

# Core Vocabularies and Grammar for Public Sector Information Interoperability



**Dr Christopher J Harding**

**Director for Interoperability**

**[c.harding@opengroup.org](mailto:c.harding@opengroup.org)**

# Agenda

- The need for a data classification system
- Data classification
- What do we have already?
  - Vocabularies
  - Grammar
- Conclusions

# Agenda

- **The need for a data classification system**
- Data classification
- What do we have already?
  - Vocabularies
  - Grammar
- Conclusions

# “Data is the New Oil”



- Extracting and refining minerals requires work
- So does providing and consuming data

# Information Re-Use Issues

- Even though valuable data is released, additional effort is required to energize external stakeholders to create something with it (Samos).
- Open data, even when freely available, is not free to use since so much time has to be spent cleaning it up, converting it, integrating and maintaining it (Lisbon).
- There is a need to describe the quality of data in a consistent manner if potential consumers are to make informed choices (Timișoara).
- Raw data is of almost no value except to a small number of people with the skills and motivation to work with it. Commercial re-users add value to the raw data by analysis, transformation, and enhancement (Krems).

# Understanding and Analysis



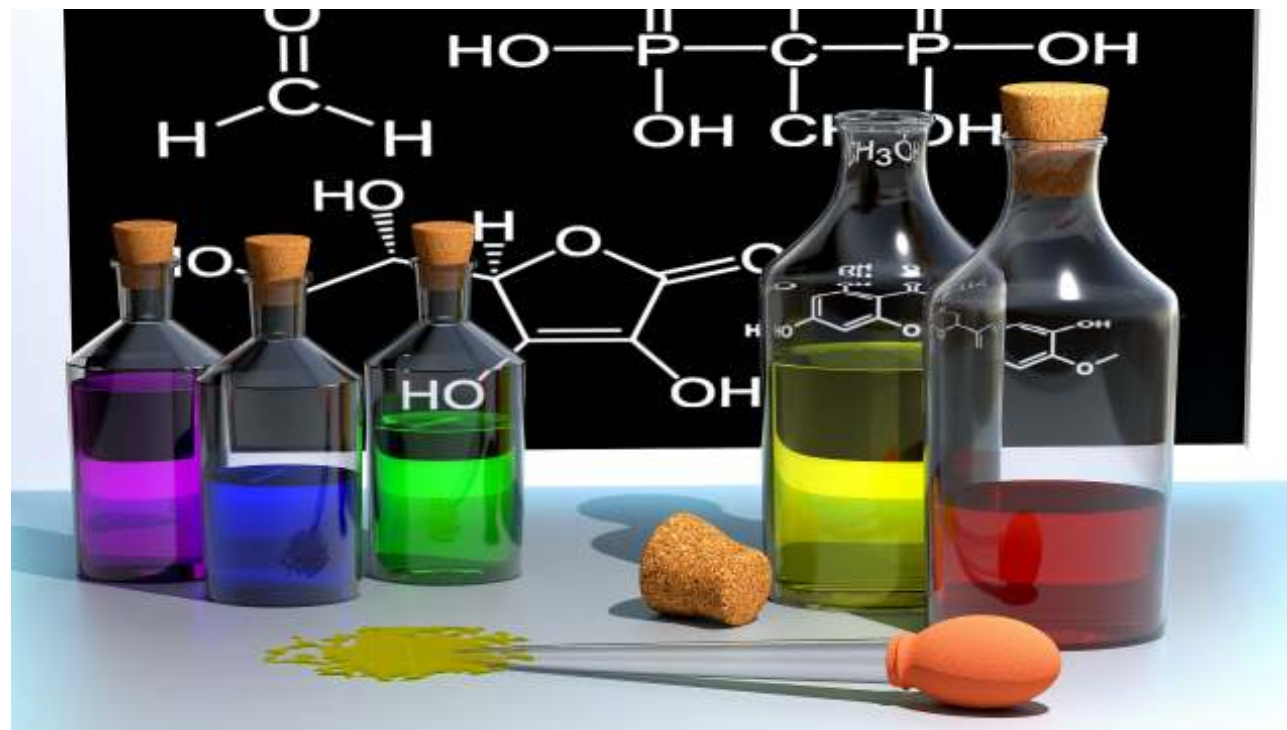
- Understanding the value of minerals in their raw state can be hard
- So can understanding the value of raw data

