The General Ontology Evaluation Framework (GOEF):

A Proposed Infrastructure for the Ontology Development Lifecycle

Rensselaer Polytechnic Institute
3 / 14 / 2013

The Problem: Ontology Elephants

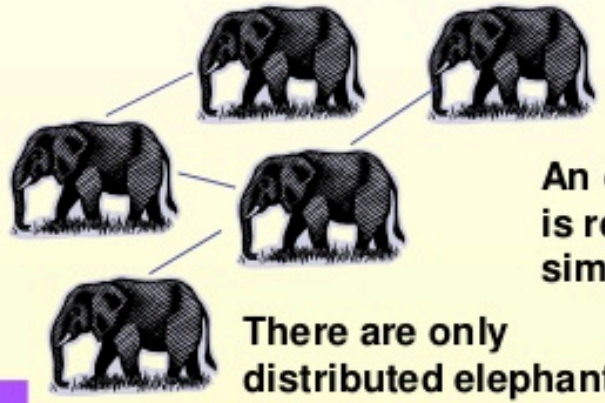
There is no single real elephant

There must be a purpose for an elephant: use cases?

An elephant is abstract

An elephant is very abstract

There must be an upper elephant



An elephant is really very simple

There are only distributed elephants & their mappings

An elephant is the result of consensus

Open vs. Closed Elephant

Prospects and Possibilities for Ontology Evaluation: The View from NCOR, Obrst, Hughes, Ray, WWW 2006, May 22–26, 2006, Edinburgh, UK.

MITRE

Semantic Technology Lifecycle

Users need a standard way to build sound ontologies and reuse them for a different purpose

Semantic Web Methodology and Technology Development Process

- Establish and improve a well-defined methodology vision for Semantic Technology based application development
- Leverage controlled vocabularies, etc.

Current Challenges

- Use case driven ontology evaluation is managed through direct inspection by subject matter experts.
- However, this is a time-consuming effort, which requires individual review of potentially multiple ontologies.
- What if we could develop a system which could take in a use case formalism, and give recommendations for ontologies to use?

General Ontology Evaluation Framework (GOEF)

Evolve towards science and engineering discipline for ontology

- Goal: Enable objective evaluation of an **ontology** with respect to a **use case**.
- Both are constructed / deconstructed to extract / expose the evaluation criteria and the ontology-encoded knowledge.
- Facilitates ontology design, modular construction, development management, and evaluation is built into the development process.

Create procedures, processes, methods to help define, adjudicate, and ensure quality of knowledge capture/ representation

Motivating Example: iChoose

CHOOSING RESPONSIBLY!

WHY USE "RiCh"?

Because "RiCh" is

"Powered by I-Choose"

I-Choose architecture guarantees that the information provided to you are TRUSTWORTHY!

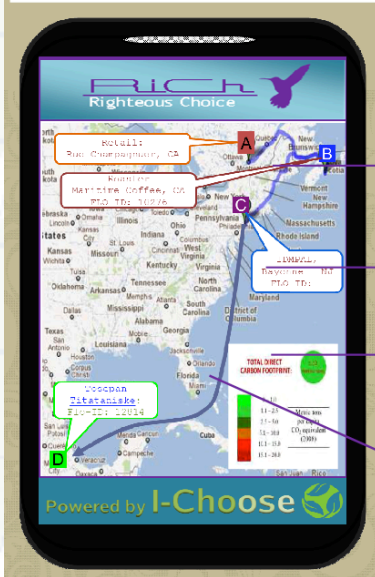
RiCh
Righteous Choice

WHAT IS "RiCh"?

"RiCh" is mobile application offering comparability and traceability functions for sustainable products.

"RiCh" provide easy to understand and trustworthy information.

"RiCh" will help you making socially and environmentally purchase that you know you can trust.



How much do you know about YOUR PURCHASING IMPACT?

With "RiCh" you can...

Trace the origin of your product. "RiCh" will enable you to trace the organization and/or persons responsible in making the products available in front of you.

"RiCh" will provide and compare the social, environmental, and health rating of the product.

Trace the carbon footprint of your product. "RiCh" will calculate the carbon footprint of the product you are interested.

"RiCh" will provide not only the expert based rating of your product but also your friend recommendation.

