



RDF Validation requirements for data about products, services and companies

Mark Harrison (University of Cambridge)
Robert Beideman (GS1 Global Office)



mark.harrison@cantab.net
robert.beideman@gs1.org



- Global not-for-profit community organization developing user-driven open standards since 1973:
 - Bar codes and product identifiers, RFID
 - Supply chain visibility
 - Product master data exchange, product categories



- Global network of research labs working together since 1999/2000 and with GS1 in development of:
 - Technology to support an “Internet of Things”
 - Supply chain visibility, security, efficiency
 - Next generation networks, sensors, etc.
- MIT, University of Cambridge, ETH Zurich / St Gallen, Keio University, KAIST, Fudan University, University of Adelaide



GS1 Digital:

Using technology such as Linked Open Data (RDF etc.) to make trustworthy data about products and services available on the web

To improve visibility of products and suppliers

To enable enhanced consumer interactivity with products and support new services / applications



LOD for products – Pre-Sale



- ① User types in a **keyword** to specify type of product
- ② Keyword is matched in GPC hierarchy and appropriate **contextual filters** are displayed to refine the search
- ③ **Additional constraints** can be specified by the user
e.g. available **budget / price** range
urgency (how long to wait?)
proximity (how far away?)

SEARCH

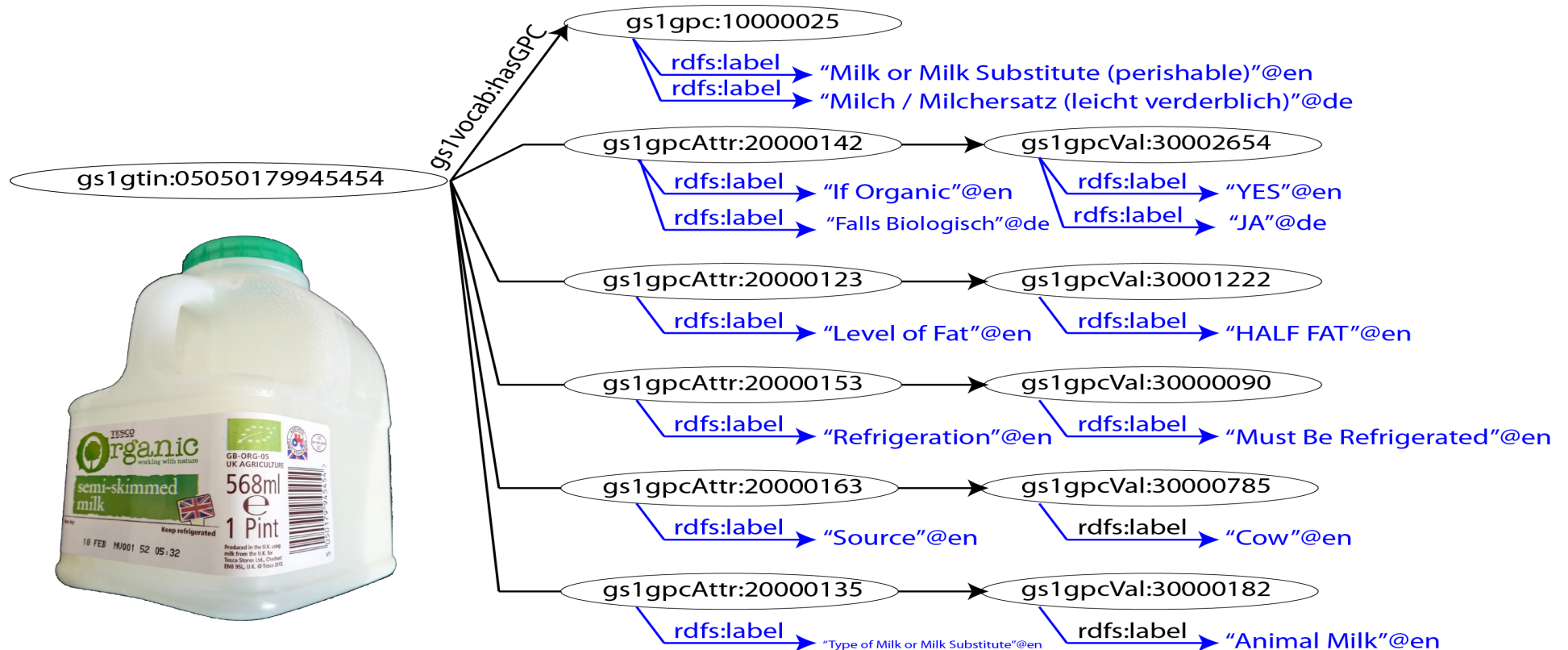


Search for Products & Services



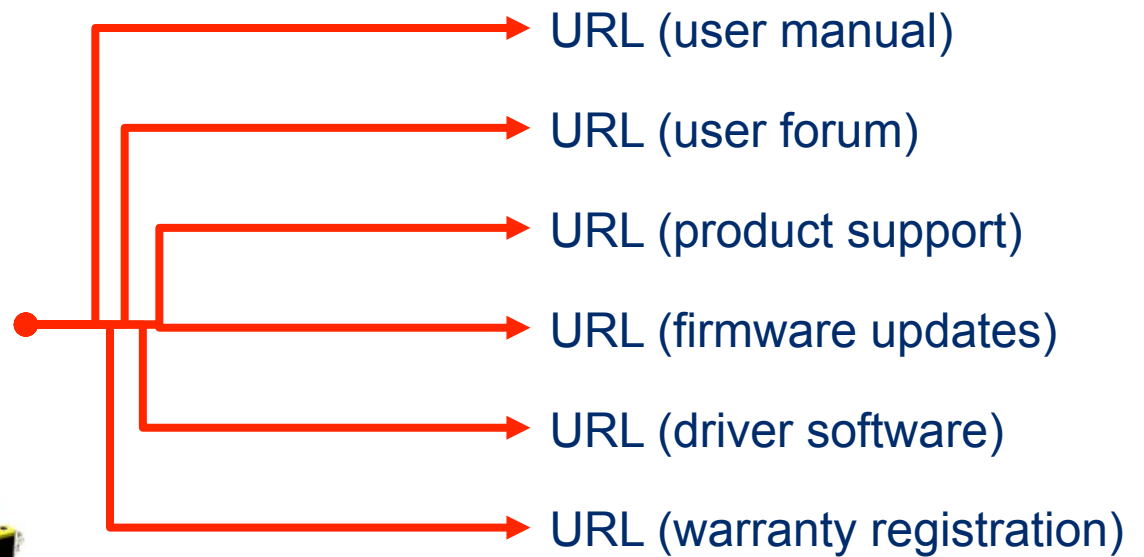


Global Product Classification as LOD





LOD for products – Post-Sale



Sourcing of structured master data



Weetabix Organic

- Nutritional Information
- Ingredients
- Allergy advice
- Stockist information

The old adage, "If it's not broken, don't fix it", could have been created especially for our beloved Weetabix. *The UK's favourite breakfast cereal recipe* has remained essentially the same since day one. And why shouldn't it? It's popular with everyone, young and old alike, and because Weetabix is made from

Ingredients

Weetabix contains the following ingredients (listed in order of amount used).

Wholegrain Wheat (95%), Malted Barley Extract, Sugar, Salt, Niacin, Iron, Riboflavin (B2), Thiamin (B1), Folic Acid.

[Back to top](#)

Allergy Advice

Information for special diets and for people with allergies.