# Classification Enables Analysis



Copyright © The Open Group 2015

# Classification Enables Analysis





# And Aids Understanding and Use

1 H Hydrogen																	2 He Helium																		
3 Li Lithium	4 Be Beryllium																	5 B Boron	6 C Carbon	7 N Nitrogen	8 O Oxygen	9 F Fluorine	10 Ne Neon												
11 Na Sodium	12 Mg Magnesium																	13 Al Aluminium	14 Si Silicon	15 P Phosphorous	16 S Sulfur	17 Cl Chlorine	18 Ar Argon												
19 K Potassium	20 Ca Calcium	21 Sc Scandium	22 Ti Titanium	23 V Vanadium	24 Cr Chromium	25 Mn Manganese	26 Fe Iron	27 Co Cobalt	28 Ni Nickel	29 Cu Copper	30 Zn Zinc	31 Ga Gallium	32 Ge Germanium	33 As Arsenic	34 Se Selenium	35 Br Bromine	36 Kr Krypton	37 Rb Rubidium	38 Sr Strontium	39 Y Yttrium	40 Zr Zirconium	41 Nb Niobium	42 Mo Molybdenum	43 Tc Technetium	44 Ru Ruthenium	45 Rh Rhodium	46 Pd Palladium	47 Ag Silver	48 Cd Cadmium	49 In Indium	50 Sn Tin	51 Sb Antimony	52 Te Tellurium	53 I Iodine	54 Xe Xenon
55 Cs Caesium	56 Ba Barium	57 La Lanthanum	72 Hf Hafnium	73 Ta Tantalum	74 W Tungsten	75 Re Rhenium	76 Os Osmium	77 Ir Iridium	78 Pt Platinum	79 Au Gold	80 Hg Mercury	81 Tl Thallium	82 Pb Lead	83 Bi Bismuth	84 Po Polonium	85 At Astatine	86 Rn Radon	87 Fr Francium	88 Ra Radium	89 Ac Actinium	104 Rf Rutherfordium	105 Db Dubnium	106 Sg Seaborgium	107 Bh Bohrium	108 Hs Hassium	109 Mt Meitnerium	110 Ds Darmstadtium	111 Rg Roentgenium	112 Uub Ununbium	113 Uut Ununtrium	114 Uuq Ununquadium	115 Uup Ununpentium	116 Uuh Ununhexium	117 Uus Ununseptium	118 Uuo Ununoctium
58 Ce Cerium	59 Pr Praseodymium	60 Nd Neodymium	61 Pm Promethium	62 Sm Samarium	63 Eu Europium	64 Gd Gadolinium	65 Tb Terbium	66 Dy Dysprosium	67 Ho Holmium	68 Er Erbium	69 Tm Thulium	70 Yb Ytterbium	71 Lu Lutetium						90 Th Thorium	91 Pa Protactinium	92 U Uranium	93 Np Neptunium	94 Pu Plutonium	95 Am Americium	96 Cm Curium	97 Bk Berkelium	98 Cf Californium	99 Es Einsteinium	100 Fm Fermium	101 Md Mendelevium	102 No Nobelium	103 Lr Lawrencium			

# We Need a Data Classification System

- To enable analysis
- To aid understanding and use
- To make interoperability possible

# Agenda

- The need for a data classification system
- **Data classification**
- What do we have already?
  - Vocabularies
  - Grammar
- Conclusions

# Data Classification

We want to describe units of data to enable

- Analysis
- Understanding
- Integration

# Data Element

- A unit of data that is considered in context to be indivisible. [ISO 2382]

# Data Element Contexts



Copyright © The Open Group 2015

14

Is the poem the context

- Wordsworth wandered
- Wordsworth was alone
- . . .

Or is it an object in another context?

- Title: I wandered Lonely as a Cloud
- Creator: William Wordsworth
- . . .