1123 STREET AVENUE • ANYTOWN, ST 54321 • 888·765·4321 • www.ri-ch.com

Got More Questions?

"RiCh" CAN ANSWER YOUR OTHER QUESTIONS TOO!

Customer

- Who roasted the coffee packaged in this bag?
- What is the country of origin of the coffee beans in this bag?
- How much money was paid to the workers who picked the coffee from the plants?
- Who certified this coffee as organic or fair trade?
- Is this coffee labeled as Fair trade because the organization who sold it is FTF certified?
- What is the difference between the Rainforest Alliance seal and the Bird Friendly Seal from the Smithsonian?
- What are the principles that this certification implies?
- Are the principles from this certification verified by an independent third party?

Retailer /Roaster

- What is the contact information for organizations that have FLO/FTF/USDA Certification?
- For how long this company has held this certificate?
- What criteria should I meet to carry fair trade/organic products?
- What are the most purchased seals and certifications among consumers in this city?
- Who are the most valuable costumers in my area?
- Which certificates involve an independent inspection?
- What principles implied by this certificate are more effective in influencing purchasing behaviors when they are displayed together with the certificate?

IN NEAR FUTURE!!!

Certifiers

- What is the market share of my seal?
- What percentage of consumers in this country is interested in buying products that comply with environmental/social issues?
- What principles implied by my certificate are more effective in influencing purchasing behaviors when they are displayed together with the certificate?
- What are the common principles between my seal and others in the market?
- What are the principles that are unique to my certificate?

Producer

- What are the certifications with the biggest market share?
- What other producers in my region or country have the same certificate that I have?
- What is the contact information of roasters looking for coffee with the certifications I have?
- What criteria should I meet to produce fair trade/organic products?
- What are the most purchased seals and certifications among consumers in this city/country?
- Which certificates involve an independent inspection?

1123 STREET AVENUE • ANYTOWN, ST 54321 • 888·765·4321 • www.ri-ch.com

Consumers want to know….

1. Who roasted the coffee packaged in this bag?
2. What is the country of origin of the coffee beans in this bag?
3. How much money was paid to the workers who picked the coffee from the plants?
4. Who certified this coffee as organic or fair trade?
5. Is this coffee labeled as Fair trade because the organization who sold it is FTF certified?
6. What is the difference between the Rainforest Alliance seal and the Bird Friendly Seal from the Smithsonian?
7. What are the principles that this certification implies?
8. Are the principles from this certification verified by an independent third party?

