Available online at www.centmapress.org

Proceedings in
System Dynamics and Innovation in Food Networks 2016

INTERNATIONAL JOURNAL ON FOOD SYSTEM DYNAMICS

HortiCube: A Platform for Transparent, Trusted Data Sharing in the Food Supply Chain

Jack Verhoosel[1], Michael van Bekkum[1] and Tim Verwaart[2]

[1] *TNO Data Science, Kampweg 5, Soesterberg, The Netherlands*
[2] *Wageningen LEI, Alexanderplein 15, The Hague, The Netherlands*
Jack.verhoosel@tno.nl

ABSTRACT

Food supply chains consist of many links and operate on a global scale with many stakeholders involved from farm to fork. Each stakeholder maintains data about food products that they handle, but this data is not transparently available to all other stakeholders in the chain due to various reasons. Trust and reciprocity for data sharing is limited and there is insufficient clarity in data ownership and possible legal consequences. However, various stakeholders could benefit from making data available across the supply chain. Food producers are very interested in consumer demands and trends. Growers also want to guide their supply based on the potential demand for specific food products in the near future. In addition, there are various other data sources that contain interesting data for these same stakeholders, such as import/export transactions, production (forecast) data, parcel crop information, local weather predictions and social media streams. To make all stakeholders in the food chain benefit from these data sources and to share data more transparently, the Dutch horticulture and food domain is developing the HortiCube platform via which various data sources are made accessible to application developers using a secure, linked data application interface. This paper describes the development of the HortiCube, the technologies being used and the lessons learned on aligning semantics by mapping of terms and the implementation performance using an example demo application.

Keywords: *HortiCube, data sharing, trust and security, semantic alignment*

1 Context, problem and approach

Food supply chains consist of many links and operate on a global scale with many stakeholders involved from farm to fork. Each stakeholder maintains data about food products that they handle, but this data is not transparently available to all other stakeholders in the chain due to various reasons. Trust and reciprocity for data sharing is limited because of fear of sensitive data becoming available to competitors and the risk of negative publicity that can have impact on trade and export. In addition, government does not easily share data that it maintains on national and international food production and import/export because of insufficient clarity in data ownership and possible legal consequences.

Although the sector currently operates under this limited transparency of data, various stakeholders could benefit from making data available across the supply chain. Food producers are very interested in how consumers in which countries eventually consume their products and what consumer's demands and trends are. Growers also want to guide their supply based on the potential demand for specific food products in the near future. If such data is available to them, they can optimize their production. Traders of food products are oriented towards transactions and trading and in general only want to open up part of their data to a few trusted partners. Thus, a clear incentive to open up their data for others should be defined.

In addition to data being maintained by the stakeholders in the food chain, there are various other data sources that contain interesting data for these same stakeholders. Examples include import/export transactions, production (forecast) data, parcel crop information, local weather predictions and social media streams. Most of these data sources, however interesting, are not easily accessible on-line for

DOI 2016: pfsd.2016.1642

continuous, real-time usage and automated consumption by growers and other stakeholders in the chain.

In conclusion, there are clear benefits for all stakeholders in the food chain to share data more transparently and gain access to thus far untapped data sources. The key enablers however are ease of access and security and trust. In order to address these issues, the Dutch horticulture and food domain is developing the HortiCube platform via which various data sources are made accessible to application developers using a secure, linked data application interface (see Figure 1).

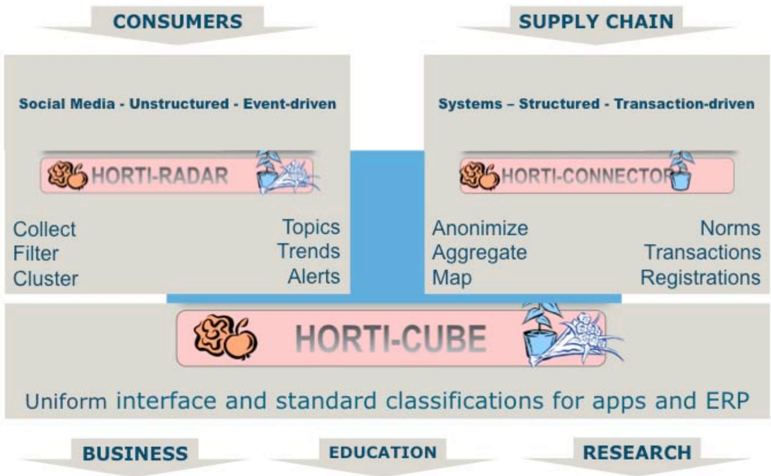

Figure 1. The HortiCube platform high-level architecture.

