

# EDEN (Environmental Data Exchange Network)

## *An InfoSleuth Application*

*Hannu Saarenmaa, European Environment Agency  
Bruce Bargmeyer, US Environment Protection Agency  
Warren Meekins, US Department of Defence  
Robert Murray, US Department of Energy  
Greg Pitts, MCC*

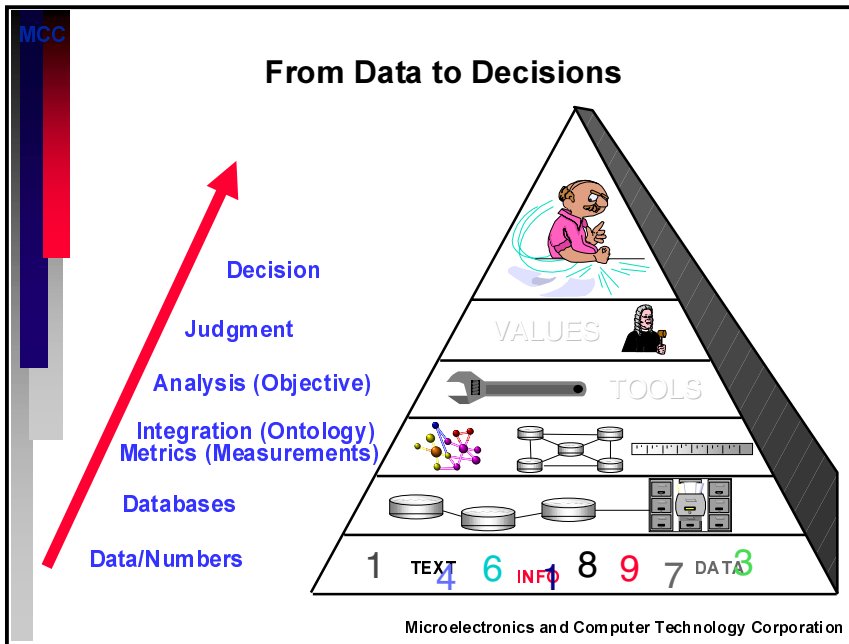
**Resource Technology '98 Symposium  
Rovaniemi, Finland, June 8-12, 1998**

Microelectronics and Computer Technology Corporation

## The Problem: On-line Access to Heterogeneous Databases on Internet

- Internet is valuable because locating information is easy
  - Universal Resource Locators
  - AltaVista Search, Yahoo!, HotBot, etc, have created a new culture
- Also locating data may be possible under some circumstances
  - GILS, G7-ENRM GELOS, Catalogue of Data Sources, Environment Data Registry
- However, *accessing* data is not readily available now
  - Data harmonization needed, differences in data models, coding and concepts to be overcome
  - Distributed data warehousing: access data in place - not only its copy

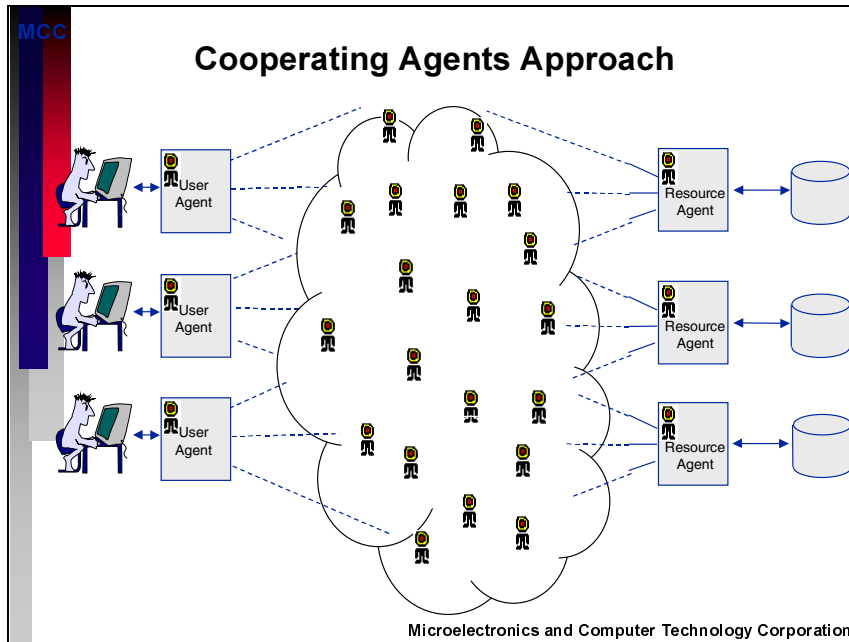
Microelectronics and Computer Technology Corporation