Nutritional Information

NUTRITION	PER 37.5g SERVING	100g
Energy (Kilojoules)	568kJ	1516kJ
Energy (Calories)	134kcal	358kcal
Protein	4.3g	11.5g
Carbohydrate	25.7g	68.6g
(of which sugars)	1.7g	4.4g
Fat	0.8g	2g
(of which saturates)	0.2g	0.6g
Fibre	3.8g	10g
Sodium	0.10g	0.26g
Salt	0.24g	0.65g

TESCO

Home > Food & Drink > Fresh Food > Cereals > Weetabix Organic Whole Grain 24PK

Weetabix Organic Whole Grain 24PK

£3.09 (12.50p/box)

Quantity: 1 Add

Product Details

A serving contains the following of your guideline daily amount

	Carbs	Fat	Fibre	Protein	Salt	Sugar
%	7%	2%	1%	1%	4%	1%

Organic

With 100% wholegrain

Low fat & sugar

High fibre

Suitable for vegetarians

Organic whole wheat cereal

Fuel your day...

Weetabix packed with wholegrain + fresh milk = slow release energy*

Our organic ingredients are grown using only natural methods with no artificial chemical fertilisers, herbicides or pesticides carefully cultivated for their taste, giving you the energy you need to take on the day.

*2 Weetabix with 125ml semi skimmed milk has a low Glycaemic Index of 47, helping keep you fuller for longer releasing its energy as part of a healthy diet and lifestyle.

Allergy Information

Contains: Barley, Wheat

Other Information

Recycling Info: Recyclable Pack

Recycling Info: Not Recyclable Film

Ingredients

Organic Wholegrain Wheat (95%), Organic Malted Barley Extract, Organic Sugar, Salt

Sainsbury's

Home > Food & Drink > Fresh Food > Cereals > Weetabix Organic Cereal x24

Weetabix Organic Cereal x24

£2.99/unit (12.50p/box)

Item code: 3112913

Description

Organic

With 100% wholegrain

Low fat & sugar

High fibre

Suitable for vegetarians

Organic whole wheat cereal

Fuel your day...

Weetabix packed with wholegrain + fresh milk = slow release energy*

Our organic ingredients are grown using only natural methods with no artificial chemical fertilisers, herbicides or pesticides carefully cultivated for their taste, giving you the energy you need to take on the day.

*2 Weetabix with 125ml semi skimmed milk has a low Glycaemic Index of 47, helping keep you fuller for longer releasing its energy as part of a healthy diet and lifestyle.

The pack contains twelve 2-biscuit servings.

Weetabix Organic is made from agricultural ingredients certified organically grown according to standards set by the Soil Association.

GB-ORG-05

EU Agriculture

www.organicfood.com

Box - Card - widely recycled

Box - Plastic - not currently recycled

Source - 95% recycled card

Box - 95% recycled card

Soil Association Organic (20 Organic Certification 5)

Ingredients

Organic Wholegrain Wheat (95%), Organic Malted Barley Extract, Organic Sugar, Salt

Other Information

Organic

Suitable for Vegetarians

Contains: Barley, Wheat

ASDA

Home > Food & Drink > Fresh Food > Cereals > Weetabix Organic

Weetabix Organic

24's

£2.68 (11.12p per heath)

Product Code: 998412

THE WEETABIX FOOD COMPANY. ORGANIC WHOLE WHEAT CEREAL.

Weetabix Organic, at the taste and quality you would expect - naturally bursting with goodness. Our ingredients are grown using only natural methods with no artificial chemical fertilisers, herbicides or pesticides - carefully cultivated for their great taste to make a healthy and satisfying start to the day. Good for you, good for the environment. Weetabix Organic is made from agricultural ingredients certified organically grown according to standards set by the Soil Association. Nestled deep in the heart of the English countryside, Weetabix have been creating delicious breakfast cereals since 1952, using only the finest ingredients. Whenever you see our sign you know you've found the goodness of wholegrain. All our cereals are made with natural wholegrains which provide a valuable source of essential nutrients, ensuring that you get the nutritional value of every part of the grain.

Ingredients

[Organic Wholegrain Wheat (95%), Organic Malted Barley Extract, Organic Sugar, Salt]

Other Information

Allergy Advice

Contains Barley, Contains Wheat.

Dietaries

Organic. Suitable for Vegetarians.