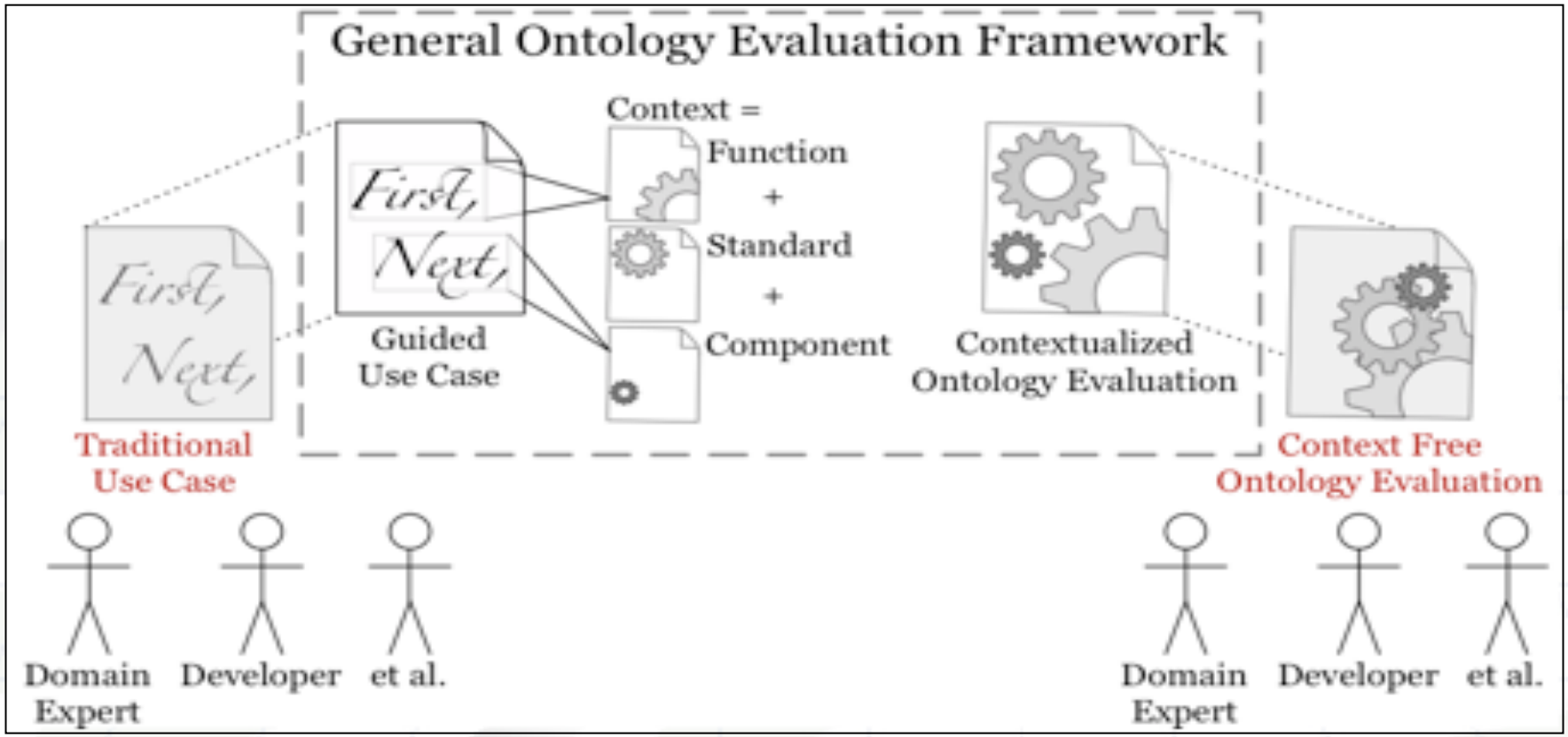
GOEF Approach

Two stages:

- Recast use case into its components:
  - Functional objective
  - Design objective and requirements specification
  - Semantic components required to achieve above
- Evaluate components using objective metrics
  - Place existing evaluation methods in context by utility

Can be used for incremental design, development and testing

Current Infrastructure



Function Level

- Represents the top level of the use case.
  - i.e. the function of the intended use (for search, for integration, for gene annotation)
- Additionally, the primary characteristics that define the classification of the domain of the ontology (organism, aircraft, instrumentation, etc.).

Standard Level

- Represents the quality or standard that has to be met by the application (e.g. for legal, interoperability, function, compliance, etc.)
- Further specifies the domain characteristics.