THE *Open* GROUP

# Structured Data

familyName	givenName	gender	dateOfBirth
Delarue	Mathieu	M	01-Jan-00
Doe	Jane	F	20-Feb-97
Mustermann	Erika	F	25-Sep-03
Nordmann	Kari	F	18-Dec-11
Rossi	Mario	M	07-Jul-88
Smith	John	M	03-Mar-73

- Relational databases
- Spreadsheets
- Triples
- Messages
- APIs

# Interoperability: An Internet of Things Example

- Smart City public data
- User-developed application
- BioTope will develop more significant examples
  - Domain-specific pilots
  - Cross-domain smart city pilots
  - <http://biotope-h2020.eu>



**Building an IoT OPEN  
innovation Ecosystem for  
connected smart objects**

*bioTope*

THE *Open* GROUP

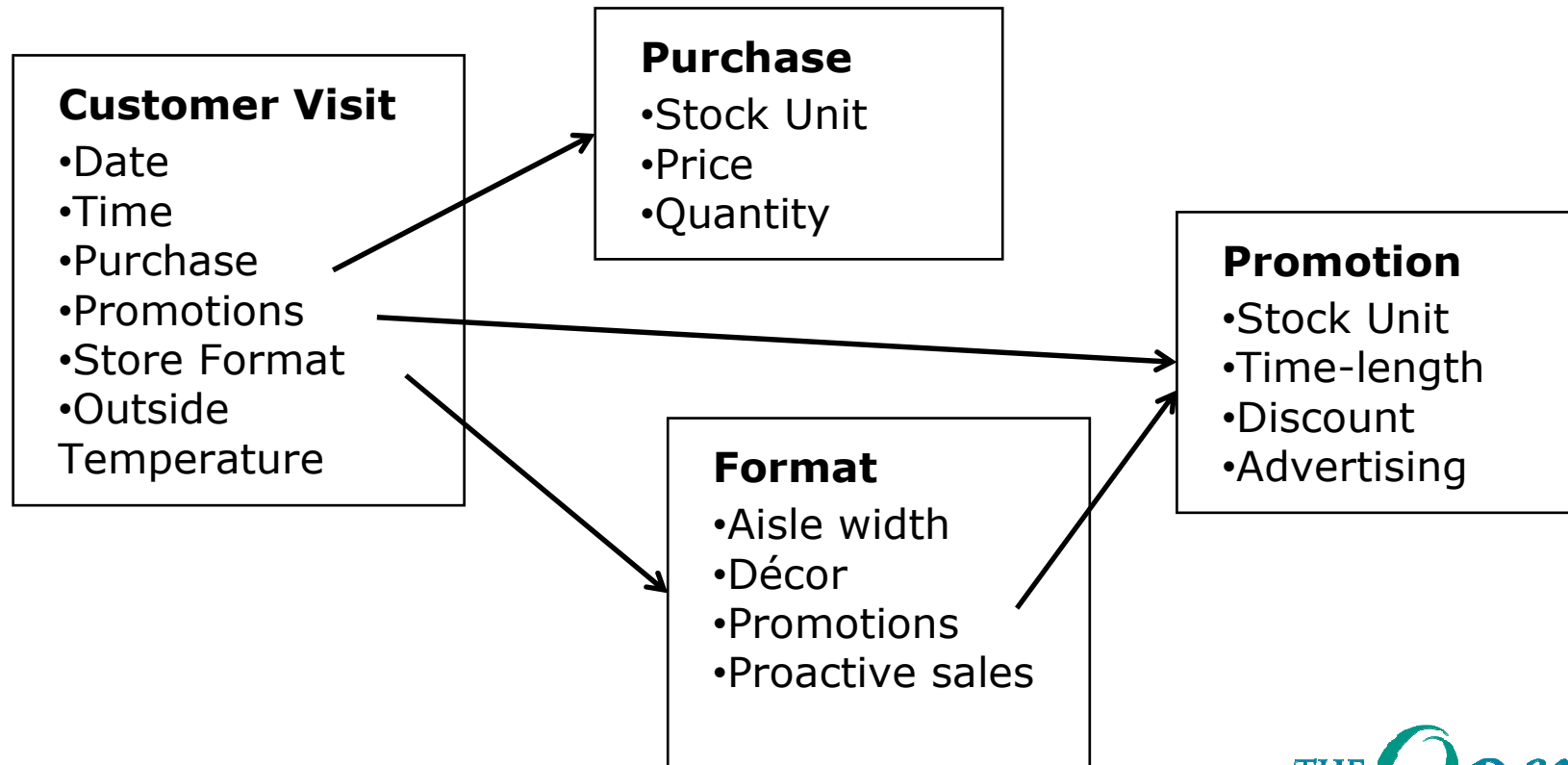


# Example of Data Re-Use



- Analysis of influence of customer experience on purchasing
- Outside temperature is a factor