2 The HortiCube platform

The HortiCube platform is an infrastructure that enables automated access to a variety of data sources by IT systems. Via the HortiConnector, structured transaction-oriented data sources are incorporated, such as import/export data from stakeholders but also open data on weather conditions. Via the HortiRadar, unstructured event-driven data sources are linked into the platform, such as social media sources from consumers. These data sources are made accessible via linked data web-based mechanisms. In addition, the platform incorporates security mechanisms to ensure that each stakeholder that has added a data source remains in control of who gets access to the stakeholders' data. Furthermore, anonymization and aggregation rules are applied to data in order to ensure the privacy of the specific stakeholders when applicable.

At first, the HortiCube focuses on data sources that contain or are closely related to market information. Using this information, the stakeholders in the food supply chain can enhance their market orientation and information-based supply planning. In addition, by combining data sources, the HortiCube allows for applications of big data analysis, pattern searching and thus better decision-making on when to produce what quantities of which food products.

When combining data from different sources, semantics becomes an important aspect of this platform because the meaning of similar terms in different data sources needs to be aligned. A specific challenge in this context is the alignment of the semantics of food product identification, because different product coding lists and levels of product aggregation are in use along the food supply chain. Apart from the generic GPC product coding of GS1, there are various local and sectorial product classifications in use. The Horticube platform provides semantic, ontological mappings between these product classifications (see Figure 2).

DOI 2016: pfsd.2016.1642

hoofdgroep	groep	hoofdvarieteit	KCB code	KCB naam MIT database	KCB naam GroentenFruit Huis	WAPA NH Variety	WAPA ZH Variety	GPC Value Code	GPC Value Title
Hardfruit	Appelen		2500	Appelen, alle soorten	**Appel, alle soorten**	Total	Total	30000881	DESSERT
Hardfruit	Appelen		2525	Appelen, diversen	Appel, overige	Other	Other	30002515	UNCLASSIFIED
Hardfruit	Appelen							30002518	UNIDENTIFIED
Hardfruit	Appelen							30000720	COMBINATION
Hardfruit	Appelen	Alkmene						30015000	ALKMENE
Hardfruit	Appelen	Alkmene Cevaal							
Hardfruit	Appelen	Altess							
Hardfruit	Appelen	Appelmix							
Hardfruit	Appelen	Appels (diversen)							
Hardfruit	Appelen	Bellefleur						30015018	BELLE FLEUR DOUBLE
Hardfruit	Appelen	Benoni	2510	Appelen, benoni	Appel, Benoni				
Hardfruit	Appelen	Bloemee zoet							
Hardfruit	Appelen	Braeburn	2517	Appelen, breaburn	Appel, Braeburn	Braeburn	Braeburn	30015026	BRAEBURN
Hardfruit	Appelen	Campagne zoet							
Hardfruit	Appelen	Canada Grise				Reinette Grise du Canada		30015216	REINETTE GRISE DU CANADA
Hardfruit	Appelen	Civni	2571	Appelen, rubens	Appel, Rubens			30015036	CIVNI
Hardfruit	Appelen	Collina							
Hardfruit	Appelen	Contento							
Hardfruit	Appelen	Cox	2520	Appelen, cox's orange	Appel, Cox's Orange Pippin	Cox Orange		30015039	COX'S ORANGE PIPPIN
Hardfruit	Appelen	Cripps Pink				Cripps Pink	Crippspink	30015041	CRIPPS PINK
Hardfruit	Appelen	Crown Apple							
Hardfruit	Appelen	Dalinbel							
Hardfruit	Appelen	Dalinco							
Hardfruit	Appelen	Delbarjubilee							
Hardfruit	Appelen	Delblush						30015052	DELBLUSH
Hardfruit	Appelen	Delcorf						30015053	DELCORF
Hardfruit	Appelen	Dijkmans						30015062	DYKMANNS ZOET
Hardfruit	Appelen	Discovery	2522	Appelen, discovery	Appel, Discovery			30015059	DISCOVERY
Hardfruit	Appelen	Diversen (hardfruit)							
Hardfruit	Appelen	Early Scarlet						30015817	EARLY
Hardfruit	Appelen	Elan						30015064	ELAN
Hardfruit	Appelen	Elise						30015065	ELISE
Hardfruit	Appelen	Elsmi							
Hardfruit	Appelen	Elstar	2530	Appelen, elstar	Appel, Elstar	Elstar		30015068	ELSTAR
Hardfruit	Appelen	Elstar *	2530	Appelen, elstar	Appel, Elstar	Elstar		30015197	RED ELSTAR

Figure 2. Example in tabular form of a mapping between product classifications of apples.