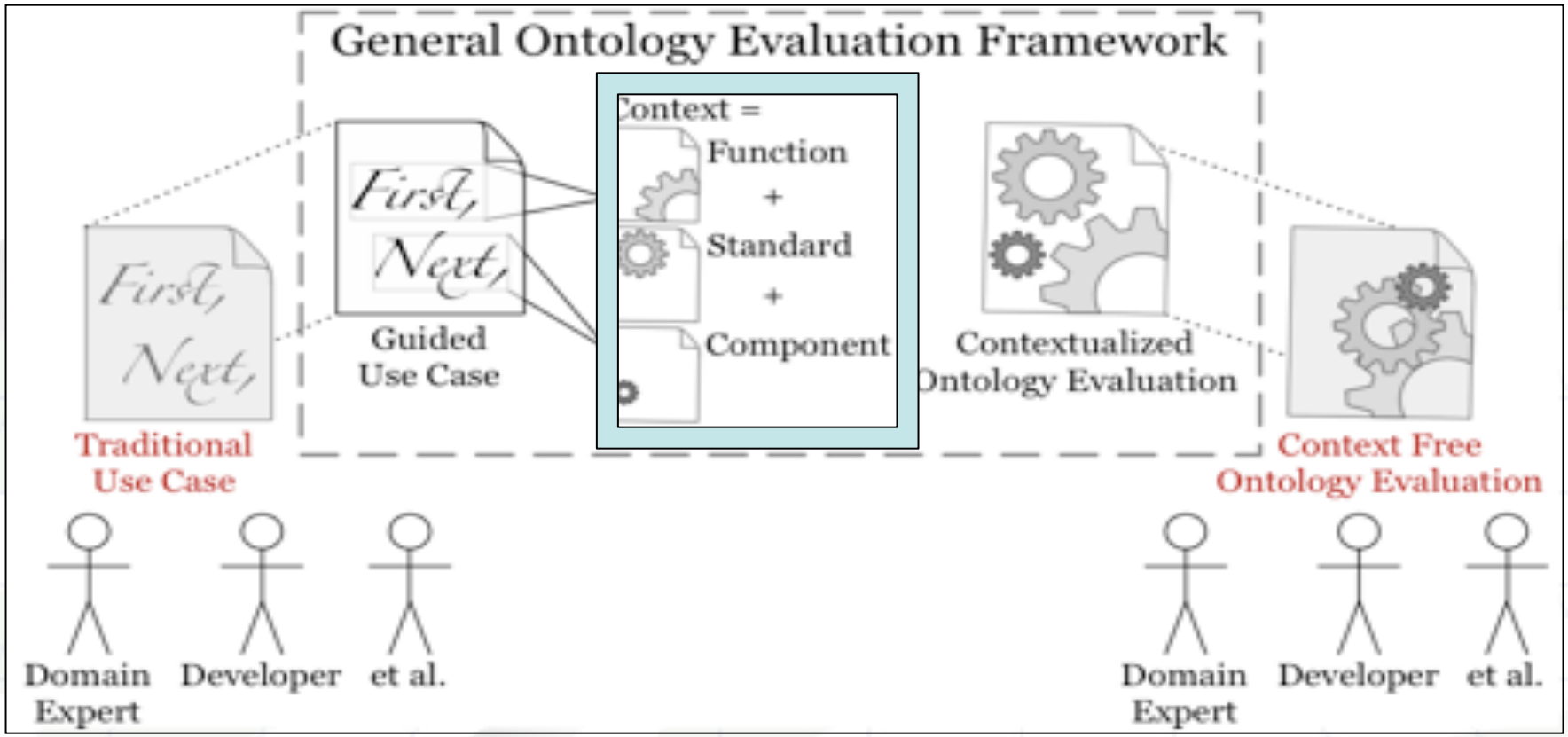
Component Level

Identifies ontology fragments that are needed in order to achieve compliance with the standard and fulfill the function.

