Health and Safety on the Semantic Web

Automated Completion of COSHH Risk Assessment Forms

Mark Borkum
m.i.borkum@soton.ac.uk
RSC CICAG Meeting
21st October 2014, Burlington House
Outline

- Background
- Problem Statement
- Method
- Results
- Discussion
- Conclusion
Outline

- Background
- Problem Statement
- Method
- Results
- Discussion
- Conclusion
GHS

- = Globally Harmonized System of Classification and Labelling of Chemicals
- International standard, created by the UN
- Aims to supersede the various standards currently used in different countries, providing a new *lingua franca*
GHS (cont.)

- Classification elements:
  - Hazard classes
  - Hazard categories
  - Notes

- Labelling elements:
  - Hazard pictograms
  - Signal words
  - Hazard and precautionary statements

http://www.orosha.org/resource-newsletter/2013/06-2013/image/chemical-x-medium.jpg
CLP Regulation (EC) No 1272/2008

- Classification, Labelling and Packaging

- Aligns older EU systems with GHS, e.g., Dangerous Substances Directive 67/548/EEC

- Came into force in January 2009

- Mandatory from June 2015
COSHH

- Control and Substances Hazardous to Health Regulations 2002 (UK statutory instrument)

- Requires employers to protect their employees from the effects of exposure to hazardous chemical substances
**COSH/ASSESSMENT FORM**

<table>
<thead>
<tr>
<th>SUBSTANCE NAME</th>
<th>PHYSICAL FORM</th>
<th>QUANTITY</th>
<th>NATURE OF HAZARD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>liquid</td>
<td>1000 ml</td>
<td>None</td>
</tr>
<tr>
<td>Dextrose</td>
<td>solid</td>
<td>&lt; 20 g</td>
<td>Possible irritation to eyes and skin</td>
</tr>
<tr>
<td>Caffeine</td>
<td>solid (tea)</td>
<td>&lt; 1 g</td>
<td>Harmful if swallowed, induce vomiting.</td>
</tr>
<tr>
<td>Milk</td>
<td>liquid</td>
<td>&lt; 100 ml</td>
<td>No particular hazards</td>
</tr>
</tbody>
</table>

**NATURE OF PROCESS**

Liquid extraction of caffeine, followed by combination with dextrose to produce a sweet drink.

Is there a less hazardous substance? _No_

If so, why not use it?

**CONTROL MEASURES REQUIRED**

( Local exhaust ventilation, personal protection, etc. )

No specific measure required

---

c/o SmartTea
COSHH Risk Assessment

0. Describe scientific experiment

1. Enumerate chemical substances

2. Discover health and safety information

3. Interpolate template

---

**COSHH ASSESSMENT FORM**

<table>
<thead>
<tr>
<th>SUBSTANCE NAME</th>
<th>PHYSICAL FORM</th>
<th>QUANTITY</th>
<th>NATURE OF HAZARD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>liquid</td>
<td>1000ml</td>
<td>None</td>
</tr>
<tr>
<td>Pentrose</td>
<td>solid</td>
<td>≤20g</td>
<td>Possible irritation to eyes and skin</td>
</tr>
<tr>
<td>Caffeine</td>
<td>solid (tea)</td>
<td>≤1g</td>
<td>Toxic if swallowed, induce vomiting</td>
</tr>
<tr>
<td>Milk</td>
<td>liquid</td>
<td>≤100ml</td>
<td>No particular hazards</td>
</tr>
</tbody>
</table>

**NATURE OF PROCESS**

Liquid extraction of caffeine, followed by combination with pentrose to produce a sweet drink.

**CONTROL MEASURES REQUIRED**

No specific measure required.