**What is InfoSleuth?**

- Integration of new technological developments for:
  - Locating relevant information based on an advertise/discover model
  - Querying based on a common domain vocabulary
- Software system utilizing intelligent agents, providing capabilities for:
  - Data analysis, simulation, what-if analysis
  - Derived information, aggregation, and drill-down
  - Data mining, pattern/trend detection, knowledge discovery
  - Subscription/Notification services
  - Event detection and monitoring

Microelectronics and Computer Technology Corporation



**MCC**

### MCC InfoSleuth Development

- Result of a multi-year research and technology development project funded by industry and government contracts
  - Sponsors: Computing Devices Int'l, Eastman Chemical, Hughes, National Security Agency, Texas Instruments, Motorola
- Tested in strategic applications, including:
  - Eastman Chemical: daily use by non-computer specialists to access polymer databases
  - NIST Health care: access to health care information including federal, state and local information sources
  - NSA: integration of multimedia databases and unstructured (multilingual) text documents
  - DARPA: BADD (satellite-based information dissemination service) and Advanced Logistics Program

Microelectronics and Computer Technology Corporation

MCC

## What is EDEN?

- EDEN is application of InfoSleuth technology to environmental problems, because...
- Solving environmental problems requires integration of data, information and knowledge from a multitude of sources and from different networks
- There are multiple heterogeneous information resources
  - Diverse hardware and software
  - Heterogeneity in representation of information
  - Permanent, dynamic and opportunistic data
- There is a need to support multiple (and conflicting) classes of applications:
  - Reporting, analysis, research, enforcement, etc.
- There is a need to respect autonomy of information owners
  - Integrated and federated data warehouses are expensive

Microelectronics and Computer Technology Corporation

MCC

## Institutional Objectives of EDEN

- Collaborate on the development and demonstration of an improved means for sharing environmental data
- EPA
  - Take full advantage of existing investments in information and data without duplication of effort at each user site
- DoD
  - Harness the technology of the military for environmental cleanup, pollution prevention, etc.
- DoE
  - Share information without incurring the cost of restructuring existing data systems or building new ones
- EEA
  - Harmonization of diverse European data, maximizing use of the Catalogue of Data Sources

Microelectronics and Computer Technology Corporation

MCC

## Common Set of Requirements

- Reduce the reporting burden imposed by the parties on each other
- Sharing of best available and most timely information
- Enable users to access information from multiple sources
- Coordinate only the common vocabulary – not the end use of information resources; focus on the inputs with each participant; individually interpreting and communicating outputs

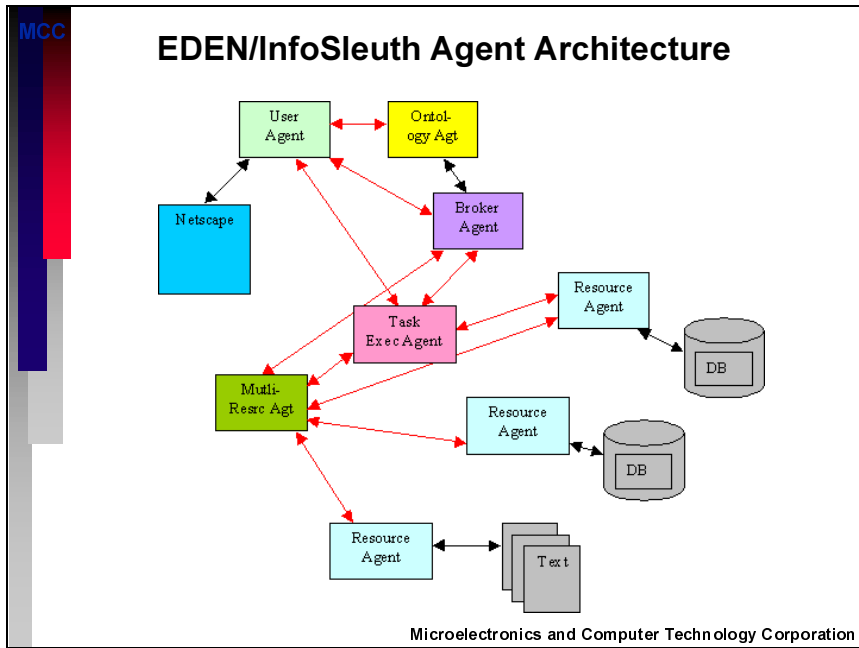
Microelectronics and Computer Technology Corporation