Example: Flo-Cert FLO Standard 3.1.1

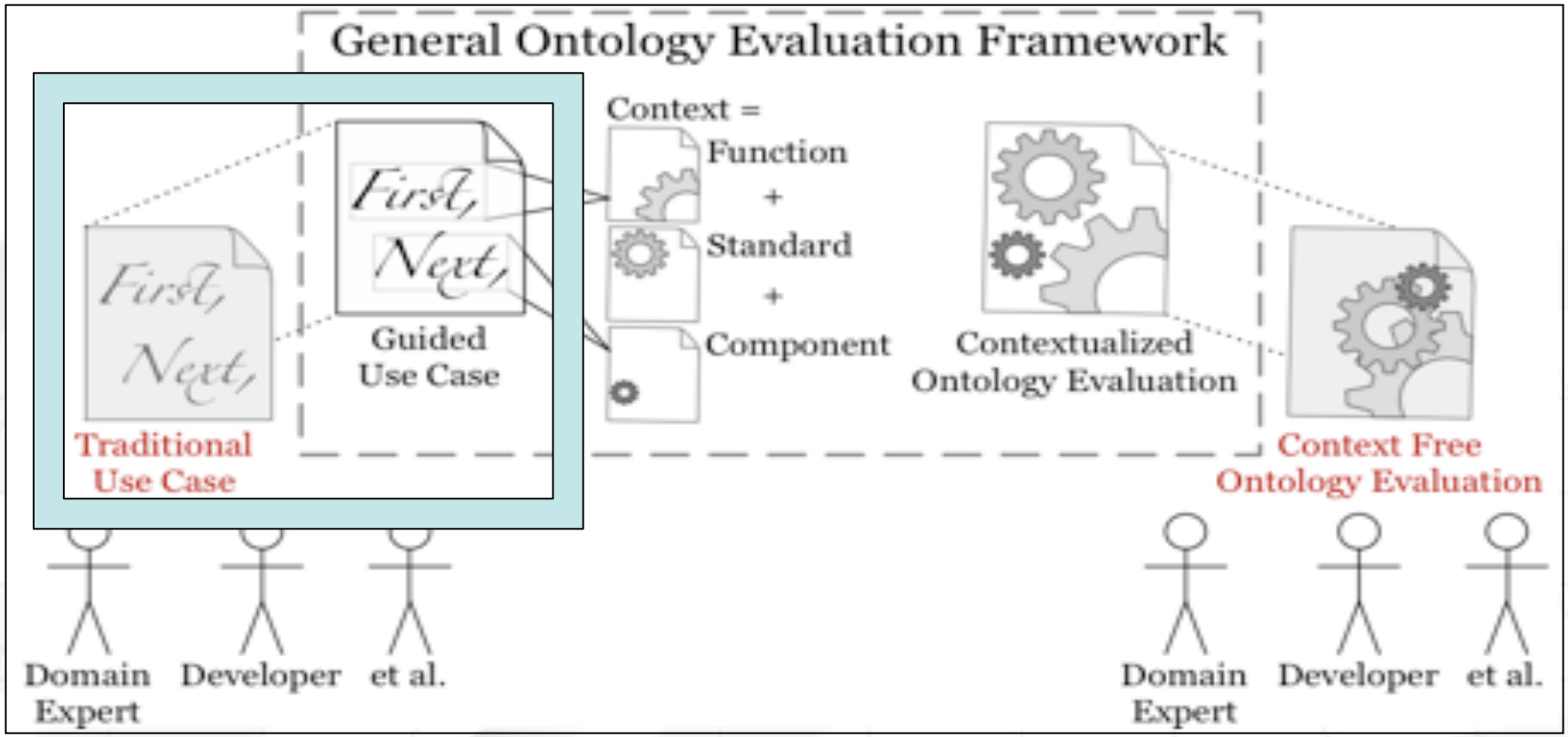
FLO Standard [FLO e.V.] or other	Applicable for:	CC No.	FLO-CERT Compliance Criteria	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Time	Criteria Type
3.1.1	1st grade, 2nd grade, 3rd grade Not applicable for: Shrimps, Timber,	3.1.0.01	(1 July 2012) You have informed your members about the Fairtrade standards for environmental and labour practices.	The organisation did not provide any information to members.	Some members were informed but less than 80% OR information provided did not cover all Core requirements.	At least 80% of members were informed once about at least Core requirements AND content of information was sufficient AND trainings are planned/scheduled for the remaining producers.	At least 80% of members were informed annually about at least Core requirements AND training materials are clear and correct AND there are informative materials.	At least 80% of members are informed annually about all requirements AND training materials are clear and correct AND there are informative materials.	0	C