Waitrose

Home > Food & Drink > Fresh Food > Cereals > Weetabix organic 24s

Weetabix organic 24s

£3.09 (12.50p each)

BRAND PRICE MATCH - CHECKED TESCO.COM 28/08/2013

Product Summary

Organic. With 100% wholegrain. Low fat & sugar. High fibre. Suitable for vegetarians. Organic whole wheat cereal. Fuel your day... Weetabix packed with wholegrain + fresh milk = slow release energy*. Our organic ingredients are grown using only natural methods with no artificial chemical fertilisers, herbicides or pesticides - carefully cultivated for their taste, giving you the energy you need to take on the day. *2 Weetabix with 125ml semi skimmed milk has a low Glycaemic Index of 47, helping keep you fuller for longer releasing its energy as part of a healthy diet and lifestyle.

Quick Facts

Contains Barley, Contains Wheat, Organic, Suitable for Vegetarians

Similar products

Heart Health / Wheat

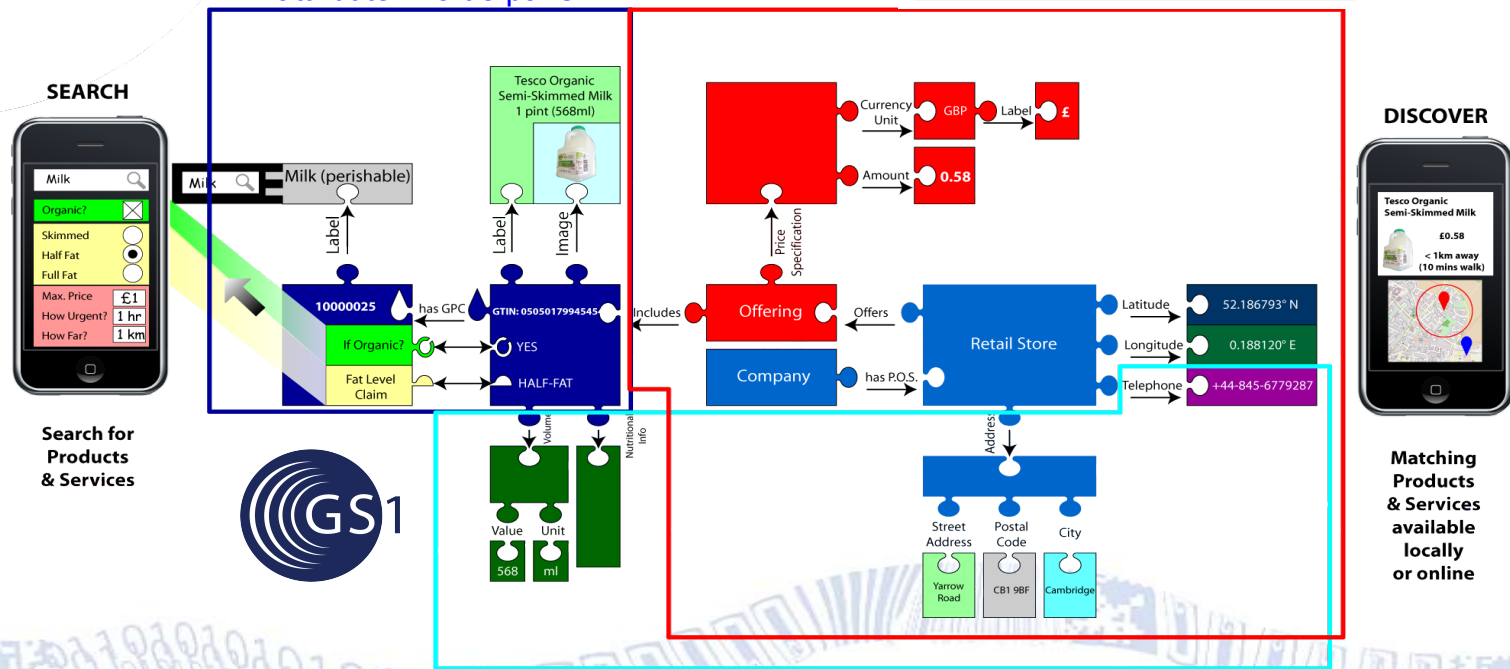


Sourcing of structured master data

GS1

Global Product Classification
(GPC)
+ attribute - value pairs

schema.org





EU Food Labelling legislation

- **EU 1169/2011 Regulation** includes Article 14 on **Distance Selling**
- Mandatory details to be freely available to consumers before purchase
- Grocery websites will typically need to provide info about
 - Ingredients
 - Potential allergens
 - Nutritional information
- Legislation does not require this information to be machine-readable
- However, EU 1169/2011 is driving a major compliance effort to make this kind of information available on manufacturers' / retailers' websites – and there is an **opportunity** to make it available as linked data at the same time, to open up much richer search applications for food products.



Liability issues

- Data needs to be correct, up-to-date, trustworthy
- Units of measure need to be precise, unambiguous
e.g. grams per 100g / 100ml of product
(also micrograms, milligrams per 100g / 100ml)
- Changes to product formulation (ingredients) could introduce new allergens (risk of death) or make a product no longer suitable for those on specific diets (vegan, kosher, etc.)