As is apparent from the figure, the product classifications differ in names, codes as well as languages used. In addition to a mapping ontology for product classifications, the platform defines an ontology for each data source that is incorporated and provides a mapping between this ontology and a common ontology for the horticulture domain. As a result, the HortiCube can offer a uniform interface with standard terms and classifications to application developers. Applications developed on top of the HortiCube use the semantically rich, linked data interface to access the data in the platform and support decision-making. This should give a boost to applications that make use of and combine market data and should minimize investment for individual stakeholders doing so.

3 Apple/pear use case

In order to show the feasibility of the HortiCube, we have defined a use case in the fruit sector, specifically on apple/pear products (see Figure 3). In this use case, we have made linked-data interfaces with ontologies on four important data sources for the apple/pear sector:

- WAPA[*], which contains the yearly apple/pear production and forecast figures per EU country over the years 2006 to 2015.
- KCB[†], which contains the accumulated value of export transactions of vegetables/fruit from The Netherlands to other countries over the years 2010 to 2014.
- RVO[‡], which contains areal information on grown crop for all open ground parcels in The Netherlands for 2014 and 2015.
- KNMI[§], which contains daily weather information for all measuring stations in The Netherlands.

[*] World Apple and Pear Association
[†] Dutch Quality Control Bureau
[‡] Federal Organization for Dutch Companies
[§] Dutch National Weather Measurement Institute

DOI 2016: pfsd.2016.1642

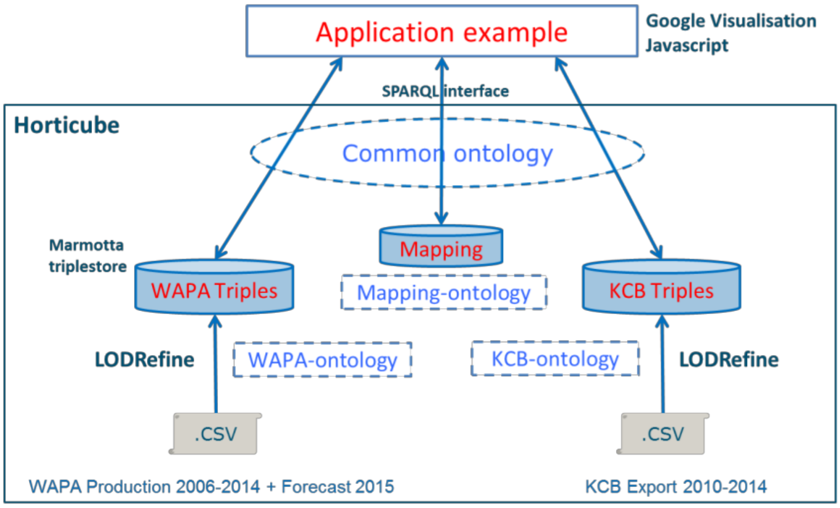

Figure 3. The HortiCube for a use case on apple/pear production and export data.

Each of the data sources was initially delivered in CSV-format. We used the OpenRefine[**] tool to import and clean the data. Based on the structure of the data we used the TopBraid[††] composer tool to model an ontology that captures the semantics of the data source. This ontology was then used to generate a set of RDF triples with the LOD-plugin of the OpenRefine tool. This resulted in a data source in RDF-format that was stored in an Apache Marmotta[‡‡] triple store. In addition, we have made a mapping between the apple/pear product classifications used in WAPA, KCB, RVO and GPC, so that it is possible to combine the apple/pear data from these sources in a semantically correct way and reason about the combined data. This mapping has been modelled in an ontology and stored into the Apache Marmotta triple store.

