IARPA

BE THE FUTURE

FUSE

Foresight and Understanding from Scientific Exposition

Dewey Murdick, Program Manager

FUSE Proposers' Day, 17 June 2010

Disclaimer

- This presentation is provided solely for information and planning purposes.
- The Proposers' Day Conference does not constitute a formal solicitation for proposals or proposal abstracts.
- Nothing said at Proposers' Day changes the requirements set forth in a BAA.
- BAA supersedes anything presented or said at the Proposers' Day by IARPA

Proposers' Day Goals

- Familiarize participants with IARPA's interest in time-dependent, pattern-based indicators of scientific and technical capability emergence using information from the publicly available technical literature.
- Please ask questions and provide feedback; this is your chance to alter the course of events.
- Foster discussion of synergistic capabilities among potential program participants, AKA teaming. Take a chance: someone might have a missing piece of your puzzle.

Schedule

- Once BAA is released, questions can only be answered in writing on the program website.
- Full Proposals are due ~45 days after BAA is published.

FUSE Overview

- **Goal:** Automated detection of emerging scientific and technical capabilities using the worldwide scientific, technical, and patent literature
- **Key Technical Challenges:**
  - Identify and extract observables from tens of millions of full-text documents in multiple languages
  - Generate and disambiguate features across a massive, growing, noisy, and incomplete set of documents
  - Develop reliable models of capability emergence
  - Nominate technical areas with emerging capability

Complete, Continuous, Unbiased

Today	FUSE
Manual	Automatic
Selected coverage	100% literature coverage
Updated infrequently	Updated monthly
Months to produce (for one technical area)	24hrs to produce (for all technical areas)
Ad hoc evaluation	Formal models of emergence

Current Situation – Emerging Capabilities

- **Search Limitations**
  - Data often not comprehensive
    (using 10k metadata-only records not 10M full-text docs)
  - Search for key phrases on pre-identified technical areas
  - Evaluate one area at a time (many manual steps)
  - Cross-field connections often missed
  - No multivariate temporal pattern “queries” at scale
  - No reliable models for important pattern combinations

- **Full-text NOT being used systematically at scale**
  (exception: information retrieval systems)

A time-consuming, domain-specific, expert-intensive process, frequently done under severe time constraints without a systematic, reproducible audit trail or bias control using limited tools against an overwhelming information deluge

Use Case: Analysts need…

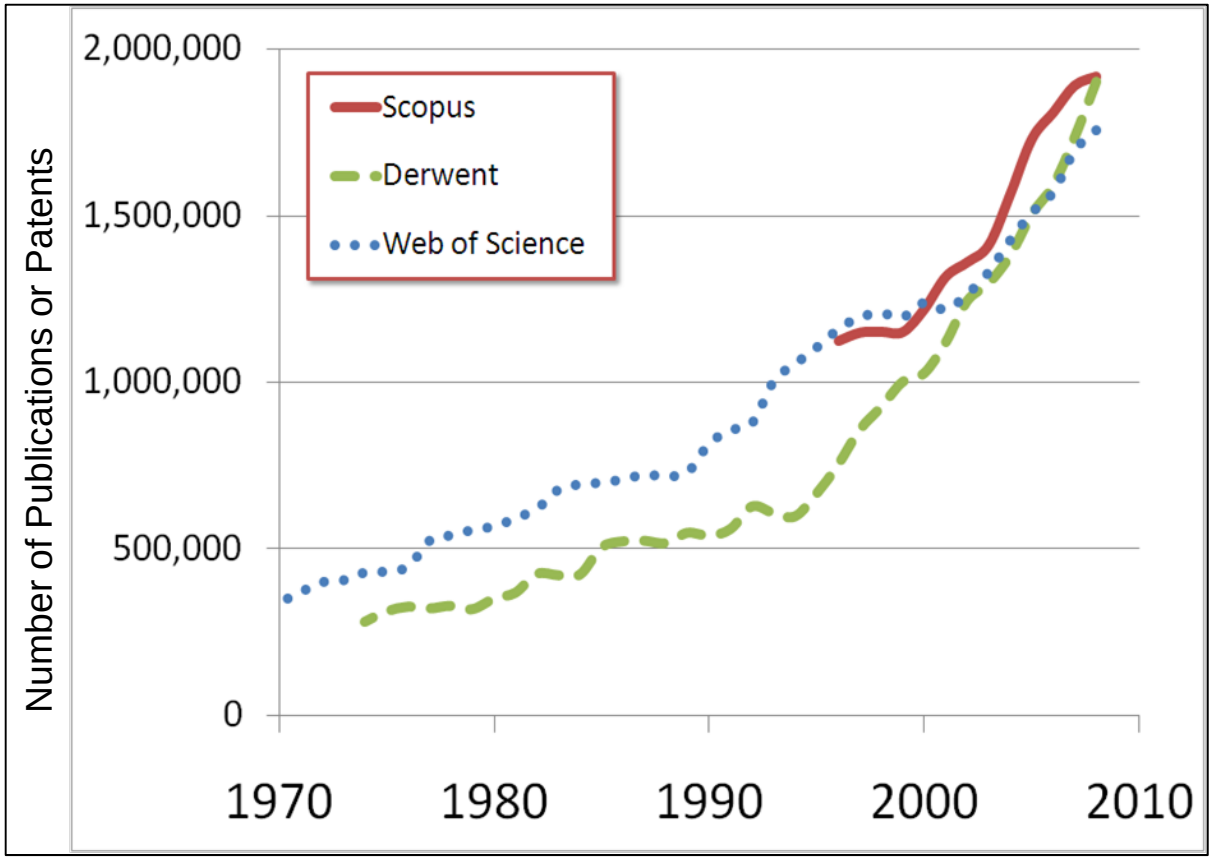
- An automated, reliable, and transparent tool that nominates important technical areas for further analysis based on realistic and validated measures of emerging capability
- Alerts for unknown technical areas with sufficient auditable evidence to support further exploration
- Discovery of trends and connections in an exponentially growing flood of data at a speed, scale, and comprehensiveness that exceeds human capacity
- Pathways to find supporting evidence and to explore related model sensitivity to particular evidence… must provide a path to understand the “black boxes”

