

Common Ground: Agile Data Sharing

Ken Smith

(703) 983-6115 • kps@mitre.org

MITRE Sponsored Research



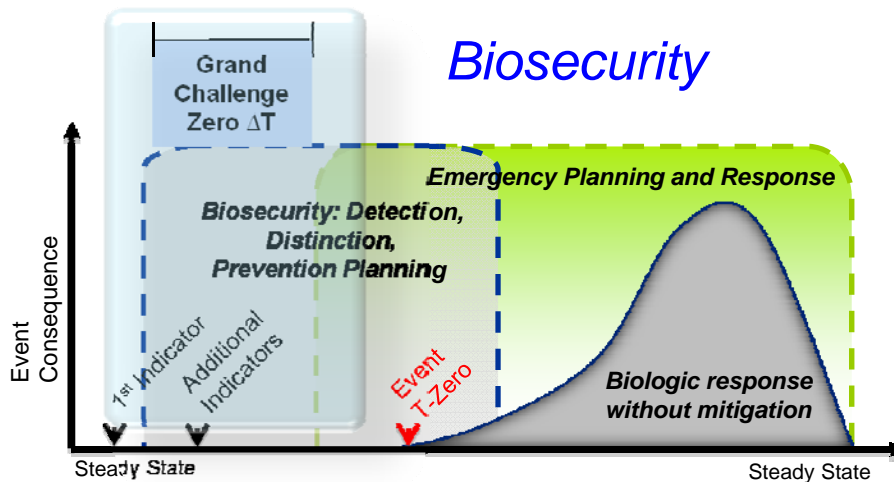
Problem



- **Data sharing is often prohibitively expensive in time and labor.**
- **Yet, many high-value data sharing activities are urgent, or funding-constrained.**
- **How can we “lower the activation energy” for data sharing?**

Background

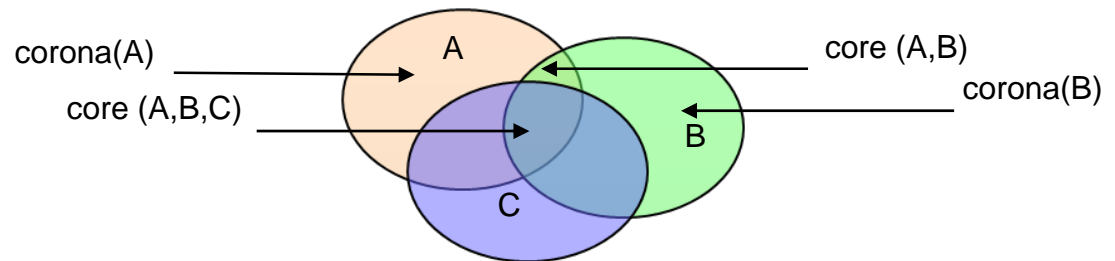
Producing exhaustive agreement disallows vital sharing!



Emergency Preparedness & Response



But sharing can start sooner



Objective



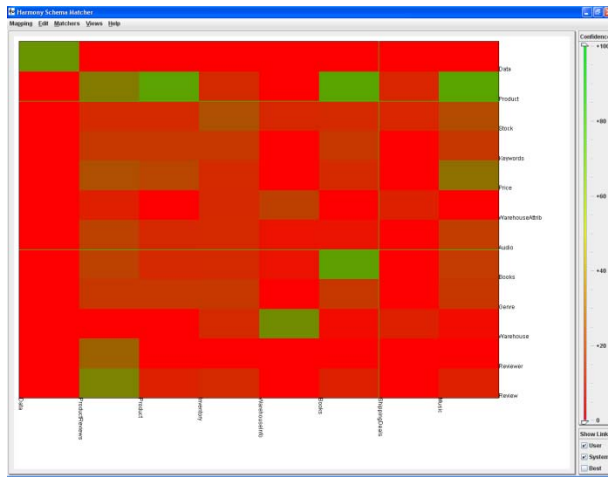
- **Rapidly identify a semantically-important “core” set of attributes from the data models of sharing partners, thus:**
 - a) Reducing the problem size
 - b) Focusing limited integration resources
 - c) Building trust and momentum by “getting in the game” sooner.
- **Design and build flexible “Common Ground Workshop”**
 - Tools enabling rapid development of “starting point” core att.s
 - Enable effective sharing during both preparation & response.
- **FY09 Foci:**
 - Harmony 2.0, Visualizations, CG Workshop design / prototyping.

Activities



- **Developed, distributed [Harmony 2.0](#)**
 - Incorporated schema repository, graph-based memory interface
 - Substantially enhanced GUI, based on user-feedback
 - Tested extensively on massive customer (JFCOM) schemas
 - Distributed to waiting customers, released on *code.google.com*.
- **Designing [Common Ground Workshop](#)**
 - Tool for scaling data sharing to many partners
 - Steps: 1) cluster & visualize, 2) match within cluster, 3) form & visualize comprehensive vocabulary
 - Integrating via OpenII / Eclipse framework.
- **Researching [schema cluster visualizations](#)**
 - Examined 12 strategies; selected: a) interactive dendrogram, b) Jaccard / non-metric MDS, c) topographic view.

Highlight: Match Visualizations

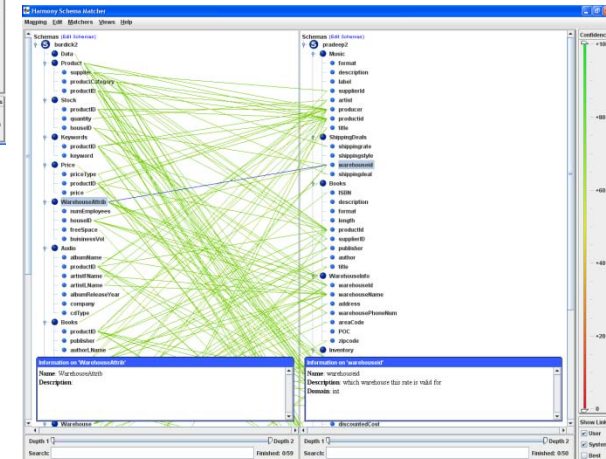


HeatMap View (Aggregated) enables big picture analysis for large matches, highlights regions of intense matching (i.e., green), assists in schema summarization, entity resolution, & understanding.



