



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 732064



This project is part of BDV PPP

PUBLICATION OF INSPIRE-BASED AGRICULTURAL LINKED DATA

Raul Palma¹, Tomáš Řezník², Karel Charvát², Soumya Brahma¹, Dmitrij Kozuch², Raitis Berzins²

¹Poznan Supercomputing and Networking Center, Poland

²WirelessInfo, Czech Republic

Linked Open Data in Agriculture

MACS-G20 Workshop in Berlin (Germany), 27 – 28 September 2017

This document is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under agreement No 732064. It is the property of the DataBio consortium and shall not be distributed or reproduced without the formal approval of the DataBio Management Committee. Find us at www.databio.eu.

Motivation

- Farm management context
 - Multiple activities and stakeholders
 - Multiple applications, tools and devices
 - Multiple data sources, data types and data formats

Challenge

 To combine/integrate those different and heterogeneous data sources in order to make economically and environmentally sound decisions



Data Integration in relevant projects (context)

- DATABIO Data-driven Bioeconom
- Data integration challenges have been the focus of relevant projects



EU FP7, ICT CIP, 2014-2017

SDI4Apps aimed at building a cloudbased framework with **open API** for **data integration** focusing on the development of six pilot apps, drawing along the lines of **INSPIRE**, **Copernicus** and **GEOSS**





EU FP7, ICT CIP, 2014-2017

FOODIE aimed at building an **open and interoperable** cloud-based platform addressing among others the **integration of data** relevant to **farming** production including their **geo-spatial dimension**, as well as their **publication as Linked data**.

DataBio aims at showcasing the benefits of **Big** Data technologies in the raw material production from agriculture & others for the bioeconomy industry; deploying an interoperable platform on top of the existing partners' infrastructure.

reproduced without the formal approval of the DataBio Management Committee. Find us at www.databio.eu.

DataBio aims at delivering solutions for big data mgmt., including i) the **storage** and **querying** of **various big data sources**; ii) the harmonization and **integration** of a **large variety of data** from many sources, using **linked data** as a federated layer

Linked data publication in agriculture



- LD is increasingly becoming a popular method for publishing data on the Web
 - Improves data accessibility by both humans and machines, e.g., for finding, reuse and integration
 - Enables to **discover** more **useful data** through the links, and to exploit data with **semantic queries**
- Growing number of datasets in the LOD cloud
 - > 1100 by 22nd August 2017
- Coverage of the LOD cloud
 - Large cross-domain datasets (dbpedia, freebase, etc.)
 - Domain coverage varies (e.g., large number of datasets in Geography, Government, BioInformatics)
- What about Agriculture?
 - Only few examples (AGRIS biblio records, AGROVOC thesaurus + other thesaurus like NALT)
 - Farming data?

This document is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under agreement No 732064. It is the property of the DataBio consortium and shall not be distributed or reproduced without the formal approval of the DataBio Management Committee. Find us at www.databio.eu



Linked data publication process overview

- Simple set of principles & technologies
 - URI, HTTP, RDF, SPARQL
- Involves a set of tasks





Reference Linked data publication pipelines



Linked data publication technologies overview

- Used technologies:
 - D2RQ for transforming Relational Databases as Virtual RDF Graphs
 - RDF for the representation of data
 - Farming ontology providing the underlying vocabulary and relations
 - Virtuoso for storing the semantic datasets
 - Silk for discovery of links
 - Sparql for querying semantic data
 - Hslayers NG for visualisation of data
 - Metaphactory for visualisation of data







Datasets identification

- Goal: to publish linked data from pilots in FOODIE project (available in PostgreSQL database):
 - Precision viticulture (Spain)
 - Delivered a web-based solution providing advisory services in different aspects related to winegrowing, like disease prevention, production estimation or harvesting schedule
 - Open Data for Strategic and Tactical planning (Czech Republic)
 - Delivered two main applications, one for farm telemetry and other for estimation of yield potential