---

C/O SmartTea
Outline

- Background
- Problem Statement
- Method
- Results
- Discussion
- Conclusion
Problem Statement

- For humans, performing COSHH risk assessment has two key drawbacks:
  1. Reduced working time
  2. Inherent danger
COSHلاح Risk Assessment Hazards

0. Incorrect description of scientific experiment

1. Incorrect enumeration of chemical substances

2. Incorrect health and safety information

3. Transcription errors
Problem Statement (cont.)

- An incorrect COSHH risk assessment is a potential source of danger, i.e., a hazard.

- Using automation, the associated risk can be managed, e.g.,
  - Input
    - Description of scientific experiment
    - Template
  - Output
    - Interpolated template
Outline

- Background
- Problem Statement
- Method
- Results
- Discussion
- Conclusion
Method

- Identify/generate required information resources:
  - Formalisation of CLP Regulation
  - Database of classified* chemical substances
  - Templates

- Implement Web application

* As specified by CLP Regulation
RDF

- Resource Description Framework
- Family of W3C specifications for modelling information as Web resources
- Used in knowledge management systems
- Core data model is based upon making “subject—predicate—object” statements
RDF (cont.)

![RDF Diagram](http://www.w3.org/TR/rdf11-primer/example-graph.jpg)
SKOS

- = Simple Knowledge Organisation System
- W3C recommendation for representation of controlled vocabularies
- Built upon RDF and RDF Schema
Method (cont.)

- Use RDF, RDF Schema and SKOS
- Transfer definitions verbatim*
- Automate* extraction and enrichment of classifications

* Facilitates curation of provenance information
Outline

- Background
- Problem Statement
- Method
- Results
- Discussion
- Conclusion
Results

- Machine-accessible representation of GHS/CLP Regulation
  - Definitions (annexes I—V)
  - Classified chemical substances (annex VI)

RDF Schema for GHS/CLP Regulation
RDF for EC (No) 007-001-01-2
Results (cont.)

- The dataset is interesting, but not practical
- It only contains ~3000 classifications
Observation

- Many researchers procure their chemical substances from specialist suppliers
Benzene
analytical standard

CAS Number 71-43-2 | Empirical Formula (Hill Notation) C₆H₆ | Molecular Weight 78.11
Bellstein Registry Number 969212 | EC Number 200-753-7 | MDL number MFCD00000309
PubChem Substance ID 24847664

POPULAR DOCUMENTS: FTNMR (PDF)

Documents
Bulk Quote-Order Product MSDS
Certificate of Analysis
Enter Lot No.

Safety Information
Symbol

Signal word
Hazard statements
Precautionary statements
Personal Protective Equipment
Hazard Codes (Europe)
Risk Statements (Europe)

Danger
P201-P210-P301 + P310-P305 + P351 + P338-P308 + P313-P331
Eyeshields, Faceshields, full-face respirator (US), Gloves, multi-purpose combination respirator cartridge (US), type ABEK (EN14387) respirator filter
F, T
45-46-11-36/38-48/23/24/25-65
Excerpt of RDF for <fluka/12540>

@base <http://www.sigmaaldrich.com/catalog/product/> .
@prefix chemaxiomprop: <http://www.polymerinformatics.com/ChemAxiom/ChemAxiomProp.owl#> .
@prefix chemdomain: <http://www.polymerinformatics.com/ChemAxiom/ChemDomain.owl#> .
@prefix ghs: <http://xmlns.com/ghs/0.1/> .
@prefix sial: <http://www.sigmaaldrich.com/ns#> .