# Customer Experience Data Model



# Smart City Provider Data Model

## **Reading**

- Type
- Value
- Date
- Time
- Location

# Smart City Provider Data Model

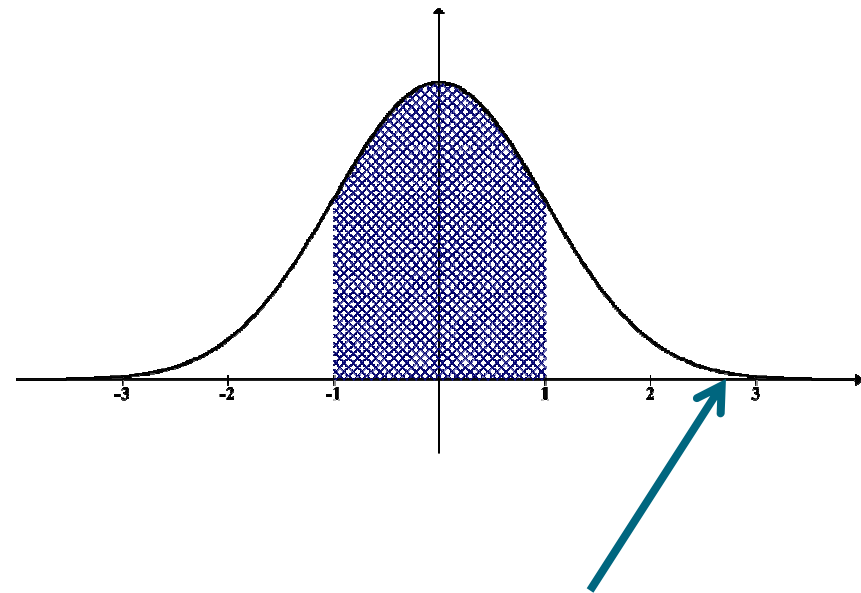
## **Reading**

- Type
- Value
- Date
- Time
- Location

- Models are very different
- They use different terms
- They assume different information structures
- Each is “right” for its application

# The “Long Tail”

- Some applications are heavily used
  - They justify major development effort
- Others are hardly used at all
  - They are only viable if they need minimal effort
- Software productivity and rapid development enable a “long tail” of lightly-used applications
- The customer experience application is near the end of the tail



# Agenda

- The need for a data classification system
- Data classification
- **What do we have already?**
  - **Vocabularies**
  - Grammar
- Conclusions

# Dublin Core® Metadata Initiative



- Properties for use in resource description
  - Title: I wandered Lonely as a Cloud
  - Creator: William Wordsworth
  - . . .
- Standardized as ISO Standard 15836:2009
- <http://dublincore.org/>

# ISA Core Vocabularies

familyName	givenName	gender	dateOfBirth
Delarue	Mathieu	M	01-Jan-00
Doe	Jane	F	20-Feb-97
Mustermann	Erika	F	25-Sep-03
Nordmann	Kari	F	18-Dec-11
Rossi	Mario	M	07-Jul-88
Smith	John	M	03-Mar-73

- Re-usable and extensible standard data models
  - Person
  - Registered Organization
  - Location
  - Public Service
- [http://ec.europa.eu/isa/read-y-to-use-solutions/core-vocabularies\\_en.htm](http://ec.europa.eu/isa/read-y-to-use-solutions/core-vocabularies_en.htm)



# INSPIRE



- Infrastructure for Spatial Information in the European Community
- <http://inspire.ec.europa.eu/>

# UNSPSC®



- United Nations Standard Products and Services Code®
- Extensive and widely used classification of products and services
- Owned by the United Nations Development Programme, managed by GS1 US
- <http://www.unspsc.org/>

# SNOMED-CT



- Systematized Nomenclature of Medicine-Clinical Terms
- Extensive and widely-used clinical health terminology
- Owned by the International Health Terminology Standards Development Organisation (IHTSDO)
- <http://www.ihtsdo.org/snomed-ct>