Worldwide Sci, Tech & Patent "Lit. Space"

Top Languages
(English Language Indexers*)

Language	Pubs/Patents
English	~55M / 6M
German	~1.5M / 3M
French	~1M / 1M
Russian	~1M / 1.5M
Chinese	~650k / 2M
Japanese	~350k / 8M
Spanish	~300k / 500k

*Many additional non-English collections



Source: Thomson Reuters Web of Science® (>10k journals & >100k conference proceedings, 1900-present) and Derwent World Patents Index® (41 patent issuing authorities, 1970s-present) , Elsevier B.V. Scopus® (18k journals & 3.6M conference papers, 1996-present)

“Lit. Space” (Continued)

- **Formal technical discourse has natural (somewhat standard) structure**
  - Prior Art / Background
  - Motivation / Goal
  - Applications (potential or realized)
  - Methods / Procedures / Equipment / Data
  - Claims
  - Observation / Evidence
  - Conclusions / Result
  - Collaborators, Organizations, and Geographic Locations
  - Funding sources
  - References
- **Common structure of these materials can be leveraged**

Definitions within FUSE

Real World

- **Technical Capability** is the ability to organize knowledge about the world, condense it into testable laws and theories, and apply it in an algorithmic or physical way
- **Technical Capability Emergence** is the process of discovery / invention, exploration, adoption, application, and diffusion of a technical capability

Literature Space

- **Technical Area (TA)** is a cohesive, time-evolving concept or domain that is represented in lit. space by a set of related documents (100 documents or orders of magnitude larger)
- **Evidence of Technical Capability Emergence** are observable evolutionary or discontinuous indicators that can be measured in lit. space *and may occur within a single TA / similar set of TAs (Emergence) or across a set of dissimilar TAs (Co-Emergence)*

Example: Genetic Algorithms

"Genetic algorithms are evolutionary inspired techniques used in computing to find exact or approximate solutions to optimization and search problems by using inheritance, mutation, selection, and crossover."