<fluka/12540> a chemdomain:NamedChemicalSpecies;
  chemdomain:hasIdentifier [ a sial:ChemicalNumber;
    chemdomain:hasValue "042802"],
  [ a chemdomain:Name;
    chemdomain:hasValue "Benzene"],
  [ a chemdomain:CASNumber;
    chemdomain:hasValue "71-43-2"],
  [ a chemdomain:SMILES;
    chemdomain:hasValue "c1ccccc1"],
  [ a chemdomain:MolecularFormula;
    chemdomain:hasValue "C6H6"];
ghs:hasHazardCategory <http://id.unece.org/ghs/hazard_categories/Flam_Liq_2>,
  <http://id.unece.org/ghs/hazard_categories/SPEC_RE_1>,
  <http://id.unece.org/ghs/hazard_categories/Skin_Irrit_2>,
  <http://id.unece.org/ghs/hazard_categories/Carc_1A>,
  <http://id.unece.org/ghs/hazard_categories/Asp_Tox_1>,
  <http://id.unece.org/ghs/hazard_categories/Eye_Irrit_2>,
  <http://id.unece.org/ghs/hazard_categories/Muta_1B> .
Red square = Chemical substance
Orange square = Mixture part
Green square = PPE product, e.g., gloves
Blue square = GHS hazard category
Web Application Protocol

1. Load Web application
2. Describe scientific experiment
3. Select template
4. Accept terms and conditions
5. Receive interpolated template
1) Load Web Application

“Generate” button is disabled
2) Describe Scientific Experiment

Dynamic search with auto-completion
2) Describe Scientific Experiment

Selected product is added to list

Link to Sigma-Aldrich product page
3) Select Template

List of Templates
4) Accept Terms and Conditions

“Generate” button is enabled
The University requires that assessment forms are printed and signed.

<table>
<thead>
<tr>
<th>Substance Name</th>
<th>Physical Form</th>
<th>Quantity</th>
<th>Nature of Hazard</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Benzene</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Nature of Process**

<table>
<thead>
<tr>
<th>Control Measures Required</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Benzene</strong></td>
</tr>
<tr>
<td>- PS21: Obtain special instructions before use.</td>
</tr>
<tr>
<td>- PS22: Do not handle until all safety precautions have been read and understood.</td>
</tr>
<tr>
<td>- PS31: Keep away from heat, sparks, open flames, and sewage. No smoking.</td>
</tr>
<tr>
<td>- PS32: Keep area well-ventilated.</td>
</tr>
<tr>
<td>- PS33: Use only non-flammable solvents.</td>
</tr>
<tr>
<td>- PS41: Use non-flammable cleaning equipment.</td>
</tr>
<tr>
<td>- PS42: Store in suitable well-ventilated area.</td>
</tr>
</tbody>
</table>

**Declaration**

<table>
<thead>
<tr>
<th>Name of Assessor</th>
<th>Name of Supervisor (or students only)</th>
<th>Head of Department</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

One row per product

Editable text area for “Nature of Process”

Editable text fields for “Physical Form” and “Quantity”
Toggle visibility of classification and labelling elements

Integration with purchasing platform
Web Application Limitations

- Descriptions of scientific experiments are just sets of “Chemical substance—Phase—Quantity” triples

- To enable more detailed analysis, we need richer descriptions
This plan describes **3 activities**, **2 entities**, and **3 usages**.

- **Initialize** [edit] [destroy]
  - New Usage
  - Generates: Acetaldehyde puriss. p.a., anhydrous, >=99.5% (GC) [destroy]
  - Generates: Benzene analytical standard [destroy]
  - New Entity

- **Mix** [edit] [destroy]
  - Uses: Initialize → Acetaldehyde puriss. p.a., anhydrous, >=99.5% (GC) (Role: Primary Target) [edit] [destroy]
  - Uses: Initialize → Benzene analytical standard (Role: Secondary Target) [edit] [destroy]
  - New Usage
  - New Entity

- **Non-Linear Activity** [edit] [destroy]
  - Uses: Initialize → Acetaldehyde puriss. p.a., anhydrous, >=99.5% (GC) [edit] [destroy]
  - New Usage
  - New Entity

- **New Activity**

---

**Figure**: Workflow-style provenance graph for *Eric’s Plan*.
Sigma-Aldrich One Page Safety Data Sheets (“One Pagers”)
**Sigma-Aldrich “Quick Order Form”**