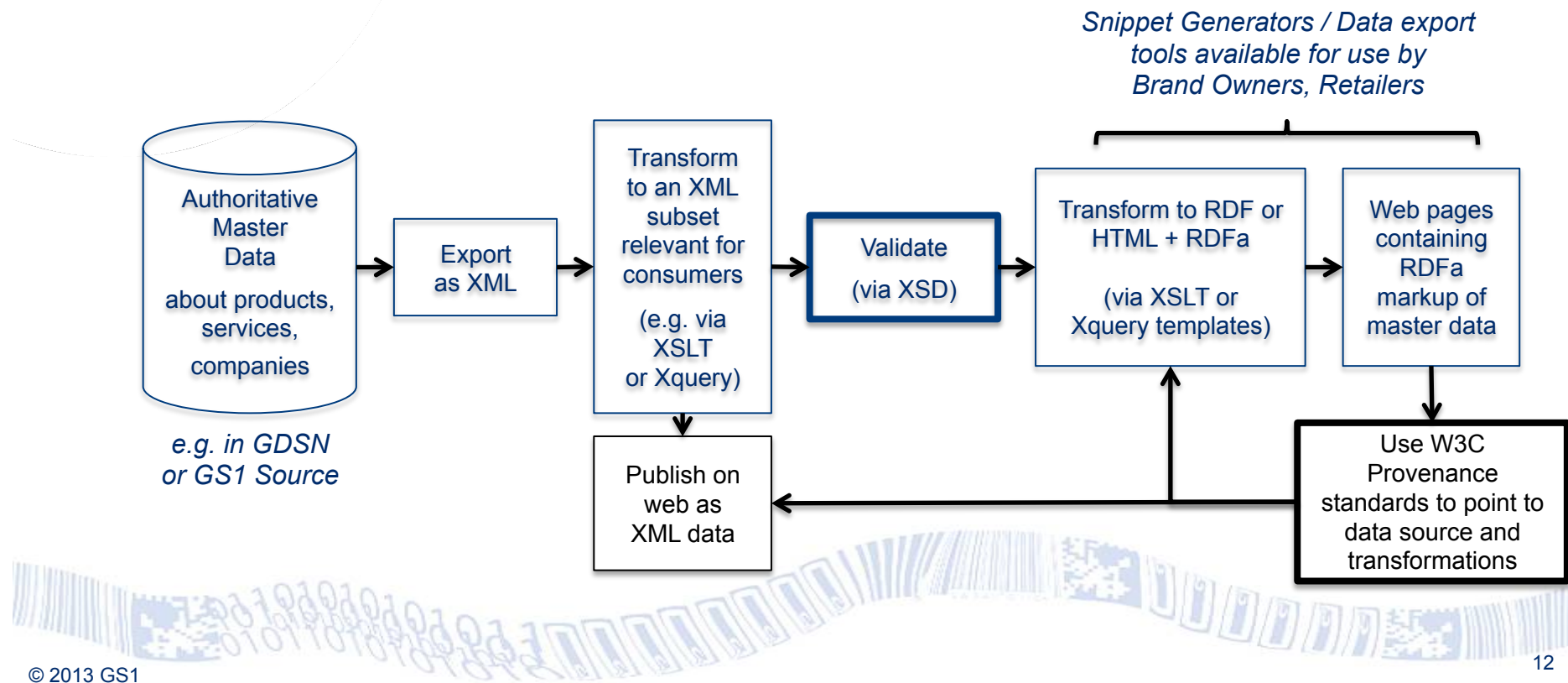


Technical validation requirements

Type of requirement	Example for product master data
Cardinality constraints / Completeness	Are all mandatory attributes are specified? (with non-null values) Which attributes are optional?
Cardinality constraints within complex data structures such as Quantitative Values	Are required units of measure specified precisely?
Range checking for codified values. Values of specific attributes are members of a defined code list or controlled vocabulary	Units of measure or product categories, attributes and values take code values or URIs from a defined code list or controlled vocabulary, where expected.
Checking for current reachability of online resources (when the Object of an RDF triple is a URL)	Use RDF validation to check for broken hyperlinks (e.g. RDF links to product manuals etc.)
Canonicalization of a specified ordered set of RDF triples in order to use digital signatures to detect any deviations from a corresponding set of RDF triples from a source that is considered to be authoritative	Manufacturers might publish master data about products and retailers might replicate this data in their websites. Potential liability issues from inaccurate product information if a product's ingredients list changes and introduces an allergen / unacceptable ingredient



Maybe not for validating source data





Improve markup / validation tools

TESCO [Tesco.com](#) | [My Orders](#) | [My Account](#) | [Sign In](#) | [Help](#)

Groceries | Favourites | Recipes | Special Offers | **Wimbledon** | Delivery Saver

Fresh Food | Bakery | Food Cupboard | Frozen Food | Drinks | Baby | Health & Beauty | Pets | Household | Home & Ents

Groceries > Nestle Cheerios 600G [Back to previous page](#)

Nestle Cheerios 600G

£3.19 (53.17p/100g)

Quantity: 1 [Add](#)

Product Details

A serving contains the following of your guideline daily amount:

Calories	Sugar	Fat	Saturated	Salt
175	12g	3g	2g	0.5g
9%	14%	5%	8%	8%

Deliciously crunchy Os, packed with 4 whole grains. Say Yes to Cheerios.

It's simple, everybody knows that the whole family love Cheerios. The perfect balance of health and taste.

8 Vitamins & Iron. Source of Calcium & High in Fibre.

Check out our Facebook page @ [www.facebook.com/cheeriosUK](#) for exclusive Cheerios content.

