityIdentifier
IR-003	Prop	beneficiaryName	Name of the beneficiary	1	LegalEntityLegalName
IR-004	Prop	beneficiaryType	The beneficiary can be identified by one of the two values: SME, or large enterprise.	1	LegalEntityCompanyType
IR-005	Prop	Sector	The economic sector in which the beneficiary is active. It is expressed as a NACE code.	1	LegalEntityCompanyActivity
IR-006	Prop	location/region	The location of the beneficiary expressed as a NUTS 2 code: http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_CLS_DLD&StrNom=NUTS_33&StrLanguageCode=EN	1	AddressAdminUnitL2
IR-007	Prop	aidInstrument	Type of the aid measure. For example: tax reduction, etc.	1	
IR-008	Prop	objective	Main objective to be achieved by the beneficiary.	1	
IR-009	Prop	amount	Value of the aid, awarded to the beneficiary by the granting authority.	1	Value
IR-010	Prop	currency	The corresponding currency of the value of the aid awarded. NAL: http://publications.europa.eu/resource/authority/currency/	1	Currency
IR-011	Prop	language	The language of all the names in this form (beneficiaryName, grantingAuthority, entrustedEntity, financialIntermediary) NAL: http://publications.europa.eu/resource/authority/language/	1	Language
IR-012	Prop	dateGranted	Date of the aid award.	1	
IR-013	Prop	grantingAuthority	Name of the MS's authority that grants the aid	1	LegalEntityLegalName
IR-014	Prop	entrustedEntity		0..1	LegalEntityLegalName
IR-015	Prop	financialIntermediary		0..1	LegalEntityLegalName
IR-016	Prop	aidMeasure	Identification of the aid measure. This is the EC's	1	

internal identifier,
automatically generated by
the EC's state aid case
management system.

3.4. Step 4: Syntax binding

At this step, we choose the syntax, and complete the mapping with the Core Vocabularies. RDF was chosen as the representation format.

The ISA Programme has developed guidelines on developing RDF vocabularies from conceptual data models [9]. These guidelines were followed in this work.

Following the general agreement among the stakeholders to use RDF as a representation format for the state aid transparency vocabulary, we reused the RDF representation of the Core Vocabularies:

- Registered Organization Vocabulary <http://www.w3.org/TR/vocab-regorg/>
- Location Core Vocabulary <http://www.w3.org/ns/locn>
- Core Public Service Vocabulary Application Profile <http://purl.org/vocab/cpsv>

From the Core Public Service Vocabulary Application Profile we used the Cost and Currency properties of the class Cost, to map the properties Amount and Currency of the newly created class State Aid Award from the state aid transparency vocabulary.

An overview of reused classes and properties can be found in Table 9.

We then minted new terms for any new property which did not match an existing property from the Core Vocabularies. An overview of minted classes and properties can be found in Table 10.

Drawing on the recommendations of the PURI study [6], URIs were also defined for all newly minted terms following the pattern:

`http://data.europa.eu/sat`

Such URI must be registered to be used for referring the data assets of the state aid transparency domain. The "sat" namespace is only a recommendation. Any other string which follows the recommendations of the PURI study [6] can be used. Regardless the chosen namespace, this must be registered with the PURI service in order for the RDF vocabulary to be used in any up-and-running application.

The State Aid Transparency RDF vocabulary can be found in the Annex II of this document.

3.5. Step 5: Syntax documenting and mapping

We used the mapping spreadsheet of the Core Data Model Mapping Directory to create the mappings. The result is the "Mappings" worksheet with the full list of properties identified for the State Aid Transparency domain. Annex I of the current document displays the "Mappings" worksheet.

The table below lists all the fields of the Annex III of the Commission Regulation (EC) 651/2014 [2], and their syntax mappings.