Data model for farming data

- Goal: (i) to define the application vocabulary covering the different categories of information dealt by the farm mgmt. tools/apps (in FOODIE) (ii) in line with existing standards and best practices
- **INSPIRE** directive is an EU initiative that aims at building a Pan-European spatial data infrastructure (SDI) requiring
 - EU Member States to make available spatial data, from multiple thematic areas, according to established implementing rules using appropriate services.
 - Based on ISO/OGC standards for geographical information, i.e., ISO 19100 series standards
- Hence, FOODIE data model builds on
 - the INSPIRE specification for agricultural and aquaculture facilities theme AF (for agricultural data), and
 - the INSPIRE data specification for themes in annex I for for geospatial data

This document is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under agreement No 732064. It is the property of the DataBio consortium and shall not be distributed or reproduced without the formal approval of the DataBio Management Committee. Find us at www.databio.eu

*consulted with experts from various institutions, e.g., EU DG JRC, EU Global Navigation Satellite Systems Agency (GSA), Czech Ministry of Agriculture, Global Earth Observation System of Systems (GEOSS), German Kuratorium für Technik und Bauwesen in der Landwirtschaft (KTBL).

Data model for farming data

- INSPIRE data specifications are defined as UML models and are available in different XML-based formats
- FOODIE extensions on the UML model developed by domain experts*
- Challenge: Transform model into OWL ontology

*consulted with experts from various institutions, e.g., EU DG JRC, EU Global Navigation Satellite Systems Agency (GSA), Czech Ministry of Agriculture, Global Earth Observation System of Systems (GEOSS), German Kuratorium für Technik und Bauwesen in der Landwirtschaft (KTBL).

Transformation from UML model to OWL ontology

- Followed a semi-automatic approach
 - ShapeChange tool that implements ISO 19150-2 standard rules for mapping ISO geographic information UML models to OWL ontologies.
- Required different processing tasks:
 - Pre-processing
 - Source model preparation
 - ShapeChange tool configuration: encoding rules; mappings UML classes OWL elements; namespaces definition
 - Base ontologies fixes (INPSIRE common, ISO 19100 series standards)
 - Post-processing tasks
 - Manual fixes in the ontology
 - Manual creation of ontology elements of the base INSPIRE schemas (AF)

DATABio

Palma R., Reznik T., Esbri M., Charvat K., Mazurek C., An INSPIRE-based vocabulary for the publication of Agricultural Linked Data. Proceedings of the OWLED Workshop: OWL Experiences and Directions, collocated with the 14th International Semantic Web Conference (ISWC-2015), Bethlehem PA, USA, October 11-15, 2015

Ontology for farming data - overview

- ShapeChange output
 - UML featureTypes and dataTypes modelled as classes, a their attributes as datatype or object properties
 - UML codeLists modelled as classes/concepts, and their attributes as concept members
 - Cardinalities restrictions defined on properties (exactly, min, max)

iso19150-2:Datatype

- DataType properties ranges defined according to model/mappings
- Object properties ranges defined according to model/mappings
- Object properties inverseOf defined

This document is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under agreement No 732064. It is the property of the DataBio consortium and shall not be reproduced without the formal approval of the DataBio Management Committee. Find us

Ontology for farming data – main classes overview

- For the purposes of FOODIE, we found the lack of a feature on a more detailed level than **Site** that is already part of the INSPIRE AF data model.
- Main concept: Plot
 - Represents a continuous area of agricultural land with one type of crop species, cultivated by one user in one farming mode
 - Two kinds of data associated:
 - metadata information
 - agro-related information

Next concept: Management Zone

• Enables a more precise description of the land characteristics in fine-grained area

This document is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under agreement No 732064. It is the property of the DataBio consortium and shall not be distributed or reproduced without the formal approval of the DataBio Management Committee. Find us at www.databio.eu

Ontology for farming data – main classes overview

- The **Intervention** is the basic feature type for any kind of (farming) application with explicitly defined geometry, e.g., tillage or pruning.
 - Has multiple indirect associations with different concepts

