



行政院環境保護署

Environmental Protection Administration
Executive Yuan, R.O.C. (Taiwan)

An Ontology-Driven Approach for Harmonizing and Integrating Environmental Information

Yu-Chi Chu Dept. of Environmental Monitoring and Information
Management, **EPA, TAIWAN**

Su-Mei Huang Dept. of Computer Science, Soochow University,
TAIWAN

Chin-Cheng Lien **TAIWAN**

Chen-Chau Yang College of Electrical and Computer Engineering,
St. John University, **TAIWAN**

EnvirolInfo 2007 Conference
Warsaw, Poland
September 12-14, 2007

Introduction



- Government agencies and private sectors need detailed information regarding the current state of the environment
- It is very difficult to share environmental data
 - resides on disparate databases
 - heterogeneity (**syntax** and semantics)
- There is a demand for appropriate systems to provide integrated environmental information

Heterogeneity issues



- **Structural heterogeneity**
means that different information systems store their data in different structures.
- **Semantic heterogeneity**
considers the **content of** an information item and its intended meaning.

Example -1



S₁ { Factory(facId, name, address, ...)
Permit(facID, permitNo, description, ...)

S₂ { Plant(serialNo, plant-name, Plant-
location, ...)

S₃ { <Site>
 <Name> ... </Name>
 <Location>
 <latitude> ... </latitude>
 <longitude> ... </longitude>
 <Location>
 <WasteItem> ... </WasteItem>
 .
 .
</Site>

- the offices in EPA do not all use the same database schema
- with the growing amount of data on the Internet, we are facing with data that is not well designed but with little structure such as HTML pages and XML documents

Example -2



EIONET
GEMET Thesaurus

SERVICES | REPORTNET | TOOLS | TOPICS

You are here: Eionet » GEMET thesaurus

- » User directory
- » Roles
- » NFP/Eionet IG
- » Mails to NFPs
- » Eionet ABC Wiki
- » 50 latest updates
- » Priority dataflows

Find a person

Thematic listings | Alpha

Select language:

Themes list

- administration
- agriculture
- air
- animal husbandry
- biology
- building

TRS Home
SoR Home

Introduction
Search
Contact Us

SoR Resources
Recent Additions
Upcoming Events

U.S. Environmental Protection Agency

Terminology Reference System

Recent Additions | Contact Us

EPA Home » SoR » TRS

Welcome to the Terminology Reference System

TRS Transition offering associated menus. The TRS will be available to the public, the

行政院環境保護署
環境資料庫

資料庫主題

- 空氣
- 噪音
- 水
- 土壤
- 廢棄物
- 毒化物
- 資源回收
- 環境用藥
- 環境影響評估
- 非屬原子能
- 游離輻射
- 環境統計
- 列管場所
- 居家環境系統
- 地方環境資料
- 地理資訊系統
- 回首頁

國際常用環境詞彙

查詢詞彙: 於 詞彙

(選擇以五種字母字首) ABCDEFGHIJKLMNOPQRSTUVWXYZ

殘留性有機污染物質

サンリュウセイユクキオセンプラシツ 【英】Persistent Organic Pollutants 【略】POPs

解説

難分解性で環境中に残留し、食物連鎖を通じて生物に蓄積され、人の健康や生態系に影響を及ぼす性質を有する汚染物質のこと。POPsの典型的なものは、アルドリル、ディルドリン、エンドリン、クロルデン、ヘプタクロル、DDT、マイレックス、トキサフェン、PCB、ヘキサクロロベンゼン、ダイオキシンおよびフuranの12物質がPOPsとして規制されている。

最近発見して使われるDDT (POPsの1つ)は、量が多くなると大気中に気化し、冬になると地表に沈着することを繰り返して長距離を運ばれ、本来DDTが使われたことのない地域や種類の生物から高濃度で検出され、その影響が懸念された。

この解説に含まれる用語

- 北極
- 難分解性
- 南極
- 生態系
- 食物連鎖
- 異化剤
- ヘキサクロロベンゼン
- ディルドリン
- ダイオキシン
- POPsの略
- PCB
- DDT

「残留性有機汚染物質」の関連情報を検索

- 国際ニュース
- 海外ニュース
- イベント情報
- 環境Q&A
- 機関情報
- 環境年表

「残留性有機汚染物質」の関連用語を表示

There are a number of thesaurus, but without harmonization

Ontology



- Ontologies have gained importance in many areas of applications
 - interoperability, knowledge management
 - design of intelligent systems
- Used for a more precise specification of the semantic information content of the underlying data
- Domain specific ontologies are also being viewed as vehicles of capturing semantic information content

Methodology (1/2)