Table 9 – Properties of the State Aid Transparency mapped to properties of the Core Vocabularies

IRReqID	Label	Property	Vocabulary	Description
IR-001	Beneficiary national ID	rov:registration	rov	The national identifier of the beneficiary.
IR-002	Beneficiary VAT	org:identifier	org	The EU VAT number of the beneficiary.
IR003	Beneficiary name	rov:legalName	rov	The name of the beneficiary.
IR-004	Beneficiary type	rov:orgType	rov	The type of enterprise at the time of granting.
IR-006	Location	locn:adminUnitL2	locn	Region in which the beneficiary is located.
IR-005	Sector	rov:orgActivity	rov	Sector of activity.
IR-009	Amount	cpsv:value	cpsv	The committed amount of the aid award
IR-010	Currency	cpsv:currency	cpsv	The currency of the committed amount of the aid award.
IR-011	Language	dcterms:language	dcterms	The language of all the names in this form (beneficiaryName, grantingAuthority, entrustedEntity, financialIntermediary).
IR-013	Granting authority	rov:legalName	rov	The name of the authority who granted the aid award.
IR-014	Entrusted entity	rov:legalName	rov	The name of the entrusted entity.
IR-015	Financial intermediary	rov:legalName	rov	The names of the financial intermediaries.

The table below lists all the new properties to be minted.

Table 10 – New minted (defined) properties

Label	Property	Vocabulary	Description
Aid instrument			The aid instrument.
Objective			The objective of the aid measure.
Aid measure			The EC aid measure reference.
Date granted			The date when the aid award was granted.

4. PROOF OF CONCEPT

To showcase the created RDF vocabulary, we used a mock dataset and the publication tool Open Refine⁴ and its RDF extension⁵. The Open Data Support project⁶ of DG CONNECT has created a detailed example of how the EU Institutions can use Open Refine for publishing linked data⁷.

The purpose of this proof of concept is to use the state aid transparency vocabulary to publish a mock data set as open, machine-readable data. It is up to the development team of DG Competition to use the output of this proof of concept, and go one step further to develop test cases to demonstrate how the state aid transparency vocabulary fits the purpose of their Transparency project.

Table 11 – The mock data provided by DG Competition, which was provided as an Excel spreadsheet, which was transposed in this document for the sake of readability

BeneficiaryNationalID	34446632		3535353
BeneficiaryVAT	1234567890	RO4394839	
BeneficiaryName	Przedsiębiorstwo Hotelarsko-Gastronomiczne "Chata" Hotel-Restauracja "Chata Karczowiska" A. i H. Kozieja Sp.J.	Siveco company	ING Bank
BeneficiaryType	SME	SME	large
BeneficiaryLocation	Warsaw	Galati	Bruxelles
BeneficiarySector	35.2 Manufacture of gas; distribution of gaseous fuels through mains	IT	Banking
Amount	20000000	1000	43000
Currency	PLN	RON	EUR
AidInstrument	Grant	loan	tax exempt
DateOfGranting	06/09/2016	05/06/2014	07/07/2013
Objective	Regional urban development aid (Art. 16)	Investment	Economic growth
GrantingAuthority	Zarząd Województwa Lubelskiego	Ministerul dezvoltarii	Government Bruxelles
FinancialIntermediary	SANDIS LTD.		
AidMeasure	SA.45576	SA.434343	SA.656565
Language	PL	RO	FR
CountryCode	POL	ROU	FRA

⁴ <https://github.com/OpenRefine>

⁵ <https://github.com/fadmaa/grefine-rdf-extension>

⁶ <http://opendatasupport.eu>

⁷ https://joinup.ec.europa.eu/sites/default/files/7d/49/ee/D3.1.7%20-%20Introduction%20to%20linked%20data_EUI_LUX_0.02.pdf

BusinessBaseURI	http://data.europa.eu/business	http://data.europa.eu/business	http://data.europa.eu/business
------------------------	--------------------------------	--------------------------------	--------------------------------

The last two columns CountryCode and BusinessBaseURI were added to the spreadsheet to support our demo, and they are not part of the mock data received from DG Competition.