MCC

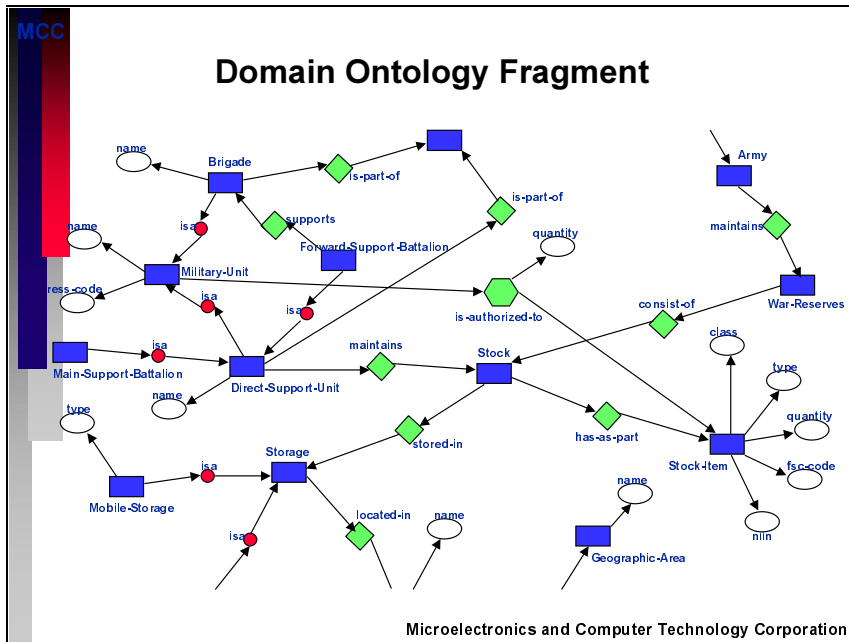
## EDEN Technical Approach

- Build upon agency developments for organizing and describing data
  - The EPA Environmental Data Registry (EDR) for describes data elements and is the foundation for a semantic, conceptual data model
  - The EEA Thesaurus and EIONET provide potential multilingual capability and
- Utilize MCC InfoSleuth technology to achieve uniform access to environmental information resources
  - Access to a specific set of geographically distributed data sources in situ with minimal impact
  - Utilize Internet-capable browsers for cross-platform compatibility

Microelectronics and Computer Technology Corporation



- EDEN Pilot Databases**
- **Hazardous Waste**
    - Target of pilot application
  - **Databases from each partner**
    - CERCLIS
    - ITT
    - HazDat
    - ERPIMS
    - Basel Convention Database
  - **Five disparate databases and data models**
    - One common ontology created
    - GEMET: Common multilingual thesaurus of environmental terms
- Microelectronics and Computer Technology Corporation



- 
- Queries Based on Ontology**
- Queries spanned the five databases
  - Tailored to address concerns of four user types:
    - Citizen, Regulator, Scientist, Site Manager
  - Examples:
    - *“cleanup technologies to treat most frequently occurring contaminants detected in various media”*
    - *“health effects of contaminants found at sites on the national priority list”*
    - *“sites where cleanup of top contaminants was completed”*
    - *“hazardous waste transported to europe”*
  - Query user interface created to launch queries
    - Results viewed in tables
    - Results can be cached and drilled down further
- Microelectronics and Computer Technology Corporation

MCC

## How to Create an EDEN Infrastructure

- Create a conceptual model, or ontology, that corresponds to the EDR and EEA Thesaurus in the specified domain
- Provide capability for browsing and populating an ontology, mapping and advertising data, query formulating and executing
- Develop resource agents to map sources of environmental information to the common vocabulary
- Develop user agents with query, browsing and data analysis applets
- Document relatability to proposed ANSI X3L8 and ISO SC14 data standards and describe the attributes
- Investigate the utility of integration of data analysis capabilities with agents and web-based text
- Demonstrate a pilot system

Microelectronics and Computer Technology Corporation

MCC

## Conclusion

- InfoSleuth is an attempt to go beyond homepages on remote data access
- Scalable architecture to tie diverse data sources to each other
- EDEN is an approach to access to heterogeneous environmental data
- Environmental professionals need data from multiple sectors - and hence EDEN
- Pilot project on hazardous waste has been promising

Microelectronics and Computer Technology Corporation