Allergy Information

Allergy Other Text: May contain nut traces

Contains: Barley, Oats, Wheat

May Contain: Nuts

Other Information

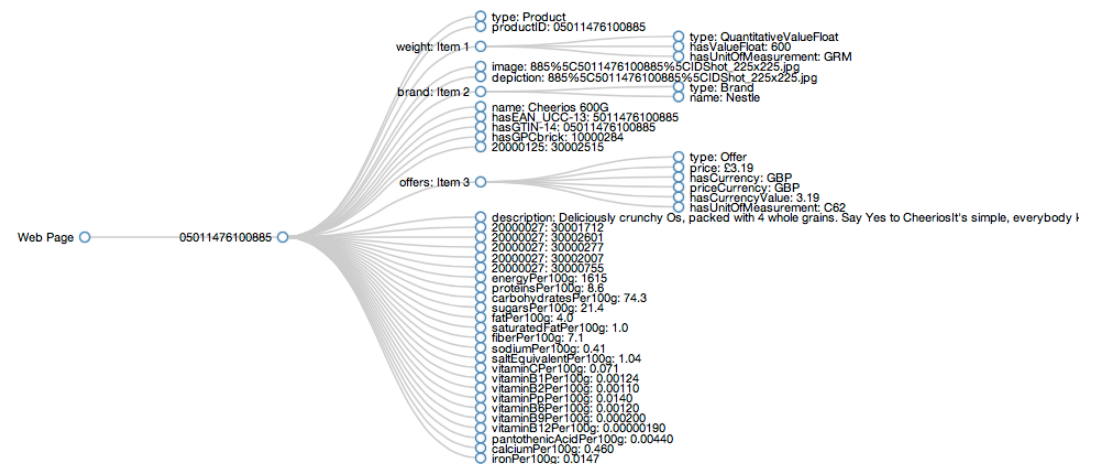
Recycling Info: Recyclable Inner

Recycling Info: Recyclable Carton

Ingredients

Cereal Grains (77.5%) (Whole Grain OATS (38.0%), Whole Grain WHEAT (18.6%), Whole Grain BARLEY (12.8%), Whole Grain Rice (5.5%)), Whole Grain Maize (2.8%), Sugar, Wheat Starch, Partially Inverted Brown Sugar Syrup, Salt, Tripotassium Phosphate, Sunflower Oil, Colours: Caramel, Annatto, Carotene, Antioxidant: Tocopherols, Vitamins and Minerals: Vitamin C, Niacin, Pantothenic Acid, Thiamin (B1), Vitamin B6, Riboflavin (B2), Folic Acid (Folacin), Vitamin B12, Calcium Carbonate, Iron, MAY CONTAIN NUT TRACES. To produce 100g of this product 77.5g of Whole Grain was included in the recipe.

Nutrition	
Typical Values	Typical values per 100g
Energy	1615 kJ
	382 kcal
Protein	8.6g
Carbohydrate	74.3g
of which sugars	21.4g
Fat	4.0g
of which saturates	1.0g
Fibre	7.1g
Sodium	0.41g
Salt Equivalent	1.04g
Vitamins & Minerals %RDA*	
Vitamin C	71.0 mg 89%
Thiamin (B1)	1.24 mg 113%
Riboflavin (B2)	1.10 mg 79%
Niacin	14.0 mg 88%
Vitamin B6	1.20 mg 86%
Folic Acid (Folacin)	200 µg 100%
Vitamin B12	1.90 µg 76%
Pantothenic Acid	4.40 mg 73%
Calcium	480 mg 58%
Iron	14.7 mg 105%



Demo at <http://www.autoidlabs.org.uk/RDFa-test/markup-rdfa.html>

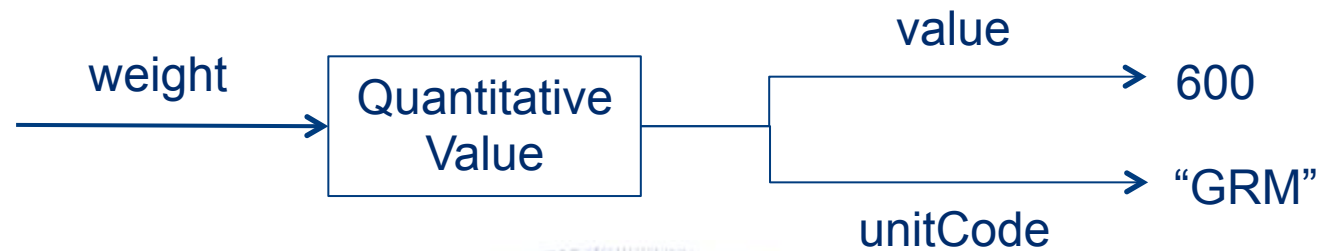
RDF Visualization: <http://rdfa.info/play>



Promote consistent convergent structured data through 'profiling'

*Which vocabularies?
Which terms?*

Promoting correct markup for complex values





Support dynamic contextual search interfaces

Milk

Organic? ☐

Skimmed ☐

Half Fat ☒

Full Fat ☐

Boolean choice

Choose 1 from N

Cardinality / Enumeration constraints can drive UI options for contextual search



Questions / Discussion