The Annex III of this report, contains a full description of the steps performed to publish the mock data received from DG Competition, as open, machine-readable data, using Open Refine. The output is a file containing the mock data in RDF format, having the structure defined by the state aid transparency vocabulary. Figure 4 below, shows a sample of this file.

```

<rdf:Description rdf:about="http://data.europa.eu/stateaidtransparency/sat/0">
  <rdf:type rdf:resource="http://data.europa.eu/stateaidtransparency/sat#StateAidAward"/>
  <sat:aidInstrument>Grant</sat:aidInstrument>
  <sat:objective>Regional urban development aid (Art. 16)</sat:objective>
  <sat:aidMeasure>SA.45576</sat:aidMeasure>
  <sat:grantingAuthority>Zarząd Województwa Lubelskiego</sat:grantingAuthority>
  <sat:entrustedEntity>SANDIS LTD.</sat:entrustedEntity>
  <cpsv:Value>Grant</cpsv:Value>
  <cpsv:Currency>PLN</cpsv:Currency>
  <dcterms:language>PL</dcterms:language>
  <dcterms:issued>Tue Sep 06 00:00:00 CEST 2016</dcterms:issued>
</rdf:Description>

<rdf:Description rdf:about="http://data.europa.eu/business/POL/3.4446632E7">
  <rdf:type rdf:resource="http://data.europa.eu/stateaidtransparency/sat#Beneficiary"/>
  <rov:registration>3.4446632E7</rov:registration>
  <dcterms:identifier>1.23456789E9</dcterms:identifier>
  <rov:legalName>Przedsiębiorstwo Hotelarsko-Gastronomiczne "Chata" Hotel-Restauracja Chata Karczowiska" A. i H. Kozieja Sp.J.</rov:legalName>
  <rov:orgType>SME</rov:orgType>
  <rov:orgActivity>35.2 Manufacture of gas; distribution of gaseous fuels through mains</rov:orgActivity>
</rdf:Description>

<rdf:Description rdf:about="http://data.europa.eu/stateaidtransparency/sat/Warshaw">
  <rdf:type rdf:resource="http://www.w3.org/ns/locn#Address"/>
  <locn:adminUnitL2>Warshaw</locn:adminUnitL2>
</rdf:Description>

<rdf:Description rdf:about="http://data.europa.eu/business/POL/3.4446632E7">
  <sat:hasAddress rdf:resource="http://data.europa.eu/stateaidtransparency/sat/Warshaw"/>
</rdf:Description>

<rdf:Description rdf:about="http://data.europa.eu/stateaidtransparency/sat/0">
  <sat:hasBeneficiary rdf:resource="http://data.europa.eu/business/POL/3.4446632E7"/>
</rdf:Description>

```

Figure 4 - Open Refine, sample output

5. CONCLUSIONS AND NEXT STEPS

Creating a new vocabulary for the State Aid Transparency domain by reusing the ISA Core Vocabularies (by restriction) demonstrates the cross-sector nature of the Core Vocabularies and their easy reuse to support different policy-domain-specific applications.

The flexibility of RDF is an enabler to the reuse of the Core Vocabularies not only at the conceptual, but also at the syntax level. RDF allows to easily reuse terms, i.e. classes and properties, from existing vocabularies to develop new ones, as long as domain and range restrictions are respected. More restrictions and considerations would have to be made, should the selected syntax have been XML.

The fact that State Aid Transparency data created as a result of this pilot is conformant with the Core Vocabularies, in particular with the Legal Entity and the Address classes, means that integrating this data with data coming from other Core-Vocabulary-conformant sources, such as BRIS, the pilots developed by ISA Action 1.1⁸, or OpenCorporates⁹, is made very easy.

There are several next steps recommended, following the development of the current version of the State Aid Transparency vocabulary:

1. Model the following elements as separate entities:
 - a. Granting authority
 - b. Entrusted entity
 - c. Financial intermediary

By publishing more information about who grants state aids, who are entrusted entities or financial intermediaries for certain state aid measures, would offer additional information, increasing the quality of the published, open data.

2. Analyse and choose identifiers for businesses. There are at least two initiatives which should be investigated:
 - a. BRIS – New EU Directive on Interconnecting Business Registers
https://joinup.ec.europa.eu/asset/core_business/news/new-eu-directive-interconnecting-business-registers
 - b. LEI – The Legal Entity Identifier Regulatory Oversight Committee
<http://www.leiroc.org/>
3. Provide access to the transparency information published as open, machine-readable data via an API, i.e. following the linked data paradigm.

