SHER: Semantic Databases using ontologies

Julian Dolby, Achille Fokoue, Aditya Kalyanpur, Aaron Kershenbaum, Li Ma, Edith Schonberg, Kavitha Srinivas

Watson/Hawthorne, China Research Lab

SHER – Semantic DB using ontologies

Breakthrough technology that:

- Is highly scalable -- reasons on 7.7M triples in 7.9 s on benchmarks, scales to 60M triples.
- Can cleanse inconsistencies in noisy data. Identifies thousands of logically inconsistent patterns in minutes.
- Provides explanations of the chain of semantic reasoning for a result set.

Who is interested in SHER?

- Government:
  - Ordnance survey, the British national mapping agency
  - NSA
- Healthcare and informatics:
  - Mayo Clinic, Chris Chute, Chair of Biomedical Informatics
  - Vanderbilt Medical Center (Dan Masys, Director of Biomedical Informatics),
  - Ohio University Medical Center (Philip Payne)
  - Columbia University Medical Center (Clinical and Translational Award Center)
- Pharmaceutical industry:
  - Pfizer
- Telecom:
  - DoCoMo, a mobile services client.
- Software/Services Vendors:
  - Clark-Parsia, a semantic consulting services company, interest in licensing/subcontracting to SHER.
  - RacerPro (Franz Inc.) interest in licensing SHER.

When are semantic DBs useful?

Complex knowledge domains, where there is a semantic gap between data and queries

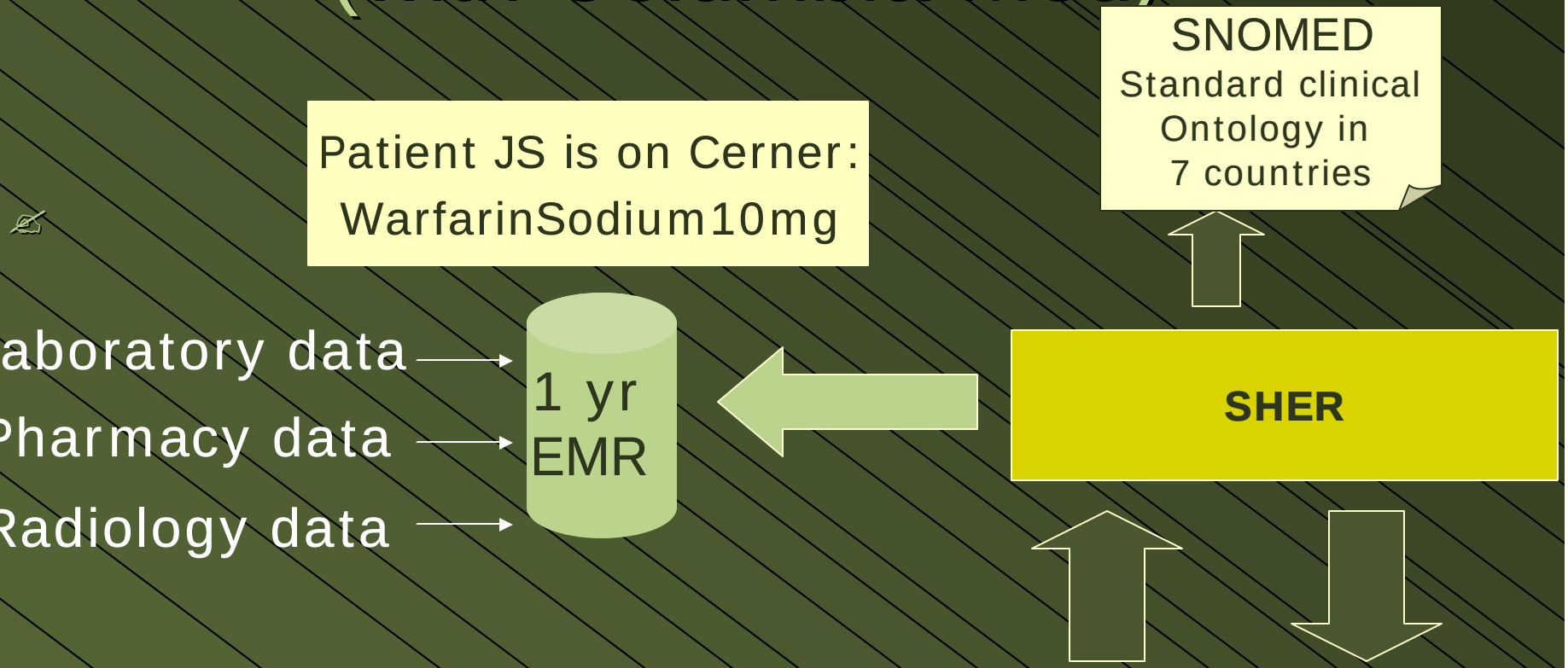
Example from healthcare domain -- matching patient records to clinical trials, clinical decision support:

Patient data:	Queries:
Patient on methotrexate	Patients on immunosuppresants
Patient tested positive for tuberculosis	Patients with
mycobacterium tuberculosis	meningitis.