- **Is there a capability development trigger?**
  - 1950s-1960s: 1st articles in evolution-inspired algorithms appear (little follow-up)
  - 1962: Crossover and recombination operators first emerge (Holland et al.)
  - 1966: Evolutionary programming concepts introduced (Fogel et al.)
  - 1975: "Adaptation in Natural and Artificial Systems" published (Holland) and dissertation shows wide variety of functionality (De Jong)

Source: http://www.talkorigins.org/faqs/genalg/genalg.html

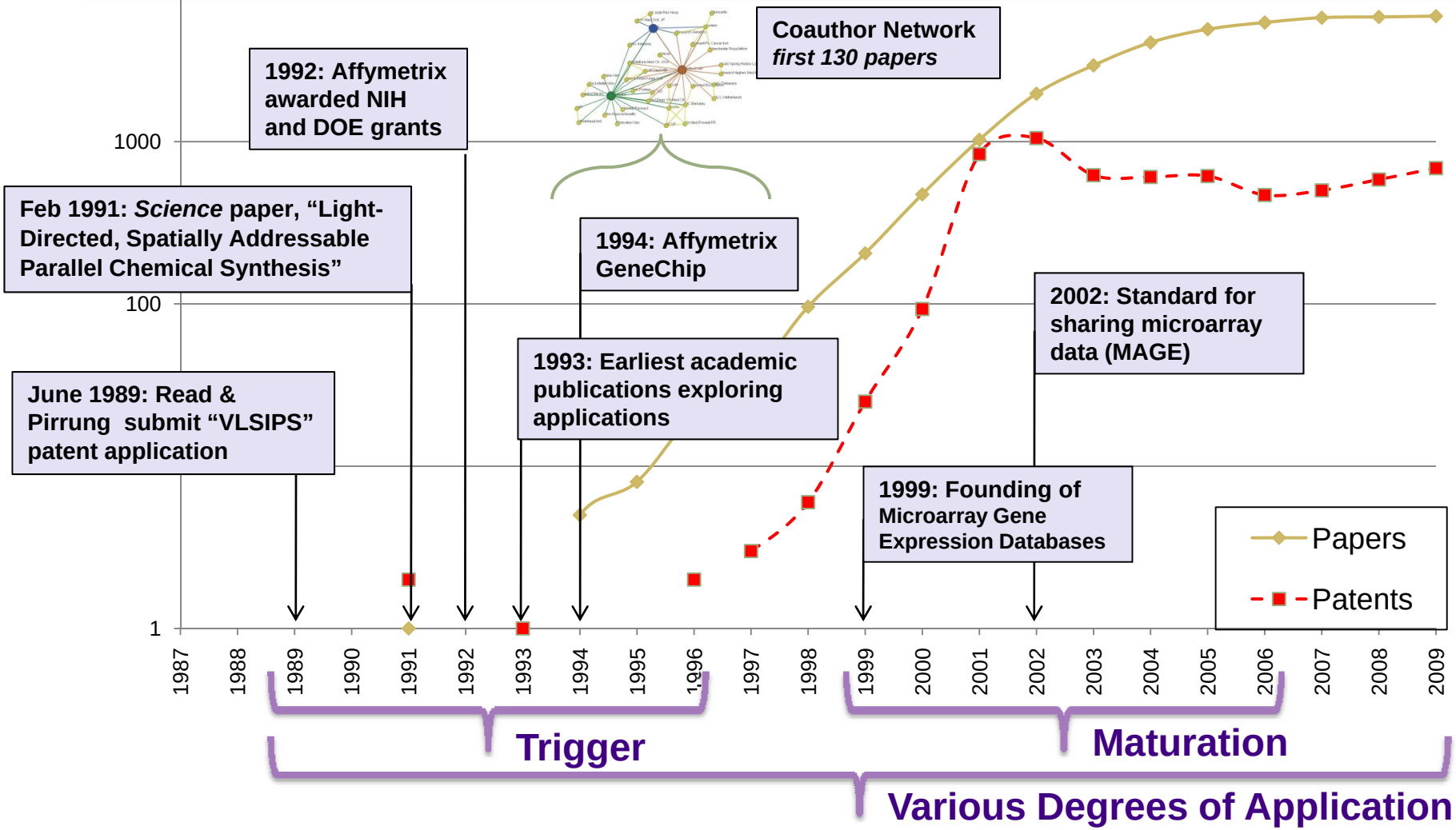
Genetic Algorithms Example (Continued)

- **Is there evidence of capability maturation and impact?**
  - 1985: 1st Int'l Conf on Genetic Algorithms and Applications
  - 1988: Machine Learning special double issue
  - July 1992: Scientific American article, excitement about capability
  - 1980s-1990s (enabling conditions): Increase in computing power
  - Increasing usage trend in technical papers as successful method
- **Is there evidence of the application of a capability?**
  - 1980s and beyond: Applied to a broad range of subjects
    - stock market prediction and portfolio planning
    - aerospace engineering
    - microchip design
    - biochemistry and molecular biology
    - scheduling at airports and assembly lines