⁸ https://joinup.ec.europa.eu/community/semic/og_page/pilots

⁹ <https://opencorporates.com/>

REFERENCES

1. Interoperability Solutions for European Public Administrations.
<http://ec.europa.eu/isa/>
2. European Commission, Commission Regulation (EU) No 651/2014 of 17 June 2014 declaring certain categories of aid compatible with the internal market in application of Articles 107 and 108 of the Treaty Text with EEA relevance.
http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2014.187.01.0001.01.ENG
3. European Commission, ISA Programme (2014). Report on high-value datasets from EU Institutions.
http://ec.europa.eu/isa/documents/publications/report-on-high-value-datasets-from-eu-institutions_en.pdf
4. Core Vocabularies are defined as simplified, re-usable and extensible data models that capture the fundamental characteristics of a data entity in a context-neutral fashion.
https://joinup.ec.europa.eu/asset/core_vocabularies/description
5. D03.03.02 – Handbook for using the Core Vocabularies.
https://joinup.ec.europa.eu/site/core_vocabularies/Core_Vocabularies_user_handbook/Handbook-for-using-the-Core-Vocabularies_v0.50.pdf
6. D7.1.3 – Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC.
<https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris>
7. OpenSpending exists to “map the money worldwide” – that is, to track and analyse public financial information globally.
<http://community.openspending.org/about/>
8. PublicSpending is a research initiative focused on engineering meaningful interconnections among public spending data to demonstrate the usefulness of economic Linked Open Data in promoting transparency and innovation.
<http://publicspending.net/about>
9. Cookbook for translating Data Models to RDF Schemas.
<https://joinup.ec.europa.eu/community/semic/document/cookbook-translating-data-models-rdf-schemas>
10. Slides 39 to 48. Introduction to Linked Data.
https://joinup.ec.europa.eu/sites/default/files/7d/49/ee/D3.1.7%20-%20Introduction%20to%20linked%20data_EUI_LUX_0.02.pdf

ANNEX I: CONCEPTUAL MAPPING BETWEEN THE CORE VOCABULARIES AND ANNEX III OF THE COMMISSION REGULATION (EC) 651/2014

Core Vocabulary Identifier	Link	Mapping relation	Data model	Identifier	Label	Definition	Mapping comment
LegalEntityLegalIdentifier	Link	Has exact match	Transparency	beneficiaryNationalID	Beneficiary national ID	The national identifier of the beneficiary (if available).	No pattern validation Business rule : national identifier or EU VAT number is mandatory
LegalEntityIdentifier	Link	Has exact match	Transparency	beneficiaryVAT	beneficiary VAT	The EU VAT number of the beneficiary (if available).	VIES pattern validation. Business rule : national identifier or EU VAT number is mandatory
LegalEntityLegalName	Link	Has exact match	Transparency	beneficiaryName	Beneficiary name	Name of the beneficiary	
LegalEntityCompanyType	Link	Has narrow match	Transparency	beneficiaryType	Beneficiary type	The beneficiary can be identified	beneficiaryType can take one of two possible values: "SME" (small and medium enterprise), or "large enterprise"

						by one of the two values: SME, or large enterprise
AddressAdminUnitL2	Link	Has close match	Transparency	location/region	Location / Region	The location of the beneficiary expressed as a NUTS 2 code. If values of AddressAdminUnitL2 could be NUTS code, then, the mapping relation could be changed from "Has close match" to "Has exact match"
LegalEntityCompanyActivity	Link	Has exact match	Transparency	sector	Sector	The economic sector in which the beneficiary is active. It is expressed as a NACE code.
	Error! Hyperlink reference not valid.	Has no match	Transparency	aidInstrument	Aid instrument	Type of the aid measure. For example: tax

						reduction, etc.
	Error! Hyperlink reference not valid.	Has no match	Transparency	objective	Objective	Main objective to be achieved by the beneficiary.
	Error! Hyperlink reference not valid.	Has no match	Transparency	aidMeasure	Aid measure	Identification of the aid measure. This is the EC's internal identifier, automatically generated by the EC's state aid case management system.
Value	Link	Has exact match	Transparency	amount	Amount	Value of the aid, awarded to the beneficiary by the Value is a property of the class Cost defined in the CPSV-AP

						granting authority.
Currency	Link	Has exact match	Transparency	currency	Currency	The corresponding currency of the value of the aid awarded. Currency is a property of the class Cost defined in the CPSV-AP
Language	Link	Has exact match	Transparency	language	Language	Identification of the aid measure. This is the EC's internal identifier, automatically generated by the EC's state aid case management system. http://publications.europa.eu/mdr/resource/authority/language/
	Error! Hyperlink reference	Has no match	Transparency	dateGranted	Date granted	Date of the aid award.