The interface to the HortiCube in this use case is formed by the SPARQL engine of the Marmotta triple store that enables the querying of each data source in terms of the common ontology. The query results are delivered in table and in JSON-format. We have built an example application on top of this interface that provides answers to the following important information questions apple/pear growers may have, using the data from these sources:

- What is the trend in apple/pear production per country and what are the consequences for the export distribution to other countries?
- What is the relation between monthly apple/pear export and local weather conditions in this time period?

Figure 4 shows the result of a question to the application on the production data of apples in two countries over the years 2006 to 2014.

The example application provides added value to both growers and traders in the fruit sector. The availability of the HortiCube platform is therefore both an incentive and an enabler for them to share their production and export data. Access to this application is given only to those that can electronically identify themselves to have certified access to the platform.

[**] http://www.openrefine.org
[††] http://www.topquadrant.com/tools/ide-topbraid-composer-maestro-edition/
[‡‡] http://marmotta.apache.org

DOI 2016: pfsd.2016.1642

Result:

Variety	Productiongroup1	Productiongroup2
Boskoop	166	136
Cox Orange	11	21
Elstar	1328	82
Golden Delicious	192	268
Jonagold	1017	1260
Other	321	199
Other new varieties	187	102

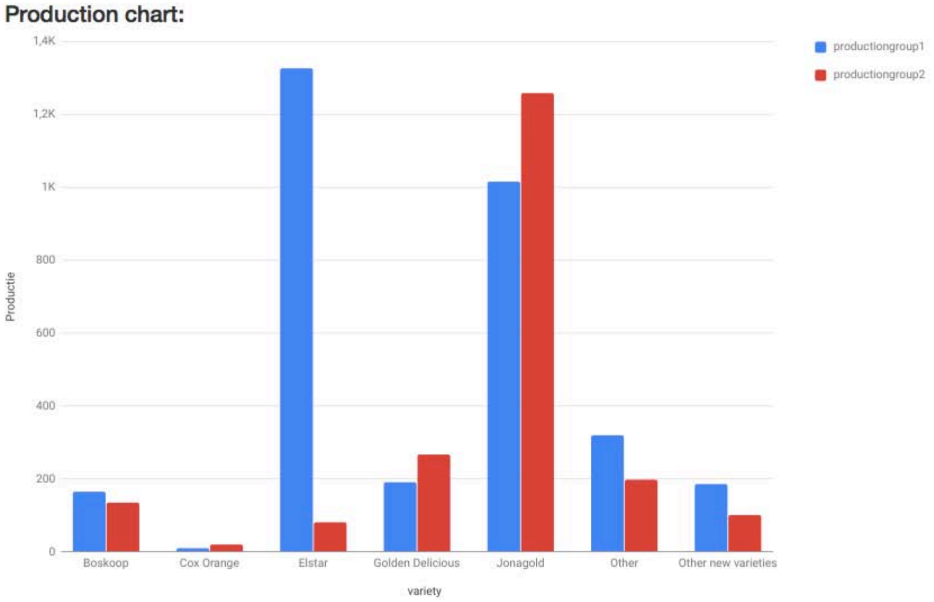

Figure 4. Output of the apple/pear example application on production data in two countries.

4 Lessons learned and future work

The HortiCube platform enables applications to be developed with unambiguous access to a variety of linked data sources of horticulture market information in a secure and transparent manner. It provides opportunities for big data analysis and better decision-making by stakeholders along the food supply chain. The use of ontologies and linked data mechanisms supports a much larger variety of questions to be answered in a semantically uniform and correct way.

Mapping of product classifications is a laborious activity, but once done it can be reused many times for alignment of various data sources. Opening up data sources and making them linkable via the HortiCube is feasible in a relatively small amount of time. Updating these data sources therefore becomes possible, especially when the incorporation step can be mostly automated. Scalability might become an issue that needs attention when large data sources are added and generic questions, like requesting all the data, are possible. In that case, intelligent performance mechanisms might be incorporated such as linked data fragments. The data sources and example application that are discussed in this paper did not show any performance issues and the response time of the HortiCube is well within boundaries.

The security mechanism now incorporated is still fairly basic, so further work needs to be done on the application of mechanisms for role-based or user-specific accessibility per specific data source or even sets of data elements. Data sharing mechanisms that allow fine granularity of access to data elements and guarantee strict maintenance of both ownership and security need to be put in place. For instance, a role mechanism might be introduced in the future to enable more complex accessibility rules on the users of the HortiCube.

DOI 2016: pfsd.2016.1642