Capability emerged from within one technical area and has been applied to many

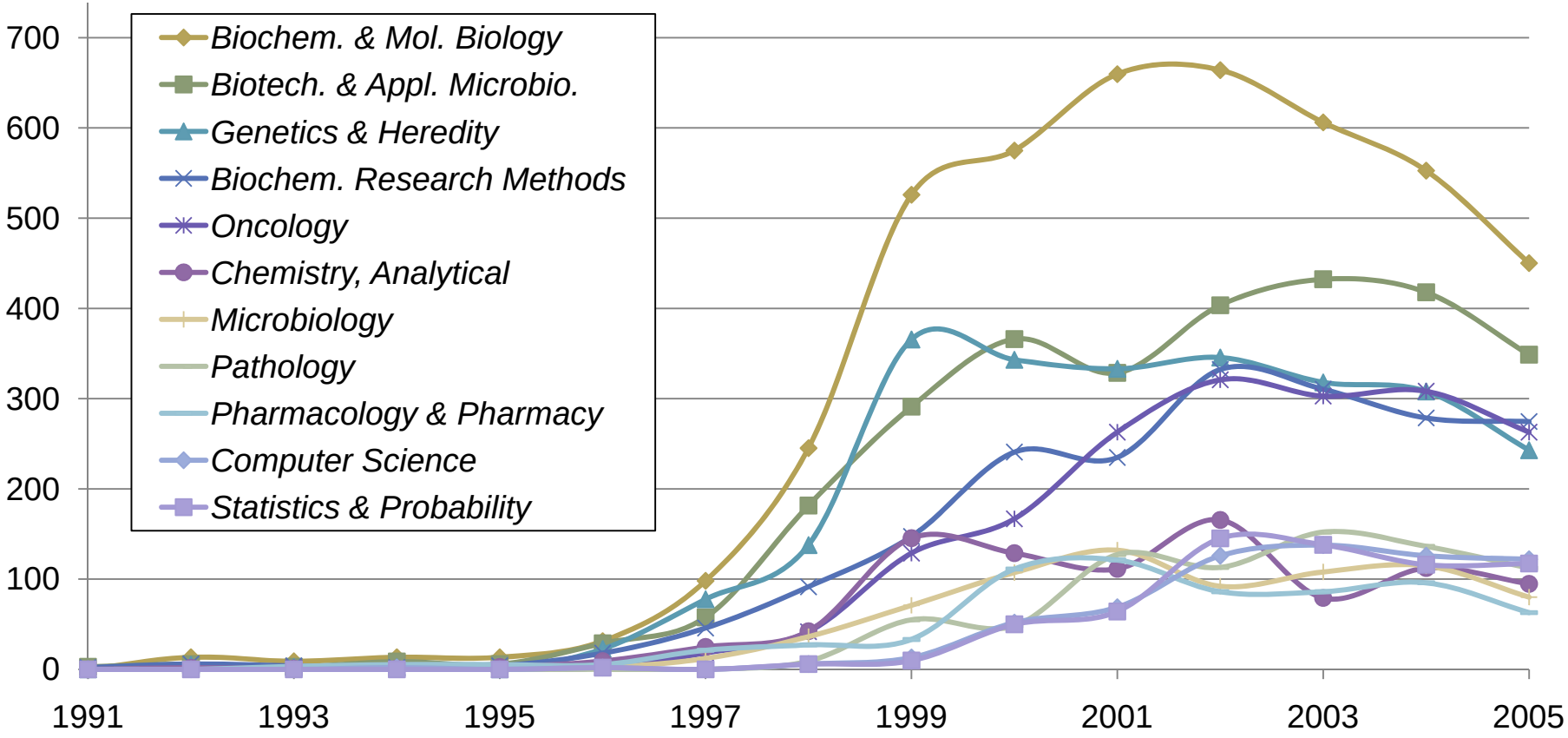
Example: DNA Microarrays

Technology used in molecular biology & medicine for massively parallel measurement of expression of genes under varying biological conditions



DNA Microarray Example (Continued)