<table>
<thead>
<tr>
<th>Select</th>
<th>Product Number</th>
<th>Description</th>
<th>Quantity</th>
<th>Your Price</th>
<th>Item Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>✅</td>
<td>FLUKA 00070</td>
<td>Acetaldehyde puriss. p.a., anhydrous, &gt;=99.5% (GC)</td>
<td>1</td>
<td>Not Calculated</td>
<td>Not Calculated</td>
</tr>
<tr>
<td>✅</td>
<td>ALDRICH Z677647</td>
<td>★ Butoject gloves</td>
<td>0</td>
<td>Not Calculated</td>
<td>Not Calculated</td>
</tr>
<tr>
<td>✅</td>
<td>FLUKA 12540</td>
<td>Benzene analytical standard</td>
<td>1</td>
<td>Not Calculated</td>
<td>Not Calculated</td>
</tr>
<tr>
<td>✅</td>
<td>ALDRICH Z677698</td>
<td>★ Vitoject gloves</td>
<td>0</td>
<td>Not Calculated</td>
<td>Not Calculated</td>
</tr>
</tbody>
</table>

Add Selected Items to Shopping Cart
Outline

- Background
- Problem Statement
- Method
- Results
- Discussion
- Conclusion
Data Stakeholders
Data Stakeholders (cont.)

- Bidirectional data exchange

- Forwards
  - Suppliers and safety officers consume the legislation
  - Researchers consume chemical information and templates

- Backwards
  - Legislators, suppliers and safety officers gather usage statistics
The Wrong Question

- Who has legal liability when users act upon the correct answer to an incorrect question?
  - USER Can I drink this poison?
  - SYSTEM Yes
  - USER Is it safe to drink this poison?
  - SYSTEM No
Issues with GHS/CLP Regulation

- Does not identify hazards associated with presence of strong magnetic fields
- Identifies hazards of reagents and products of chemical reactions, but not kinetics
- Labelling elements do not have requirement levels
- *et al*
GHS Statements

- Statements are fragments of text that describe the “nature of” and “recommended measure(s) to minimise or prevent adverse effects resulting from exposure to” hazardous chemical substances.

- Statements are either atomic or compound.
“P301: IF SWALLOWED:”

+ 

“P330: Rinse mouth”

+ 

“P331: Do NOT induce vomiting”

= 

“P301+330+331: IF SWALLOWED: Rinse mouth. Do NOT induce vomiting”
The notation suggests that there is a binary statement composition operator.

However, this is not the case.

There is a statement composition operator *per se*, but it is variadic and partial.

Hence, new compound statements cannot be derived.
Algebraic GHS Statements

- Translate phrases into controlled natural language, and then into expressions of terms and logical operators
- Terms constitute a controlled vocabulary
Controlled P301+330+331

“P301: If you swallow the substance then …”

+ “P330: You rinse mouth”

+ “P331: Not you induce vomiting”

= “P301+330+331: If you swallow the substance then you rinse mouth and not you induce vomiting”
Algebraic P301+330+331

P301 = \( \lambda x. \text{“swallow the substance”} \Rightarrow x \)

P330 = “rinse mouth”

P331 = \( \neg \text{“induce vomiting”} \)

P301+330+331 = P301 (P330 \( \land \) P331)
Outline

- Background
- Problem Statement
- Method
- Results
- Discussion
- Conclusion
Conclusion

- Given formalisation of legislation, automated completion of COSHH risk assessment forms is trivial
- Legal mitigations, such as disclaimers and curation of provenance information, are vital
- Linchpin is availability of accurate and trustworthy chemical information
- Despite the success of GHS/CLP Regulation, there still exist many opportunities for enhancement
Thank You For Listening

- Questions?
Acknowledgements

- University of Southampton
  - Jeremy Frey, David Kinnison

- Sigma-Aldrich
  - Bo Jin, Jane Murray, Tom Pieper, Judith Pruss
IUPAC Green Book

Quantities, Units and Symbols in Physical Chemistry

Third Edition