Example from pharmaceutical domain (semantic querying of the metadata on microarray gene expression data):

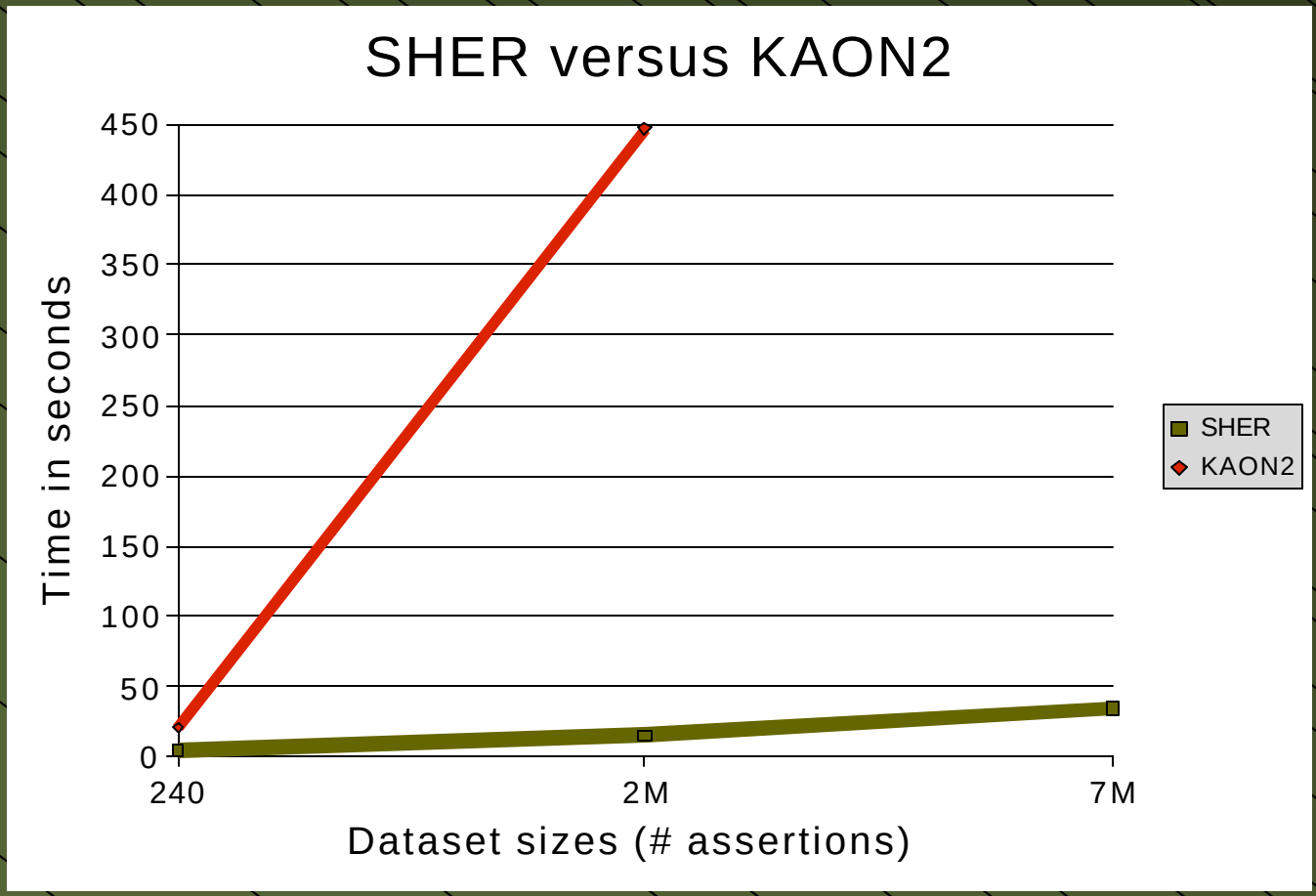
Gene Expression data	Queries
GSE1402 is data about arthritis data on disorders	Gene expression

Clinical trials matching case study (with Columbia Med)



Clinical trial queries in SNOMED: Patients on drugs with active ingredient of warfarin? JS

SHER Scalability Results vs. state of the art



Performance of SHER vs state of the art (KAON2) on OWL benchmark – KAON2 fails on 7M (112 queries) – AAAI 2007.