RDF data generation

 D2RQ requires a mapping file (in RDF) specifying how to map the content of a relational database to RDF

```
# Table foodie.site_view
map:foodie_site_view a d2rq:ClassMap;
    d2rq:dataStorage map:database;
    d2rq:uriPattern "Site/@@foodie.site_view.site_id@@";
    d2rq:class af-inspire:Site;
    d2rq:classDefinitionLabel "Site";
```

```
map:foodie_site_view__label a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:foodie_site_view;
    d2rq:property rdfs:label;
    d2rq:pattern "Site #@@foodie.site_view.site_id@@";
```

map:foodie_site_view_code a d2rq:PropertyBridge; d2rq:belongsToClassMap map:foodie_site_view; d2rq:property foodie:code; d2rq:column "foodie.site_view.code"; d2rq:datatype xsd:string;

https://github.com/FOODIE-cloud/ontology/tree/master/mappings

This document is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under agreement No 732064. It is the property of the DataBio consortium and shall not be distributed or reproduced without the formal approval of the DataBio Management Committee. Find us at www.databio.eu.

RDF data generation

 The mapping file also specifies the connection details for the source dataset

> map:database a d2rq:Database; d2rq:jdbcDriver "org.postgresql.Driver"; d2rq:jdbcDSN "jdbc:postgresql://localhost/foodie-es_db"; d2rq:username " "; d2rq:password " ";

- Based on the mapping file, the database content was dumped to an RDF file
- The RDF file was then loaded into Virtuoso triplestore

OpenLink Virtuoso version 07.20.3214 as of Dec 4 2015, on Linux (x86_64-unknown-linux-gnu), Single-Server Edition (23 GB total memory)

This document is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under agreement No 732064. It is the property of the DataBio consortium and shall not be distributed or reproduced without the formal approval of the DataBio Management Committee. Find us at www.databio.eu

Linking the generated RDF dataset

- In order to link the resulting RDF dataset with other datasets, we used Silk, but also had to do some manual entries
 - We found issues in handling large datasets in Silk, specially those accessed via SPARQL endpoint that we cannot control

- → C ☆ ① silk.foodie-cloud	1.org/linking/INSPIRE_hackathon/foodie-es-dbpedia/editor	
Silk Workbench	Start Workspace Status Editor Generate Links Reference Links About	
Undo Redo Get LinkSpo	ec Help	
roperty Paths 🤤 ource: foodie-es	sourcePath1	
(custom path)	/ <http: fc<="" foodie-cloud.com="" model="" td=""><td></td></http:>	
/rdfs:label	Path (Source) levenshteinDistance	1 🛛 🔁 🗖
/ <http: fc<="" foodie-cloud.com="" model="" td=""><td>Required 🗌</td><td></td></http:>	Required 🗌	
arget: dbpedia	targetPath1	
(custom path)	/rdfs:label Lower case (Transform) minChar 0	
/ <http: dbpedia.org="" ontology="" td="" wikil<=""><td>maxChar z</td><td></td></http:>	maxChar z	
<pre>\<http: dbpedia.org="" ontology="" pre="" wikif<=""></http:></pre>	Levensntein a	stance (compare)
ransformations Recommended \$		
Constant		
Lower case		
Tokenize		
omparators Recommended \$		
Equality		
Jaccard		
Levenshtein distance		
ggregators Recommended \$		
Average		
Maximum		
Minimum	Link Limit: unlimited 🔉 🕘 Link Type: rdfs:seeAlso	

<http://w3id.org/foodie/core/CropType/l> <http://w3id.org/foodie/core/CropType/10> <http://w3id.org/foodie/core/CropType/11> <http://w3id.org/foodie/core/CropType/12> <http://w3id.org/foodie/core/CropType/13> <http://w3id.org/foodie/core/CropType/14> <http://w3id.org/foodie/core/CropType/15> <http://w3id.org/foodie/core/CropType/16> <http://w3id.org/foodie/core/CropType/17>

<http://www.w3.org/2002/07/owl#sameAs>

<http://www.wikidata.org/entity/Q448487> . <http://www.wikidata.org/entity/Q1474000> <http://www.wikidata.org/entity/Q1065249> <http://www.wikidata.org/entity/Q63319> . <http://www.wikidata.org/entity/Q251054> <http://www.wikidata.org/entity/Q1293271> <http://www.wikidata.org/entity/Q123388> <http://www.wikidata.org/entity/Q657693>