Papers on microarrays published in different subfields over time

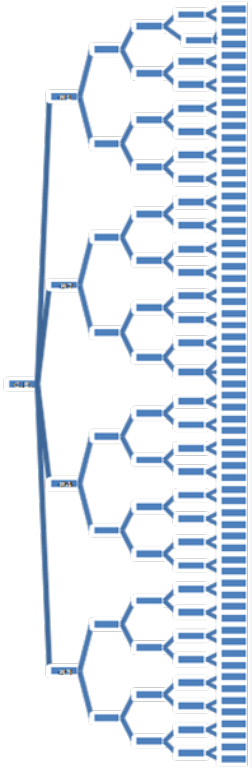


Multiple fields combined to "co-emerge" technical capability

FUSE "Indicator Framework" is a …

Structured set of key aspects of capability emergence

1. Mediates mapping of *lit. space* signals to humanly understandable indicators, by identifying the supporting or contradicting evidence
2. Forms the basis for comparing results across SMEs, performers
3. Evolves iteratively throughout Program, with inputs from performers & stakeholders



Initial FUSE Questions: Is there…

1. Confirmed evidence of a novel capability or development trigger?
2. Evidence of capability maturation and impact?
3. Evidence of the application of a capability or combination of capabilities?

Evidence found in, for example…

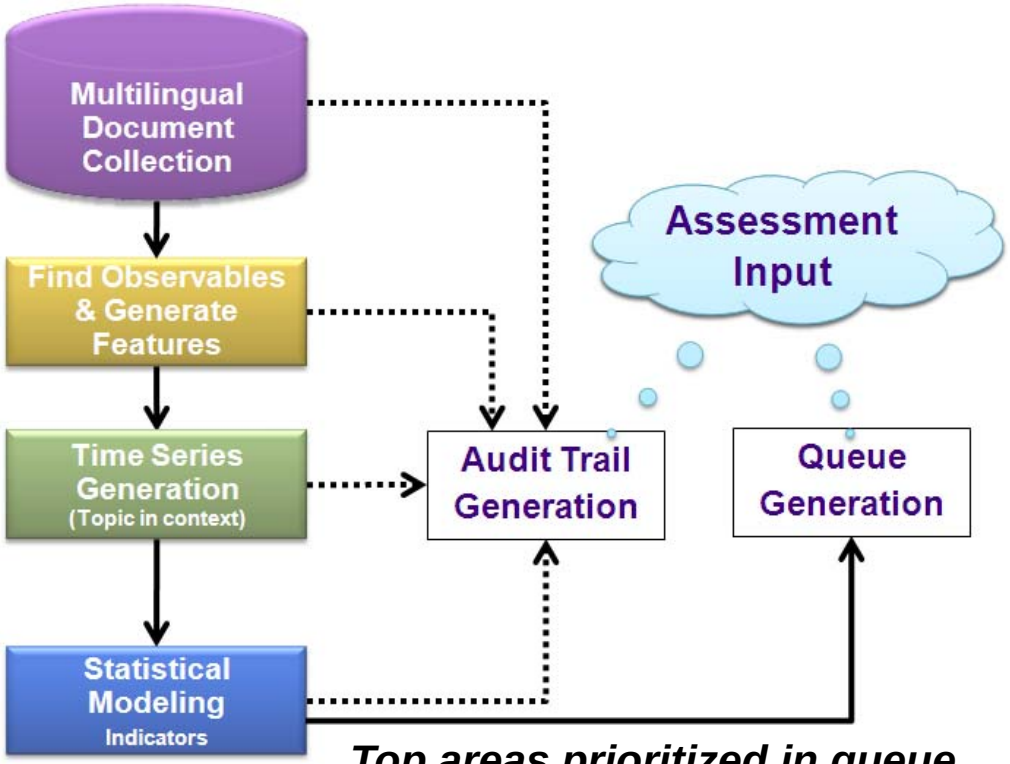
1. Temporal patterns detected in features or combinations of features related to motivation, method/procedures, equipment/data, results, etc.
2. Significant events related to forum (e.g., new journal, special issue, workshop, conference), key paper or patent, funding initiative, etc.
3. Social dynamics in collaboration, knowledge transfer, etc.
4. Sentiment in reference to citations, concepts, groups, etc.
5. Funding and infrastructure investment and availability
6. Activity in reliable places, e.g., patents, or peer-reviewed journals