not valid.						
LegalEntityLegalName	Link	Has close match	Transparency	grantingAuthority	Granting authority	Name of the Member State's authority that grants the aid
LegalEntityLegalName	Link	Has close match	Transparency	entrustedEntity	Entrusted entity	To be defined
LegalEntityLegalName	Link	Has close match	Transparency	financialIntermediary	Financial intermediary	To be defined

ANNEX II: THE STATE AID TRANSPARENCY RDF VOCABULARY

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:xmlns:adms = "http://www.w3.org/ns/adms#"
  xmlns:cc = "http://creativecommons.org/ns#"
  xmlns:dcterms = "http://purl.org/dc/terms/"
  xmlns:foaf = "http://xmlns.com/foaf/0.1/"
  xml http://www.w3.org/TR/vocab-org/
  xmlns:owl = "http://www.w3.org/2002/07/owl#"
  xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs = "http://www.w3.org/2000/01/rdf-schema#"
  xml http://www.w3.org/ns/regorg#
  xmlns:schema = "http://schema.org/"
  xmlns:skos = "http://www.w3.org/2004/02/skos/core#"
  xmlns:voaf = "http://labs.mondeca.com/vocab/voaf#"
  xmlns:vso = "http://www.heppnetz.de/ontologies/vso/ns#"
  xmlns:xsd = "http://www.w3.org/2001/XMLSchema#"
  xml http://www.w3.org/ns/locn#
  xml http://purl.org/vocab/cpsv#
  xml = "http://data.europa.eu/stateaidtransparency/sat#">

<owl:Ontology rdf:about="sat:">
  <rdfs:label xml:lang="en">State aid transparency vocabulary</rdfs:label>
  <rdfs:comment xml:lang="en">This file specifies the State Aid Transparency
  (sat) vocabulary of DG COMP</rdfs:comment>
  <owl:versionInfo xml:lang="en">v0.03</owl:versionInfo>
  <dcterms:modified rdf:datatype="xsd:date">2015-02-25</dcterms:modified>
  <rdf:type rdf:resource="voaf:Vocabulary"/>
  <rdf:type rdf:resource="adms:SemanticAsset"/>
  <dcterms:publisher>
    <rdf:Description rdf:about="http://publications.europa.eu/resource/authority/corporate-body/COM">
      </rdf:Description>
    </dcterms:publisher>
  <dcterms:creator rdf:parseType="Resource">
    <foaf:name>Alex Droscariu</foaf:name>
    <schema:affiliation rdf:parseType="Resource">
      <foaf:name>PwC EU Services</foaf:name>
    </schema:affiliation>
  </dcterms:creator>
  <dcterms:rights>Copyright 2015 - European Commission.</dcterms:rights>
  <cc:attributionName>European Commission</cc:attributionName>
  <cc:attributionURL rdf:resource="http://ec.europa.eu/">
</owl:Ontology>

<!-- Class State Aid Award -->

<rdfs:Class rdf:about="sat:StateAidAward">
  <rdfs:label xml:lang="en">State Aid Award</rdfs:label>
  <rdfs:comment xml:lang="en">Annex III of the General Block Exemption Regulation
  651/2014</rdfs:comment>
  <dcterms:identifier>sat:StateAidAward</dcterms:identifier>
</rdfs:Class>

<!-- New minted properties of the class State Aid Award -->

<rdf:Property rdf:about="sat:aidInstrument">
  <rdfs:label xml:lang="en">Aid instrument</rdfs:label>
  <rdfs:comment xml:lang="en">Type of the aid measure. There are several possible
  values, such as: tax reduction, etc.</rdfs:comment>
  <dcterms:identifier>sat:aidInstrument</dcterms:identifier>
  <rdfs:range rdf:datatype="xsd:string" />
</rdf:Property>
```

[The listing continues on the next page](#)