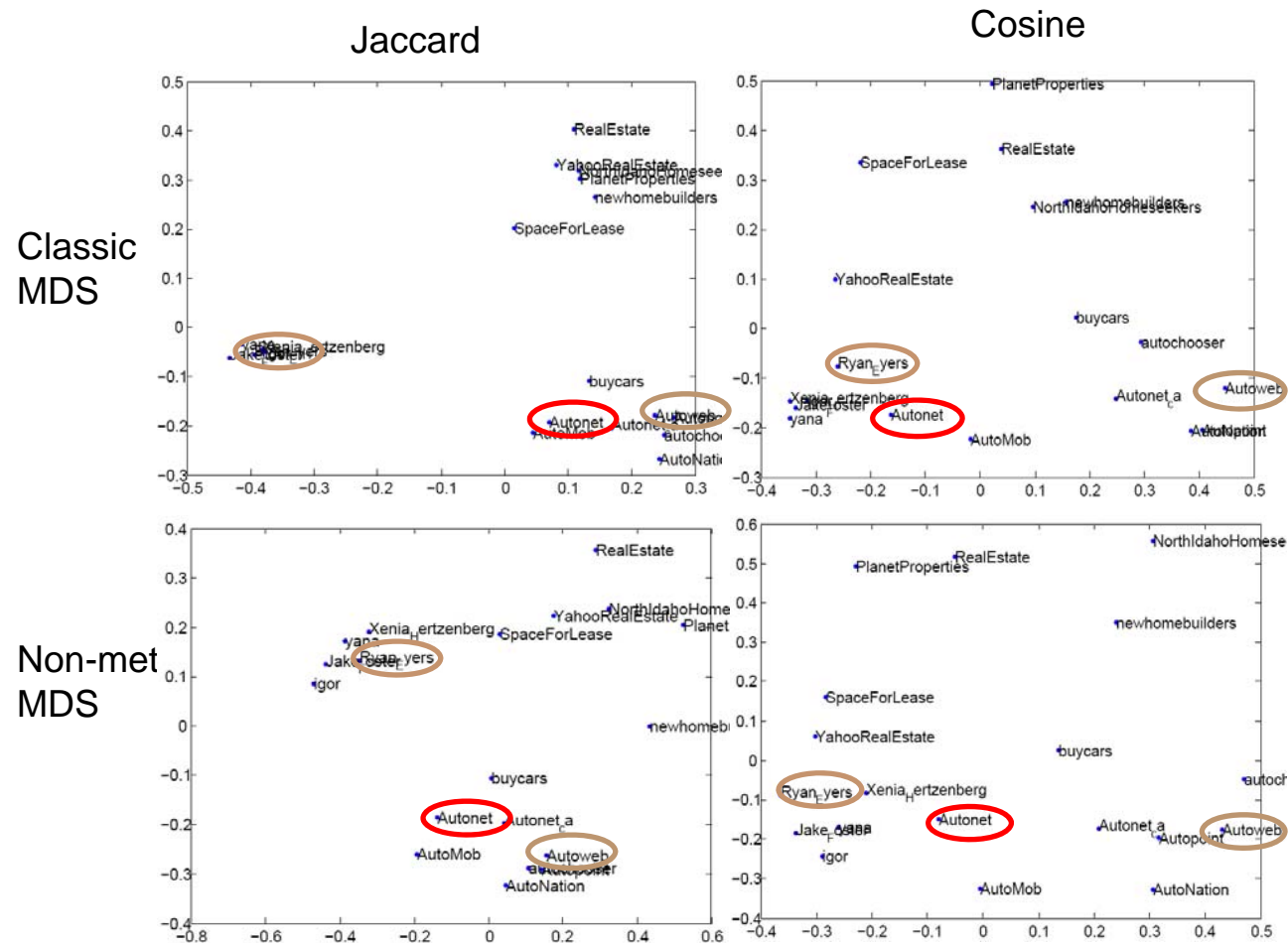
HeatMap View (Detailed) provides drill down to inspect detailed concepts in promising regions. Helps reveal meaningful match patterns (e.g., ids to ids).

Schema View traditional match view; enables focused matching on specific concepts, retains schema context of matched elements; permits inspection of documentation, annotation with match notes



Highlight: Schema Cluster Visualizations

- **Goal:** Transform high-dimensional distances to 2D/3D cluster visualizations.
- **Shown:** Four very promising visualizations.
- **Note:** “Autonet” placement shows subtle variations.



Impacts



- **The Common Ground Workshop will help MITRE’s sponsors share high value data under pressing time, money constraints.**
- **Tools (Harmony 2.0, SchemaStore) can assist any sharing effort.**
- **Presented “The Role of Schema Matching in Large Enterprises” at CIDR ‘09 (Conference on Innovative Database Research).**
- **Contributing to MITRE’s role in OpenII (with Google, UPenn, etc.).**
- **Recent user interactions:**
 - **DHS I&A (enterprise information asset awareness)**
 - **VA-DoD EHR sharing (fast, cheap adaptor construction)**
 - **USJFCOM (vocabulary alignment)**
 - **DCGS-A (data source integration planning)**
 - **JHU neuroimage consortium (heterogeneity reduction).**

Future Plans