FUSE Hypothesis

Features exist within the full-text *lit. space* that can be connected to reliably identify capability emergence

INNOVATION SOUGHT

1. Analyze full-texts to produce semantic and entity-based features
2. Calculate trends and patterns from features
3. Apply statistical models to score indicators of technical capability emergence



Top areas prioritized in queue
Analyst can audit evidence trail

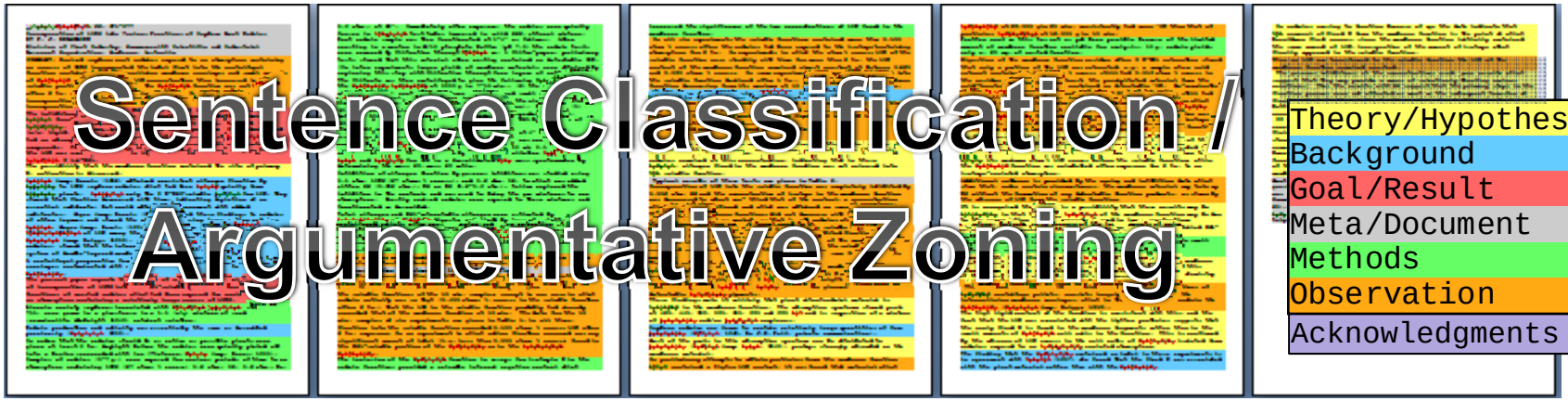
Using Full-Text to Estimate Emergence

Full-Text Document Observables → Cross-Document Features → Time-Series Landmarks & Patterns → Indicator Scores → Prioritized Queues

Observables are what can be found in the full-text of a doc

1. Sentence / Phase / Section Classification
2. Who, what, where, when, and how

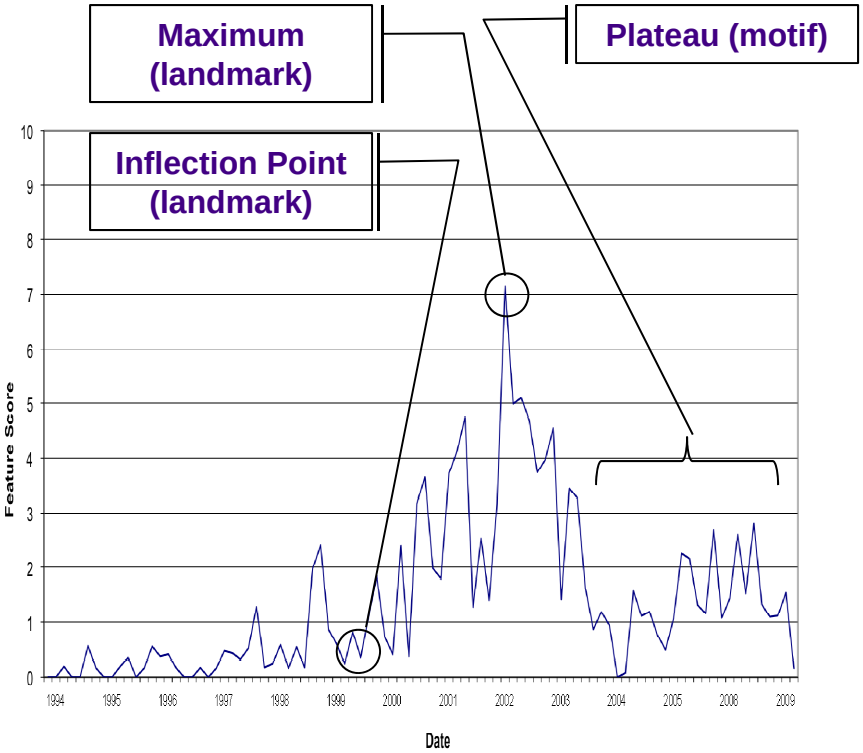
   Named-entities, Objects, Parts of speech, Geographic locations, Sentiment, Dates, Events, Relationship, etc.