Three Levels of Evaluation



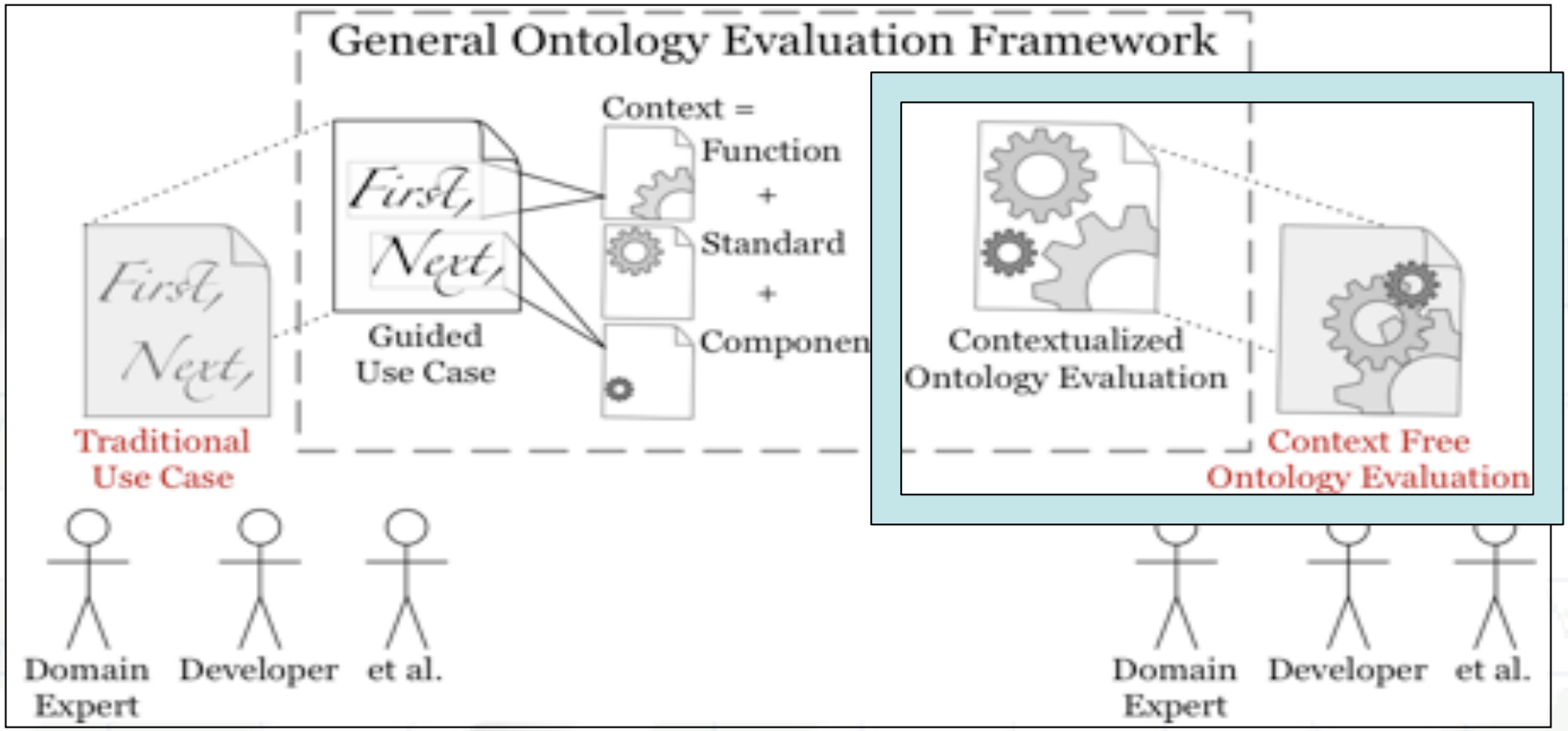
- These combine to form the context for evaluation.

Formalizing Use Cases



- Methodology for formalizing use cases still needed.
- Development – based around 3 level evaluation – will be the focus of a proposed Ontology Summit hackathon.

Evaluation Metrics



Development of metrics (to be developed or used) will follow from formalization of use case design.

Motivating Example: iChoose

Function:

Enable retrieval of specific criteria evaluations that occurred during an evaluation process of a particular product.

Design objective:

Initial system: Satisfy consensus user criteria pre-determined by survey research

Semantic components:

Compliance Criteria
- a) Pesticide
- b) Minimum Wage
- c) Child labor

Certification Body
- a) Flo-Cert
- b) Certified private inspectors
- c) Sustainable Farm Certification Intl, Ltd.

Standard
- a) FairTrade International
- b) USDA Organic
- c) Sustainable Agriculture Alliance

Product
- a) Coffee
- b) Sugar Cane
- c) Fruit

Motivating Example: i-Choose sustainable consumer choice

Correctness:

- General logical validation
- Are the right terms used (compliance criteria vs. guidelines vs. standards)
- Match information provided in the ontology to information consensus user wants (surveyed).