Exploiting the Linked Data - querying

• Sparql endpoint: <u>https://www.foodie-cloud.org/sparql</u>

Virtuoso SPARQL Query Editor									
Default Data Set Name (Graph IRI)					About Namespace Prefixes Inference rule	<u>36</u>			
Query Text PREFIX geo: <http: geosparql#="" ont="" www.opengis.net=""> PREFIX rdfs: <http: 01="" 2000="" rdf="schema#" www.w3.org=""> PREFIX plot1: <http: 1="" core="" foodie="" plot="" w3id.org=""> PREFIX foodie: <http: foodie#="" foodie-cloud.com="" model=""> PREFIX foodie-es: <http: foodie-cloud.com="" foodie-es#="" model=""> PREFIX owl: <http: 07="" 2002="" owl#="" www.w3.org=""> PREFIX iso19103: <http: 2005="" basic<br="" def.seegrid.csiro.au="" iso19103="" isotc211="">SELECT ?resource ?zone_name ?crop (bif.st_astext(?coordinates) as ?coordin WHERE {</http:></http:></http:></http:></http:></http:></http:>	#> ates_as_wkt) ?crop_description	n ?begin_date ?end_date ?campaign 7	'amount_value ?amount_uom	?image ?sameas	?seealso				
?resource rdfs:label ?label . ?resource foodie-es:managementZoneName ?zone_name . ?resource foodie:gropSpecies ?grop	н								
Presource resource geo:hasGeometry ?geo. ?geo geo:asWKT ?coordinates . ?resource foodie:holdingZone plot1: . ?product a foodie:ProductionType . (Security restrictions of this server do not allow you to retrieve remote RDF data, see details.) Results Format: HTML ‡ Execution timeout: 0 Options: Strict checking of void variables (The result can only be sent back to browser, not saved on the server, see details) Run Query Reset	http://w3id.org/foodie/core/ManagementZone/33	*C2*^^ <http: 2001="" www.w3.org="" xmlschema#string=""></http:>	http://w3id.org/foodie/core/CropType/1	coordinates_as_wkt POLYGON((-8.80489 41.939885,-8.803117 41.940124,-8.803141 41.9399785,-8.803177 41.9399785,-8.803361 41.9399785,-8.803377 41.9399785,-8.803361 41.939978,-8.803361 41.939852,-8.803171 41.938837,-8.803149 41.938652,-8.805153 41.938836,-8.805169 41.938809,-8.805046 41.93921,-8.80447 41.93988,-8.804497 41.93888,-8.80449 41.93888,-8.80449 41.93988,-8.804497	"Uva - Albariño"^^ <http: 2001="" td="" www.w3.org="" xmlschema<=""><td> ber a#string> 2017</td><td><u>gin_date</u> en .5-03- 2015 100:00:00 01Tr</td><td>5-11- 00:00:00</td><td>campa</td></http:>	ber a#string> 2017	<u>gin_date</u> en .5-03- 2015 100:00:00 01Tr	5-11- 00:00:00	campa
This document is part of a project that has received funding from the European Union's Horizon 2020 research and innovation program under agreement No 732064. It is the property of the DataBio consortium a reproduced without the formal approval of the DataBio Management Comr	http://w3id.org/foodie/core/ManagementZone/33	"C2"^^ <http: 2001="" www.w3.org="" xmlschema#string=""></http:>	http://w3id.org/foodie/core/CropType/1	POLYGON((-8.80489 41.939885, -8.803117 41.940124, -8.803141 41.940024, -8.803243 41.939917, -8.803316 41.939785, -8.803377 41.939248, -8.803361 41.939248, -8.803241 41.938652, -8.805158 41.938365, -8.805158 41.938365, -8.805158 41.938855, -8.805158 41.93885, -8.805169 41.93921, -8.8054947 41.93951, -8.80489 41.93985, -8.80489	"Uva - Albariño"^^ <http: 2001="" td="" www.w3.org="" xmlschema<=""><td>201 a#string> 017</td><td>.5-03- 2015 100:00:00 01Tr</td><td>5-11- 00:00:00 <u>h</u></td><td>ttp://w3id.org/foodie/c</td></http:>	201 a#string> 017	.5-03- 2015 100:00:00 01Tr	5-11- 00:00:00 <u>h</u>	ttp://w3id.org/foodie/c