Using Full-Text to Estimate Emergence

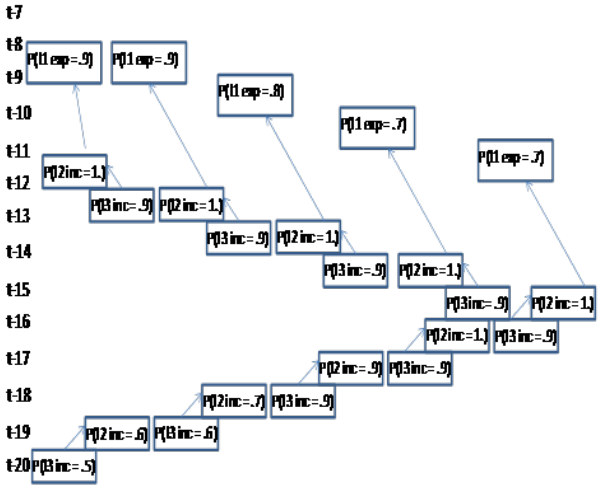
Features are discriminating concepts, named entities, etc. that are produced from correlated observables across sets of related documents

- Landmark identification and other time-series summarization (uni- & multivariate) techniques reduce time-series dimensionality
- Multiple time-series landmarks may be associated with an event or trend in the Indicator Framework
- Pattern library with indicative landmark sequences facilitates the systematic discovery of Indicator evidence



Using Full-Text to Estimate Emergence

Full-Text Document Observables → Cross-Document Features → Time-Series Landmarks & Patterns → Indicator Scores → Prioritized Queues



[Time = t_1] Prioritized Queue of TAs
1. ----------
2. ----------
3. ----------
:
:

[Time = t_2] Prioritized Queue of TAs
1. ----------
2. ----------
3. ----------
:
:

[Time = t_3] Prioritized Queue of TAs
1. ----------
2. ----------
3. ----------
:
:

Time-dependent process-based models

- Build from landmarks and patterns to score Indicators
- Estimate conditional probability (i.e., causality)

TAs prioritized for analyst attention using Indicator scores

Overview of Program Phases

Phase 1 – Full-text to Indicators (18 months):

- Demonstrate that the full-text literature can support robust indicators of capability emergence, valid across multiple TAs
- TAs are defined by case studies with associated literature sets
- Explore impact of variable signal to noise ratios

Phase 2 – TA Emergence (30 months):

- Identify and prioritize the TAs with emergent capabilities in a massive literature set, across multiple languages
- Explore impact of clutter

Phase 3 – Co-TA Emergence (12 months):

- Identify and prioritize emergence supported by multiple TAs from a massive literature set, across multiple languages
- Demonstrate robust model performance in noisy/cluttered environments
- Introduce informal full-text data sources (TBD)

Case Studies Are Basis of Evaluation

- Provided by IARPA, built using SME judgments and existing S&T literature analysis tools
- Contain wide variety of landmarks and patterns
- Form the basis for estimation of “ground truth”
  - Capability emergence properties based on Indicator Framework
  - Indicators, landmarks, and supporting evidence from *lit. space*
- Provide positive and negative examples
- Drawn from period from 1980 to present
- Improved as Indicator Framework evolves

Multi-Metric Evaluation for Queue

Metric	Function
Precision	High Priority for Review (HPR) status evaluated for top *N* of each performer queue over time based on expert judgment *(protocol published prior to evaluation)*
Recall	Evaluation of queue placement of control group (positive and negative case studies or low-effort proxy) over time
Variance	Calculate variance over time among pooled performer queues to characterize system performance and improve future iterations
Domain Diversity	Compute performance across domains to ensure domain-independent performance
Model Divergence	Computes the deviation of system output from case study findings over a range of indicative landmark characteristics at a series of time steps
Feature Impact	Analysis of how sensitive capability emergence indicators are to the inclusion or ablation of full-text derived features