```
<rdf:Property rdf:about="sat:objective">
  <rdfs:label xml:lang="en">Aid objective</rdfs:label>
  <rdfs:comment xml:lang="en">Main objective to be achieved by the beneficiary with
the use of the state aid award.</rdfs:comment>
  <dcterms:identifier>sat:objective</dcterms:identifier>
  <rdfs:range rdf:datatype="xsd:string" />
</rdf:Property>

<rdf:Property rdf:about="sat:aidMeasure">
  <rdfs:label xml:lang="en">Aid measure</rdfs:label>
  <rdfs:comment xml:lang="en">Identification of the aid measure. This is the EC's
internal identifier, automatically generated by the EC's state aid case management
system</rdfs:comment>
  <dcterms:identifier>sat:aidMeasure</dcterms:identifier>
  <rdfs:range rdf:datatype="xsd:string" />
</rdf:Property>

<rdf:Property rdf:about="sat:grantingAuthority">
  <rdfs:label xml:lang="en">Granting authority</rdfs:label>
  <rdfs:comment xml:lang="en">The name of the Member State's authority that grants
the state aid award.</rdfs:comment>
  <dcterms:identifier>sat:grantingAuthority</dcterms:identifier>
  <rdfs:subPropertyOf>skos:altLabel</rdfs:subPropertyOf>
</rdf:Property>

<rdf:Property rdf:about="sat:entrustedEntity">
  <rdfs:label xml:lang="en">has entrusted Entity</rdfs:label>
  <rdfs:comment xml:lang="en">The name of the entrusted entity of the state aid award.
</rdfs:comment>
  <dcterms:identifier>sat:entrustedEntity</dcterms:identifier>
  <rdfs:subPropertyOf>skos:altLabel</rdfs:subPropertyOf>
</rdf:Property>

<rdf:Property rdf:about="sat:financialIntermediary">
  <rdfs:label xml:lang="en">Financial intermediary</rdfs:label>
  <rdfs:comment xml:lang="en">The name of the financial intermediary.</rdfs:comment>
  <dcterms:identifier>sat:financialIntermediary</dcterms:identifier>
  <rdfs:subPropertyOf>skos:altLabel</rdfs:subPropertyOf>
</rdf:Property>

<rdf:Property rdf:about="sat:hasBeneficiary">
  <rdfs:label xml:lang="en">has beneficiary</rdfs:label>
  <rdfs:comment xml:lang="en">The link to the beneficiary of the state aid award.
</rdfs:comment>
  <rdfs:domain rdf:resource="sat:StateAidAward" />
  <rdfs:range rdf:resource="sat:Beneficiary" />
  <dcterms:identifier>sat:hasBeneficiary</dcterms:identifier>
</rdf:Property>

<rdf:Description rdf:about="cpsv:Value">
  <rdfs:label xml:lang="en">Amount</rdfs:label>
  <rdfs:comment xml:lang="en">The value of the state aid award.</rdfs:comment>
  <dcterms:identifier>cpsv:Value</dcterms:identifier>
</rdf:Description>

<rdf:Description rdf:about="cpsv:Currency">
  <rdfs:label xml:lang="en">Currency</rdfs:label>
  <rdfs:comment xml:lang="en">The currency of the value of the state aid award.
</rdfs:comment>
  <dcterms:identifier>cpsv:Currency</dcterms:identifier>
</rdf:Description>
```

[The listing continues on the next page](#)

```
<rdf:Description rdf:about="dcterms:language">
  <rdfs:label xml:lang="en">Language</rdfs:label>
  <rdfs:comment xml:lang="en">The language of all the names in this form.</rdfs:comment>
  <dcterms:identifier>dcterms:language</dcterms:identifier>
</rdf:Description>

<rdf:Description rdf:about="dcterms:issued">
  <rdfs:label xml:lang="en">Date granted</rdfs:label>
  <rdfs:comment xml:lang="en">Date when the state aid award was granted.</rdfs:comment>
  <dcterms:identifier>sat:dateGranted</dcterms:identifier>
</rdf:Description>

