

# Foundations of Next Generation Information Access

Warren Greiff

781-271-7421 • [greiff@mitre.org](mailto:greiff@mitre.org)

MITRE Sponsored Research

The logo for the MITRE Technology Program, featuring a stylized graphic of stacked blocks in yellow, orange, and blue to the left of the text.

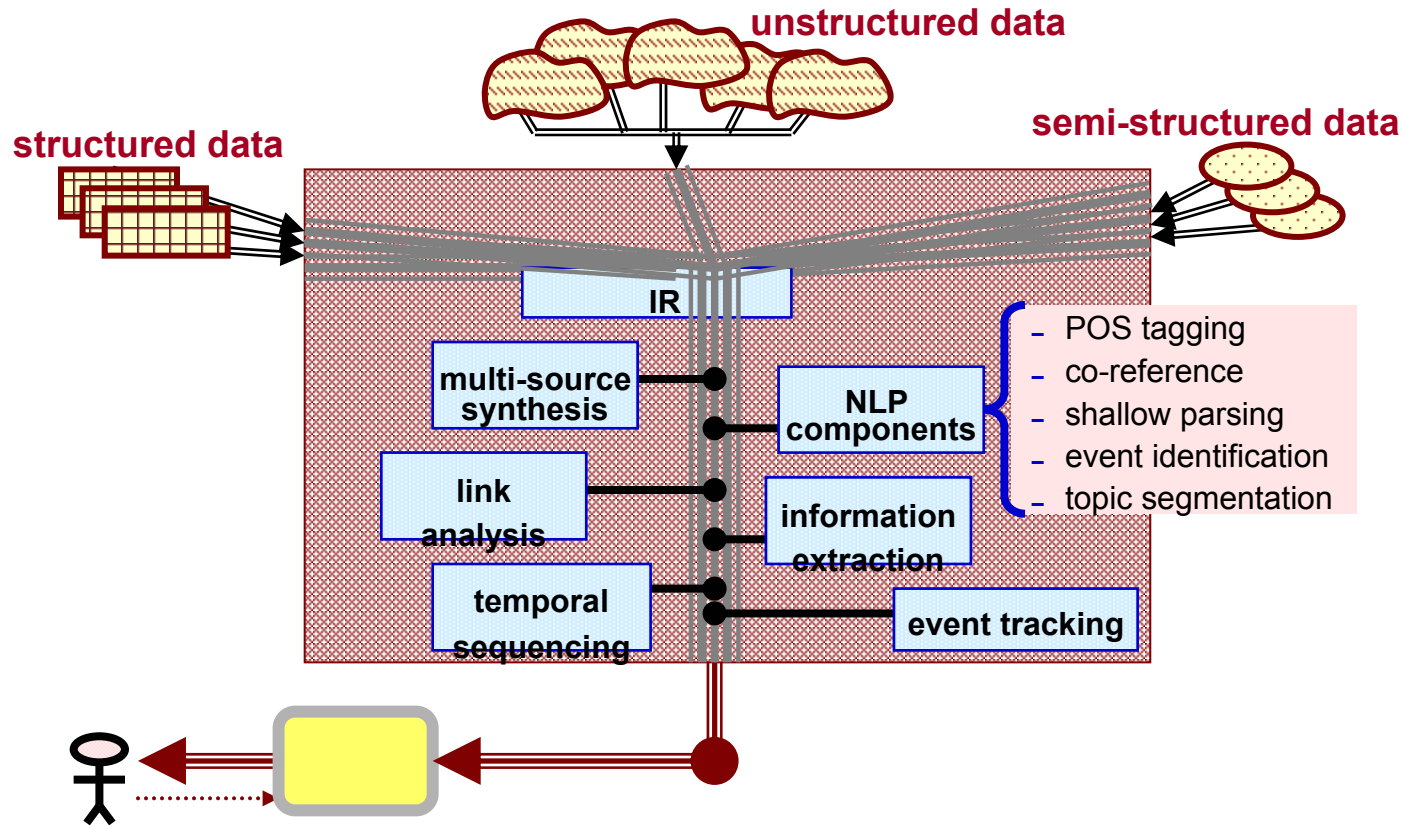
**MITRE**  
**Technology**  
**Program**

# Problem

- **Overwhelming need for computerized support for information gathering**
  - particularly in areas of:
    - military readiness
    - terrorism response
    - disease control
- **Current technology for addressing problem:**
  - fragmented across research communities
  - integration is difficult

- ***Needed:***
  - *underlying formalism*
  - *cutting across different technologies*

# Background



**Next Generation Information Access**  
will require principled approaches to the  
integration of multiple component technologies.