Completeness:

- Calculate % coverage of minimum terms
  - All “severe” pesticides listed (certain %)
  - All pesticides prohibited by U.S. EPA. Listed

Utility:

- Validate against known test sets
- Consumer Consensus Questions Satisfied

References

Jiao Tao, Li Ding, Deborah L. McGuinness. Instance Data Evaluation on the Semantic Web

Joanne Luciano – Presentation on ontology evaluation methodshttp://www.slideshare.net/joanneluciano/luciano-pr-08849ontologyevaluationmethodsmetrics-8294436

FairTrade® Standards. http://www.fairtrade.net/fileadmin/user_upload/content/2009/standards/documents/2012-07-11_SPO_EN.pdf

FloCert Compliance Criteria – Small Producer Organizations. http://www.flo-cert.net/flo-cert/fileadmin/user_upload/certification/requirements/en/Current_CC/PC_PublicComplianceCriteriaSPO_ED_7.1_en.pdf

Backup Slides

Flo-Cert Components

FLO Standard [FLO e.V.] or other	Applicable for:	CC No.	FLO-CERT Compliance Criteria	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Time	Criteria Type
Coffee 2.3.4	1st grade, 2nd grade, 3rd grade Coffee	2.3.0.04	(Coffee) The producer notified the buyer of his intention to default a contract with a minimum of 2 months prior to the shipment date.	No		Yes			0	M
Coffee 2.3.2	1st grade, 2nd grade, 3rd grade Coffee	2.3.0.05	(Coffee) When a broker is included in a commercial transaction by the producer the contract must: a) identify which party required its services and consequently will pay for them and the commission is not deducted from the FOB price. b) include agreement of the other party to the broker's participation.	No contract available.	Basic contract available, but none of the details required, no agreement of the other party.	Basic contract and agreement available which roughly mention the required details.	Good contract and agreement in place which mentions the required details in a clear manner.		0	C
Coffee 2.3.1	1st grade, 2nd grade, 3rd grade Coffee	2.3.0.06	(Coffee) In the case of price-to-be-fixed contracts, whenever the seller asks to fix the price before the coffee harvest starts, there is written agreement by the payer on doing this and a written risk management strategy agreed by both parties.	No	Only one of the two required documents available: written agreement or risk managemen t strategy.	Both documents available, agreed by both parties.			0	C
		2.3.0.07	(Coffee) The producer does not sign outright-priced-contracts with a duration longer than 1 crop period.	No		Yes			0	C
Fresh fruit 1.4.1	1st grade, 2nd grade, 3rd grade Banana	2.1.0.19	(Banana) The seller notifies the Fairtrade payer (or ripener) in writing about the arrangement of counter inspection by an authorised surveyor within 48 hours after receiving the quality claim. If the seller does not react to the quality claim report within the specified period, the buyer (or ripener) may assume that the seller accepts the refusal of the fruit.	No notifcation.	Notification made, but not in writing.	Notification made in writing within 48 hours.	Notifcation made in writing immediately .	= RANK 4 AND all communicat ion clearly documented and filed.	0	C

Flo-Cert Component

FLO Standard [FLO e.V.] or other	Applicable for:	CC No.	FLO-CERT Compliance Criteria	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Time	Criteria Type
2.1.8	1st grade, 2nd grade, 3rd grade Cane sugar, Cocoa, Fruit juices, Tea	2.1.0.14	(Cane sugar, Cocoa, Fruit juices, Tea) If you produce and process cocoa, cane sugar, juice or tea, and you sell to operators without physical traceability, you do not need to fulfill physical traceability but the volumes sold as Fairtrade do not exceed the equivalent volumes produced by your members.	There are estimated excessive sales by more than 10% OR there is no system that allows calculations.	There are estimated excessive sales by 1-10%.	No excessive sales except possible mistakes estimated up to 1% of sales.	No excessive sales with no mistakes.	No mix up of product and no mistakes AND compliant with sourcing record system with records indicating the name of the individual member, date of purchase, product name, volume and the price received by the member.	0	M