• Faceted search endpoint: <u>https://www.foodie-cloud.org/fct</u>

Text Search	Entity Label Lookup Enti	ty URI Lookup	Featured Demo Queries /	About		
	Precisio	n Search & Find				
Search Text ManagementZone #1			About: <u>ManagementZone #1</u> <u>Goto Sponge NotDistinct Permalink</u> An Entity of Type : <u>http://foodie-cloud.com/model/foodie#ManagementZone</u> , within Data Space : <u>www.foodie-cloud.org</u> Type: <u>ManagementZone</u> Command: <u>Start New Facet</u> Go			
	Hint: You can add this engine in s	earch har of an OpenSearch ext&sid=300	- canahla hrawcar		Attributes rdf:type	Values ManagementZone
					rdfs:label	ManagementZone #1
C OPENLINK S O F T W A R E					ogcgs:hasGeometry	http://w3id.org/foodie/core/ManagementZone/1/geometry
Displaying Ranked Entity N	lames and Text summaries where:				<u>code</u>	CODA1
<u>?s1</u> has <u>any Attribute</u> w	ith Value "ManagementZone #1" Drop.				creationDateTime	2015-12-01 00:00:00(<u>xsd:dateTime</u>)
View query as SPARQL Facet	<u>: permalink</u>		Go to:	Show 20 \$1 - 20 of 37 total	cropSpecies	crop_type #21
Entity http://w3id.org/foodre/N	ManagementZone/5	Title Na ManagementZone #5 http://docs.org/abs/1000000000000000000000000000000000000	amed Graph tp://w3id.org/foodie/core/es#	ManagementZone	holdingZone	<u>Plot #1</u>
http://w3id.org/foode/M http://w3id.org/foode/M http://w3id.org/foode/M	anagementZone/22 anagementZone/13 anagementZone/34	ManagementZone #22 https://document.com/second/se	<u>tp://w3id.org/foodie/core/es#</u> tp://w3id.org/foodie/core/es# ttp://w3id.org/foodie/core/es#	ManagementZone 22. ManagementZone 13. ManagementZone 34.	originType	OriginTypeValue #1
http://w3id.org/foode/M http://w3id.org/foode/M http://w3id.org/foode/M	anagementZone/25 anagementZone/32 anagementZone/16 anagementZone/32	ManagementZone #25 ht ManagementZone #32 ht ManagementZone #16 ht	tp://w3id.org/foodie/core/es# tp://w3id.org/foodie/core/es# tp://w3id.org/foodie/core/es# tp://w3id.org/foodie/core/es#	ManagementZone 25. ManagementZone 32. ManagementZone 16. ManagementZone 23	<u>zoneAlert</u>	<u>Alert #1</u> <u>Alert #5</u>
http://w3id.org/foode/M http://w3id.org/foode/M http://w3id.org/foode/M	anagementZone/12 anagementZone/19 AnagementZone/19	ManagementZone #11	tp://w3id.org/foodie/core/es# tp://w3id.org/foodie/core/es# ttp://w3id.org/foodie/core/es#	ManagementZone 11. ManagementZone 19.	http://foodie-clouanagementZoneName	A1
http://w3id.org/foodre/M http://w3id.org/foodre/M http://w3id.org/foodre/M	anagementZone/21 4anagementZone/4 apagementZone/29	ManagementZone #1 ht ManagementZone #21 ht ManagementZone #29 ht	tp://w3id.org/foodie/core/es# ttp://w3id.org/foodie/core/es# ttp://w3id.org/foodie/core/es#	ManagementZone 21. ManagementZone 21. ManagementZone 29	is <u>alertZone</u> of	Alert #1 Alert #5
http://waid.org/rodde/M http://waid.org/rodde/M http://waid.org/foodre/M	anagementZone/2 4anagementZone/2 #ManagementZone	ManagementZone #22 ht ManagementZone #2 ht ManagementZone #2 ht	tp://w3id.org/foodie/core/es# ttp://w3id.org/foodie/core/es# ttp://w3id.org/foodie/core/es#	ManagementZone 29. ManagementZone 12. ManagementZone	is <u>interventionZone</u> of	Intervention #116
http://w3id.org/foode/M http://w3id.org/foode/M http://w3id.org/foode/M	anagementZone/27 anagementZone/10 anagementZone/18	ManagementZone #18	tp://w3id.org/foodie/core/es# tp://w3id.org/foodie/core/es# ttp://w3id.org/foodie/core/es#	ManagementZone 27. ManagementZone 10. ManagementZone 18	is <u>http://foodie-clouypeManagementZone</u> of	F ProductionType #1
11.127/14/2010/19/1000	anagement2010/10	managementzone #10 <u>nt</u>	<u></u>	nanagementzone 10.	is <u>http://toodie-clouoseManagementZone</u> 01 -	StationClose #35