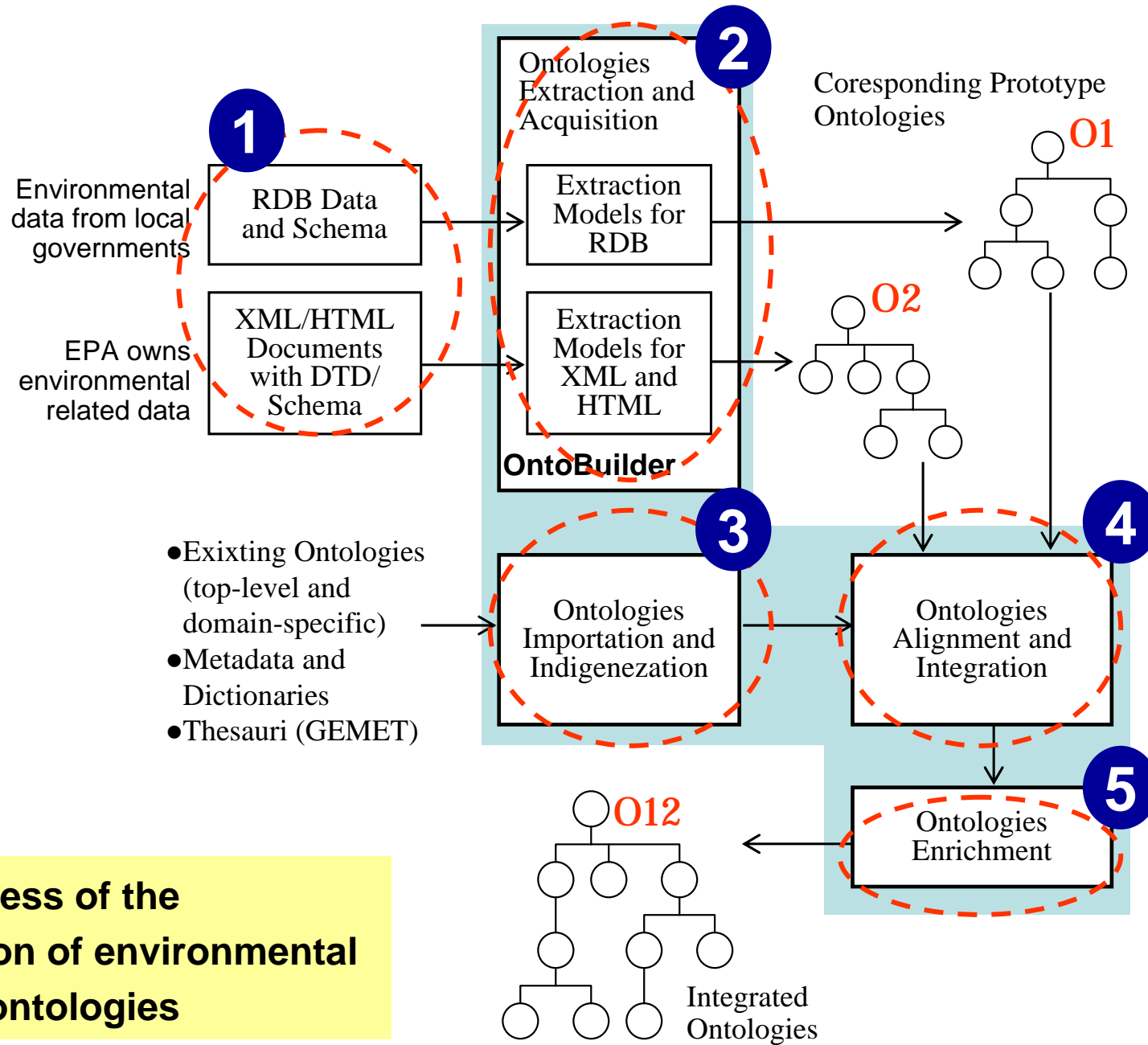


- Purpose and principles
 - ontologies to be practical and useful artefacts
 - the effort required to construct new ontologies must be minimized
 - overall effort required to construct an ontology must be amortized over multiple uses and users

Methodology (2/2)



- First, we provide an environment
 - assists users to build **prototype ontologies**
 - extracts knowledge from the existing information sources such as DB schemas or XML documents
- Secondly, for the built prototype ontologies
 - align with existing common ontologies and some of the domain specific thesauri such as GEMET
 - provide a semantic consistency among different ontologies
- Thirdly, the aligned ontologies will be merged and combined to form an **integrated ontology**



The process of the integration of environmental domain ontologies

Systematic processes (1/5)



- Selection of underlying information sources, standards, laws and regulations, classifications, etc.
 - select the sources of information relevant to the target domain
 - usually provide **taxonomy of concepts** and terminologies used in the domain

Systematic processes (2/5)



- Ontology extraction and acquisition
 - knowledge acquisition from the sources of information
 - form a prototype ontology for each knowledge source
 - employ the procedure of “**protocol analysis**” with domain experts
 - asking users to describe various types of domain applications, the data used in such applications, and the terms used in their field