# Public Service Vocabularies

- Data Catalog Vocabulary (DCAT) and the DCAT Application Profile
  - [https://joinup.ec.europa.eu/asset/dcat\\_application\\_profile/description](https://joinup.ec.europa.eu/asset/dcat_application_profile/description)
- Core Public Service Vocabulary: The Italian Application Profile
  - [http://www.w3.org/2013/share-psi/wiki/images/7/73/AgID\\_BerlinWorkshop.pdf](http://www.w3.org/2013/share-psi/wiki/images/7/73/AgID_BerlinWorkshop.pdf)
- Controlled Vocabularies and Metadata Sets for Public Sector Information Management
  - [http://www.w3.org/2013/share-psi/wiki/images/b/bc/SHAREPSI\\_Charalabidis\\_Metadata\\_v1.pdf](http://www.w3.org/2013/share-psi/wiki/images/b/bc/SHAREPSI_Charalabidis_Metadata_v1.pdf)

# Agenda

- The need for a data classification system
- Data classification
- What do we have already?
  - Vocabularies
  - **Grammar**
- Conclusions

# Basic Grammar

- Needed for consistent interpretation by people
- And especially for interpretation by machines

# Natural Language

- Good for expressing thoughts and emotions
  - “I wandered, lonely as a cloud . . .”
- Not specialized for data description
  - Lacks appropriate grammatical constructs
- Not used consistently

# Relational Database and Data Modeling

- Well established and widely used
- Specifically designed for data description
- Has international standards
  - ISO 11179 in particular
- Needs to be applied in non-RDBMS environments
  - This is possible using general concepts of *object class* and *property*



# Resource Description Framework

- Established W3C standard
- Can be supplemented by RDF Schema and the Web Ontology Language
- Defined machine-interpretable representations
- Body of sophisticated processing software
- Applying it to data description needs some thought

# Agenda

- The need for a data classification system
- Data classification
- What do we have already?
  - Vocabularies
  - Grammar
- **Conclusions**

# Conclusions



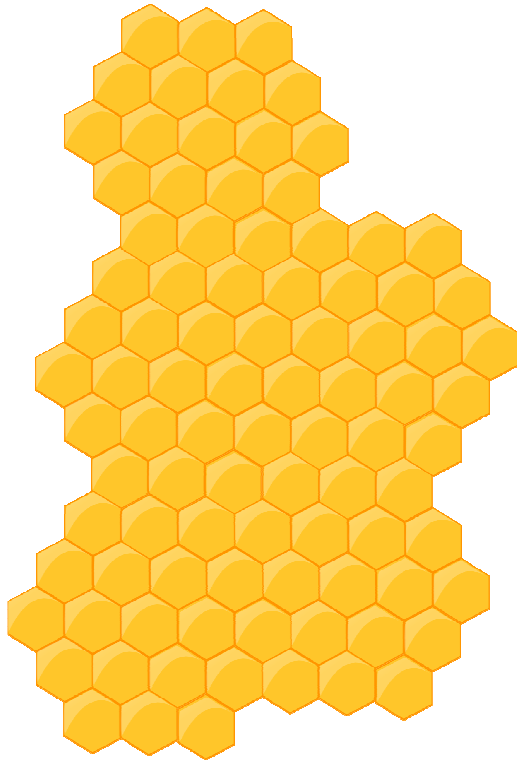
- Data is the new oil
- Full exploitation of its potential requires a data classification system

# The Data Classification System Should

STABLE  
half life more than one trillion years  
half life in range of billion years  
half life in range of million years  
half life in range of thousands of years  
half life in range of years  
half life in range of days  
half life in range of hours  
half life in range of minutes  
half life in range of seconds  
half life in range of milliseconds  
half life undetermined

- Enable use of the existing vocabularies that have been developed by industry bodies and standards organizations
- Provide a basic grammar for data descriptions
- Be consistent with relational database usage
- Be able to accommodate other data representation approaches
- Use RDF to facilitate semantic processing

# The Open Data Element Framework



- Index and method for using it to classify data elements
- Will meet the requirements
- Based on the UDEF ([www.opengroup.org/udf](http://www.opengroup.org/udf)) but
  - Has roles as well as object classes and properties
  - Has plugins
- Currently in technical review

# Core Vocabularies and Grammar for Public Sector Information Interoperability

**Thank you!**