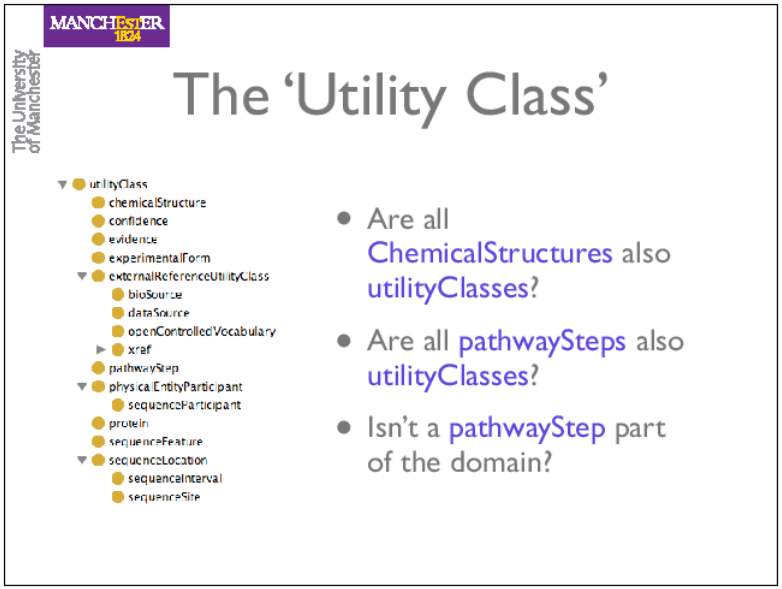
Examples:

- BioPAX (prior work)
- Habitat-Lite (subset of Environmental Ontology to support of NSF funded Mining Metadata for Metagenomics)
- Influenza Infectious Disease Ontology (for Genomics for Bioforensics MSR)

Example (1) BioPAX lack of fluency

chemical structure & pathway steps incorrectly modeled

- misunderstanding of the language (language has capability)
- modeled disjoint from the biology & chemistry
- leads to logical inconsistency



OWL has a steep learning curve, it's easy to get things wrong.

Example (2) Habitat-Lite: correctness & completeness

Objective: facilitate capture of habitat and environmental metadata on genomic sequences

Approach: select subset of terms with highest frequency and evaluate usefulness by correctness and completeness metrics

- Evaluated correctness
  - 64% agreement (84 of 132 terms) of automated and expert mapping of terms
- Evaluated coverage of terms
  - 84% exact matches ("host," "aquatic," and "soil" covered 75%)

Hirschman, Clark, Cohen, Mardis, Luciano, Kottmann, Cole, Markowitz, Kyprpides, Field
Habitat-Lite: a GSC Case Study Based on Free Text Terms for Environmental Metadata
OMICS A Journal of Integrative Biology Volume 12, Number 2, 2008 (in press)

Integrity Issues

- Unexpected Individual Type (UIT) Issue
  - rdfs:domain
  - rdfs:range
  - owl:allValuesFrom
- Redundant Individual Type (RIT) Issue
- Non-specific Individual Type (NSIT) Issue
- Missing Property Value (MPV) Issue
  - owl:cardinality
  - owl:minCardinality
- Excessive Property Value (EPV) Issue
  - owl:cardinality
  - owl:maxCardinality

Generic Evaluation Process (GEP)

- Load instance data D
  - Is loading failing?
- Parse instance data D
  - Is D **syntactically** correct?
- Load referenced ontologies O = {O_1,O_2, …}
  - Is O_i reachable? where O_i defines the terms used by D.
- Inspect logical inconsistencies in D
  - Is O_i logically consistent?
  - Merge all consistent referenced ontologies into O'
  - Are D+O' **logically** consistent?
- Inspect integrity issues in D
  - Compute DC = INF(D,O') which includes all triples in D and O', and all inferred sub-class/sub-property relations
  - Is there any **integrity** issue in D?

Jiao Tao, Li Ding, Deborah L. McGuinness
Instance Data Evaluation on the Semantic Web 2012

Example (3) Enable Influenza Research

(proposed construction and subsequent evaluation)

Function: Enable investigation of data collected on influenza strain mutations that cause death in birds

Design objective: Minimum Information about an Influenza Genotype and a Nomenclature Standard (MIIGNS)

Semantic components:

1. **biomaterial transformations**
   - **a. recombinant plasmid biomaterial transformation**
   - **b. site-directed mutagenesis biomaterial transformation**
   - **c. reverse genetic virus production biomaterial transformation**
   - **d. Mouse infection biomaterial transformation**
2. **assays**
   - **a. weight assay**
   - **b. virus replication / mouse lung assay**
   - **c. Cytokine quantification assay**
3. **data transformations**
   - **a. statistical difference evaluation**