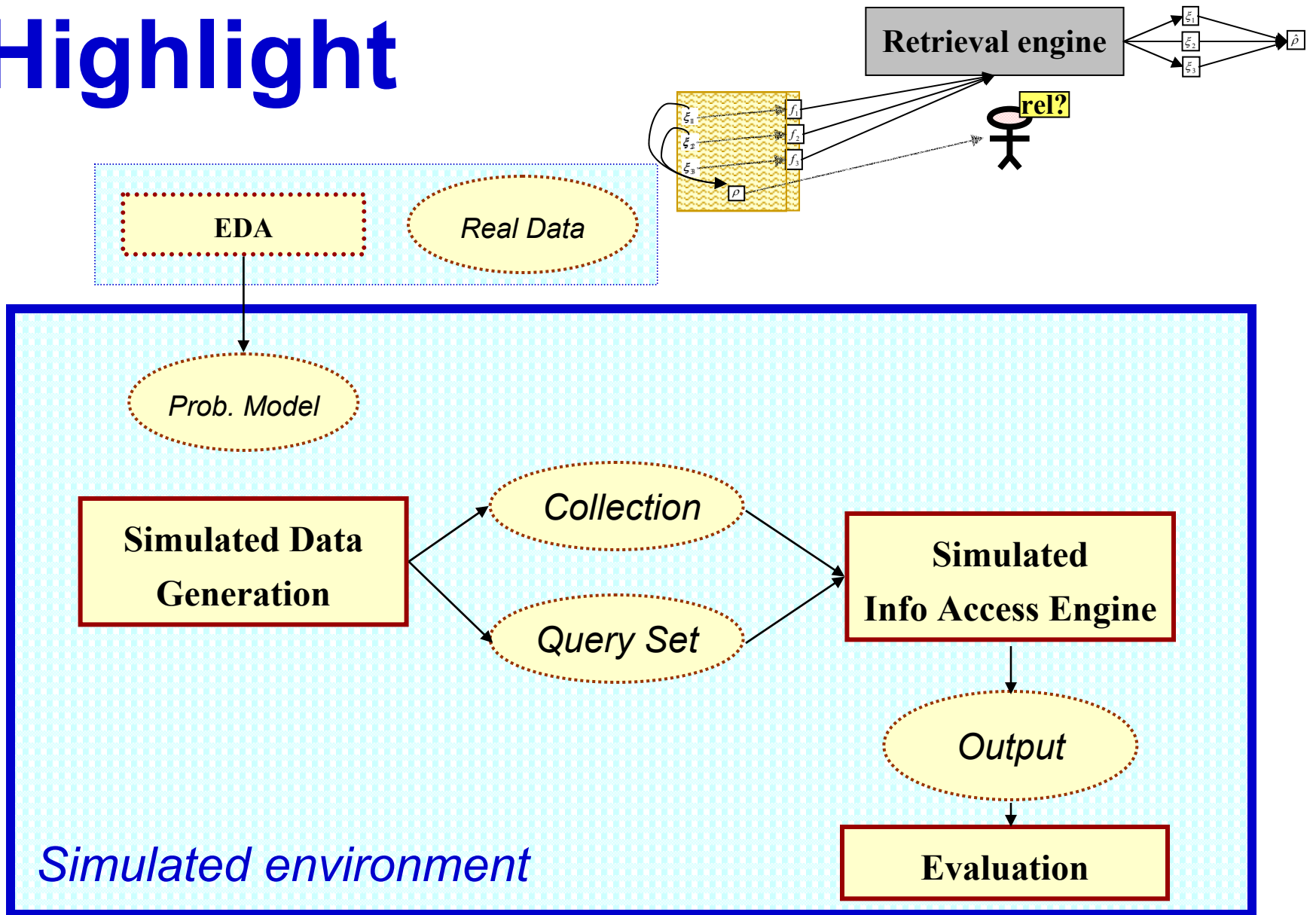
# Objective

- **Statistical foundations for information access**
  - rigorous characterizations of:
    - probabilistic modeling
    - parameter estimation
- **Principled methodologies for adapting to**
  - new languages
  - new genres
  - new information domains
  - new tasks
  - auxiliary knowledge sources
- ➔ **Exploratory analysis of retrieval data**