<!-- Class Beneficiary *** define it as a subclass of registered organization -->
<rdfs:class rdf:about="sat:Beneficiary">
  <rdfs:label xml:lang="en">Beneficiary</rdfs:label>
  <rdfs:comment xml:lang="en">The beneficiary of the aid award</rdfs:comment>
  <dcterms:identifier>sat:Beneficiary</dcterms:identifier>
  <rdfs:subClassOf rdf:resource="rov:RegisteredOrganization" />
</rdfs:class>

<!-- Properties of the Class Beneficiary -->
<rdf:Description rdf:about="rov:registration">
  <rdfs:label xml:lang="en">Beneficiary national ID</rdfs:label>
  <rdfs:comment xml:lang="en">The national identifier of the beneficiary.</rdfs:comment>
  <rdfs:domain rdf:resource="sat:Beneficiary" />
  <dcterms:identifier>rov:registration</dcterms:identifier>
</rdf:Description>

<rdf:Description rdf:about="dcterms:identifier">
  <rdfs:label xml:lang="en">Beneficiary VAT</rdfs:label>
  <rdfs:comment xml:lang="en">The EU VAT number of the beneficiary.</rdfs:comment>
  <rdfs:domain rdf:resource="sat:Beneficiary" />
</rdf:Description>

<rdf:Description rdf:about="rov:legalName">
  <rdfs:label xml:lang="en">Legal name</rdfs:label>
  <rdfs:comment xml:lang="en">The name of the organization.</rdfs:comment>
  <rdfs:domain rdf:resource="sat:Beneficiary" />
  <dcterms:identifier>rov:legalName</dcterms:identifier>
</rdf:Description>

<rdf:Description rdf:about="rov:orgType">
  <rdfs:label xml:lang="en">Beneficiary type</rdfs:label>
  <rdfs:comment xml:lang="en">The type of the beneficiary.</rdfs:comment>
  <rdfs:domain rdf:resource="sat:Beneficiary" />
  <dcterms:identifier>rov:orgType</dcterms:identifier>
</rdf:Description>

<rdf:Description rdf:about="rov:orgActivity">
  <rdfs:label xml:lang="en">Beneficiary sector</rdfs:label>
  <rdfs:comment xml:lang="en">The economical sector in which the beneficiary is active.
  It is expressed as a NACE code.</rdfs:comment>
  <rdfs:domain rdf:resource="sat:Beneficiary" />
  <dcterms:identifier>rov:orgActivity</dcterms:identifier>
</rdf:Description>

<rdf:Property rdf:about="sat:hasAddress">
  <rdfs:label xml:lang="en">has beneficiary address</rdfs:label>
  <rdfs:comment xml:lang="en">The location of the beneficiary expressed as a
  NUTS 2 code.</rdfs:comment>
  <rdfs:domain rdf:resource="sat:Beneficiary" />
  <rdfs:range rdf:resource="locn:Address" />
  <dcterms:identifier>sat:hasAddress</dcterms:identifier>
</rdf:Property>
```

[The listing continues on the next page](#)

```
<!-- Re-use the locn:Address -->
<rdf:Description rdf:about="locn:Address">
  <rdfs:label xml:lang="en">Address</rdfs:label>
  <rdfs:comment xml:lang="en">The address of the beneficiary</rdfs:comment>
  <dcterms:identifier>locn:Address</dcterms:identifier>
</rdf:Description>

<rdf:Property rdf:about="locn:adminUnitL2">
  <rdfs:label xml:lang="en">Location</rdfs:label>
  <rdfs:comment xml:lang="en">The region of the address.</rdfs:comment>
  <dcterms:identifier>locn:adminUnitL2</dcterms:identifier>
</rdf:Property>