Go to: Show 20 \$1 - 20 of 37 tota

Complete result - 37 processed in 14 msec. Resource utilization: 1.464K rnd 6 seq 1.19K same seg 20 same pg 40 same par 0 disk 0 spec disk 0B / 0 messages 0 fork is http://foodie-clou...gonManagementZone of ZonePolygon#1

• Map visualisation: http://ng.hslayers.org/examples/foodie-zones/

family

type:

Vitacea

http://foodie-

cloud.com/model/foodie#Crop

Map visualisation: <u>http://ng.hslayers.org/examples/foodie-zones/</u>

This document is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under agreement No 732064. It is the property of the DataBio consortium and shall not be distributed or reproduced without the formal approval of the DataBio Management Committee. Find us at www.databio.eu.

Point of interest

http://w3id.org/foodie/corine/clc_12/71222

http://w3id.org/foodie/corine/clc_12/71444

http://w3id.org/foodie/corine/clc_12/71319

• Metaphactory: <u>https://foodie.grapphs.com/resource/Start</u>

Welcome to DATABIO-FOODIE metaphactory

The metaphactory helps you to navigate and to visualize DATABIO-FOODIE knowledge graphs, some transformed from relational datasets, and some coming from relevant open datasets. These include: Open Land Use, Open Transport Map, Smart Point of Interest, EU NUTS classification, Corine Land Cover, Urban Atlas, Hilucs classification, Eurovoc, Agrovoc, Emergel and others

tempranillo blanco

Uva - Tempranillo blanco (http://foodie-cloud.com/model/foodie#CropType)

Tempranillo blanco (http://foodie-cloud.com/model/foodie#CropType)

See some maps belows: ement Zones in Terras Gauda (Production 2015)

This document is part of a p

Points of interests in Poznan (Stare Miasto)
 Points of interests in Prague (Staré Měste)

from the European Union's Horizon 2020 research and innovation programme under agreement No 732064. It is the property of the DataBio consortium and shall not be distributed or reproduced without the formal approval of the DataBio Management Committee. Find us at www.databio.eu.

Metaphactory: <u>https://foodie.grapphs.com/resource/Start</u>

ManagementZone #2

URI: http://w3id.org/foodie/core/ManagementZone/2

Type: ManagementZone

Graph View Statements

Outgoing Statements

Filter Results		
Property	Object	Named Graph
type	ManagementZone	http://w3id.org/foodie/core/es#
label	ManagementZone #2	http://w3id.org/foodie/core/es#
hasGeometry	geometry	http://w3id.org/foodie/core/es#
code	CODA2	http://w3id.org/foodie/core/es#
creationDateTime	2015-12-01T00:00:00	http://w3id.org/foodie/core/es#
cropSpecies	crop_type #18	http://w3id.org/foodie/core/es#
holdingZone	Plot #1	http://w3id.org/foodie/core/es#
originType	OriginTypeValue #1	http://w3id.org/foodie/core/es#
zoneAlert	Alert #2	http://w3id.org/foodie/core/es#
zoneAlert	Alert #36	http://w3id.org/foodie/core/es#

Incoming Statements

Thank you for your attention!

Contact details

This document is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under agreement No 732064. It is the property of the DataBio consortium and shall not be distributed or reproduced without the formal approval of the DataBio Management Committee. <u>Find us at www.databio.eu</u>.