# Activities

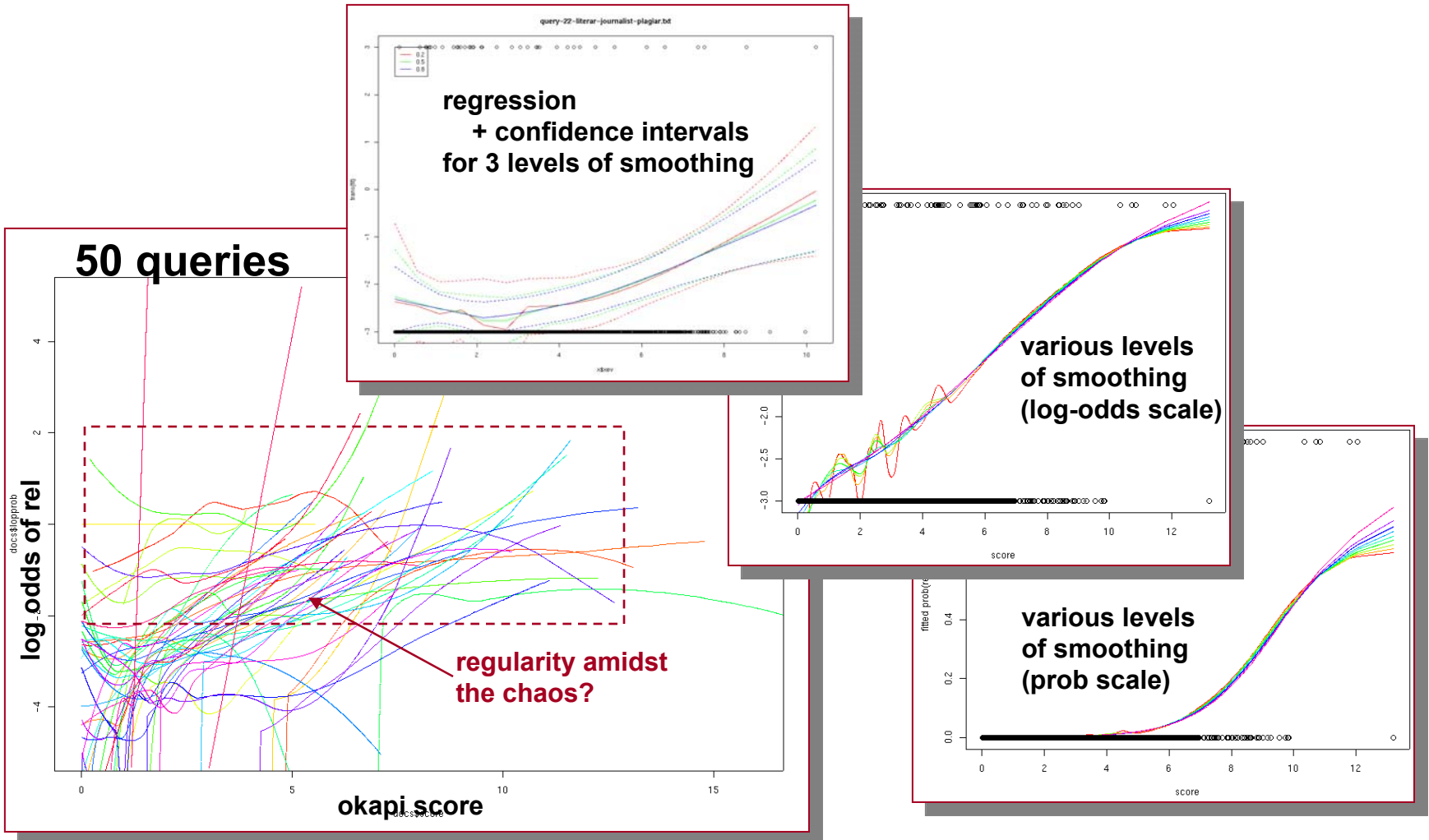
- **Simulated stochastic modeling of:**
  - latent document features
  - observable document features
  - document relevance
  - queries
- **Exploratory data analysis**
  - available research corpora
  - verification of simulation models
- **Central focus -- research into**
  - variance reduction
  - empirical Bayes techniques

# Highlight



Environment for Information Access Simulation Studies

# Highlight



**Study of probability of relevance vs. OKAPI for 50 TREC queries**

# Impacts

- **Direct relevance to existing MITRE projects**
- **Information access systems development**
  - **incorporation of new sources of evidence**
  - **tailored to specific sponsor needs**
- **Scientific contribution**
- **Community participation**
  - **CIKM, SIGIR, COLING, IPM, TREC, ARDA**
  - **NRRC summer workshop on Information Retrieval**

# Future Plans