</rdf:RDF>
```

[End of the listing](#)

ANNEX III: USE THE STATE AID TRANSPARENCY VOCABULARY TO PUBLISH A MOCK DATA SET AS OPEN, MACHINE-READABLE DATA

To publish the data sample using the Open Refine and its RDF extension, we performed the steps below:

1. Install Open Refine following the instructions in its documentation.
2. Install the RDF extension of Open Refine:
 - 2.1. Extract the archive of the `grefine-rdf-extension`.
 - 2.2. Copy the content of the un-archived extension into the `<OpenRefineHomeRootFolder>/webapp/extensions/`
 - 2.3. Restart Open Refine.
3. Transpose the table above and save it as a spreadsheet.
4. Upload the spreadsheet from step 3 into Open Refine, and create a new project.
5. Edit the RDF Skeleton for your new project.

- 5.1. Update the Base URI with the value:
<http://data.europa.eu/stateaidtransparency/sat/>

This is only as an example. In reality, it should be replaced with a persistent URI under the control of DG Competition.

- 5.2. Add the following prefixes:
 - "sat" with URI: "http://data.europa.eu/stateaidtransparency/sat#"
 - "cpsv" with URI: "http://purl.org/vocab/cpsv#"
 - "locn" with URI: "http://www.w3.org/ns/locn#"
 - "rov" with URI: "http://www.w3.org/ns/regorg#"
 - "dcterms" with URI "http://purl.org/dc/terms/"
 - 5.3. Remove all the properties created by default by Open Refine (Open Refine creates these default properties based on the column headers of the spreadsheet previously uploaded at step 4):
 - property?->BeneficiaryNationalID cell
 - property?->BeneficiaryVAT cell
 - property?->BeneficiaryName cell
 - property?->BeneficiaryType cell
 - property?->BeneficiaryLocation cell

- property?->BeneficiarySector cell
- property?->Amount cell
- property?->Currency cell
- property?->AidInstrument cell
- property?->DateOfGranting cell
- property?->Objective cell
- property?->GrantingAuthority cell
- property?->FinancialIntermediary cell
- property?->AidMeasure cell
- property?->Language cell
- property?->CountryCode cell
- property?->BusinessBaseURI cell

The arrow (->) in the list above, means that the default property "property" maps to the spreadsheet column "<name of the column>".

5.4. Add new properties from the state aid transparency RDF vocabulary, and map each of them to their corresponding columns in the spreadsheet:

- sat:aidInstrument->AidInstrument cell
- sat:objective->Objective cell
- sat:aidMeasure->AidMeasure cell
- sat:grantingAuthority->GrantingAuthority cell
- sat:entrustedEntity->FinancialIntermediary cell
- cpsv:value->Amount cell
- cpsv:currency->Currency cell
- dcterms:language->Language cell
- dcterms:issued->DateOfGranting cell
- sat:hasBeneficiary->BeneficiaryNationalID URI

Select for the BeneficiaryNationalID that the cell's content will be used as URI, and add as a custom expression the following:

```
join([cells.BusinessBaseURL.value, cells.CountryCode.value,  
if(isBlank(cells.BeneficiaryNationalID.value),cells.BeneficiaryVAT.value,cells  
.BeneficiaryNationalID.value)], '/')
```

This will allow Open Refine to display the values of the beneficiary national ID as a URI concatenating: <Base URI> + <country code> + <VAT of the beneficiary or its national ID, whichever is first provided in the spreadsheet>, all separated by "/". In our example, the beneficiary should be a URI, as in a production environment (real transparency data published as open, machine-readable data), a beneficiary will be an instance of the Beneficiary class as modelled in the State Aid Transparency RDF vocabulary (See Annex II for a complete listing of the RDF file. Example: <http://data.europa.eu/business/ROU/RO4394839>)

Under the property sat:hasBeneficiary add rdf:type sat:Beneficiary property, under which create the following new properties:

- rov:registration->BeneficiaryNationalID cell
- dcterms:identifier->BeneficiaryVAT cell
- rov:legalName->BeneficiaryName cell
- rov:orgType->BeneficiaryType cell
- rov:orgActivity->BeneficiarySector cell
- sat:hasAddress->BeneficiaryLocation cell
- Under the property sat:hasAddress add rdf:type locn:Address property, under which create the last property:
- locn:adminUnitL2->BeneficiaryLocation cell

5.5. Save all the changes.

6. Export as RDF/XML