Systematic processes (3/5)



- Ontology importation and indigenization
 - import ontologies that are existing ontologies in target domains but might usually be used in other countries or areas
 - tailor them to fit the feature in the area or country where ontologies will be applied
 - For example:
 - OECD classification
 - GEMET thesaurus

localized to fit in other countries or regions

Systematic processes (4/5)



- Ontology alignment and integration
 - align the prototype ontologies with imported ontologies or **upper level ontologies**
 - adjusting the name or terminology in the prototype ontologies
 - making them consistent with each other
 - combine and merge the prototype ontologies to form an integrated ontology

Systematic processes (5/5)



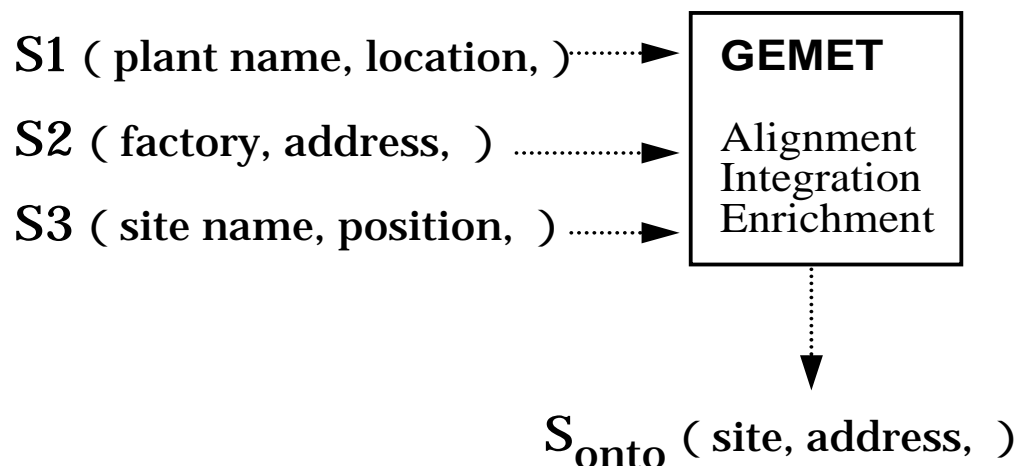
- Ontology enrichment
 - most of the ontologies merely represent taxonomy of concepts, where others may just include some attributes for them.
 - In this activity, we will enrich the integrated ontologies with extra information where available (attributes, features, ...)

Example application



- Environmental Data Repository (EDR)
 - an integrated data warehouse system that provides a single point of access to data extracted from several major Taiwan EPA databases
 - construct the integrated ontology by extracting the domain knowledge from some of the information sources and aligning the concepts in the ontology with the laws and regulations of Taiwan EPA

an example of term integration



- each information source might maintain the data regarding the potential pollution sites
- conflict between the terms that is used to identify the sites
- ontology can assist us in overcoming this conflict, and develop a **consistent view** through which information can be integrated

EDR Homepage (http://edb.epa.gov.tw)

環境資料庫 - Microsoft Internet Explorer

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

行政院環境保護署
環境資料庫

簡介 | 資訊管理規範 | 意見回饋 | 常見問答集 | 軟體下載 | 相關連結

資料庫主選

空氣
噪音
水
土壤
廢棄物
毒化物
資源回收
環境用藥
非屬原子能
游離輻射
環境統計
列管場所
地方環境資料
地理資訊系統
回首頁
回環保署首頁

按資料類型 查詢

操作說明 | 進階查詢 | 資料瀏覽

■ 環保新聞

- 環保署致力提升環境品質
- EPA CELEBRATES ITS EIGHTEENTH
- 環保署慶18周年 提出環保回收新
- 環保署十八週年慶 張國龍提三大
- 防止居家鼠害 環署提供門智小撇

更多

■ 最新消息

- 2月24日本署辦理台美合作「環境資訊系統建置與管理」研討會簡報下載。

參觀人數：0009131

精選影像

台灣地區空氣品質
_940821_Tera
除基隆台北地區臭氧濃
度稍高外，其餘地區空
氣品質良好。

更多

環境詞彙庫
Environmental Terminology
Reference List

公聽會
會議記錄

subject-oriented
information retrieval

environmental
domain ontology

Conclusions



- a methodology based on ontology-driven related technologies to integrate environmental information
- It is shown, to some degree, that ontology can provide assistance in solving the heterogeneous problems among diverse information sources
- The approach may serve as an infrastructure component for integrating environmental data with known, but differing, collections of data

Future work



- recent advancements including:
 - Web services,
 - Ajax and Web 2.0 related technologies
 - Knowledge managementmight be integrated with the proposed approach
- design and implement a more sophisticated and practical system



Thank you!