Signal to Noise Response Evaluation

- **Calculate capability emergence properties as system is perturbed**
  - Vary signal to noise ratio
  - Evaluate impact on queue
- **Potential sources of signal and noise variation**
  - *Signal:* Vetted full-text data collections
  - *Noise:* Spurious closely related sets of full-text data (replace/add), errorful full-text data, …
  - *Clutter:* presence of non-emergent topic areas, …
- **Why:** Holistic measure to explore, evaluate and demonstrate program robustness when used on real data sets

Secondary Metrics / Overall Goals

Secondary Metric	Function
Computation Time	Estimation and calculation of computational requirements and efficiency of system output at full-performance scale
Multilingual Evaluation	Evaluation of Human Language Technology algorithms' ability to ingest documents in an alternate language and compute comparable features to those in English
Internal Metrics	Quantitative measurement of the improvement in feature generation, time-series analysis, etc. (performer proposed)
Evaluation Research	Development and utilization of new methods (as needed)

Characterize 100% of the technical areas (TAs) within <24 hours of computation for an update of 30 days worth of publications / filings

- ✓ >80% of top *N* in prioritized queue judged as "High Priority for Review"
- ✓ <20% judged to be irrelevant/poor supportive evidence
- ✓ Low variance in top *N* of prioritized queue in noisy environment
- ✓ Computed from multilingual sources (English and TBD)

Goal

Overall Performance Targets

Phase	1 (18 mo)	2a (15 mo)	2b (15 mo)	3 (12 mo)
Study Domain	Case Studies	Case Studies + Many Thousands of TAs		
Primary Problem	Technical Area Emergence, Cohesive Technical Areas (TAs)			Co-emergence, Divergent TAs
Primary Metric	Multi-Metric Evaluation of Queue *(at each time step)*			
End of Phase	• Analyze impact of full-text • Evaluate metric efficacy	>60% HPR	>80% HPR	>70% HPR
		HPR = High Priority for Review (worth looking at…)		

Multi-Metric Evaluation and Signal to Noise Response Targets TBD

What's In and What's Out

- **In Scope:**
  - Semantic and entity-based feature generation exploiting full-text articles
  - Novel time-series analytics mapped to indicators
  - Development and validation of capability emergence theories
  - Computational models of capability emergence
  - Dynamic topic model development
- **Out of Scope:**
  - Heavy human-language tech innovation investment
  - Human user interfaces
  - Hardware specific solutions
  - Domain specific solutions

Government Roles

- **Government Furnished Information (GFI):**
  - Acquire data (phased acquisition of full-text)
  - Prepare data using existing technologies
  - Determine the standard Technical Areas for evaluation
- **Government Furnished Equipment (GFE):**
  - Centralized server accessed by VPN
  - Storage and computational resources, details TBD
- **Testing and Evaluation:**
  - Measure performer progress
  - Guide development of indicator framework
  - Facilitate evolving consensus for best internal and interim performance measures
  - Evaluate successful English-language and multilingual performance
  - Overall prototype system integration

Eligibility Information

- Other Government Agencies, Federally Funded Research and Development Centers (FFRDCs), University Affiliated Research Centers (UARCs), and any other similar type of organization that has a special relationship with the Government, that gives them access to privileged and/or proprietary information or access to Government equipment or real property, are not eligible to submit proposals under this BAA or participate as team members under proposals submitted by eligible entities.
- Non-US organizations and individuals may be able to participate.
  - Must comply with Non-Disclosure Agreements, Security Regulations, Export Control Laws, etc., as appropriate
  - Specific guidance for non-US participation will be provided in the BAA

Proposal Guidance

- Your proposal should include a full discussion of the technical approach that will be used to meet the program goals.
- Programmatic issues that should be addressed in the proposal:
  - Your team's current technical capabilities
  - Key resources needed (not currently available to your team), to include capital equipment and special expertise (teaming will likely play an essential role in providing special expertise). The risk in acquiring these key resources, and mitigation strategies, should be indicated as well.
  - A teaming plan along with the roles and responsibilities of each member of the research team
  - End of phase and some intermediate milestones are set, but it is expected that other intermediate milestones that are on the critical path of the proposed approach will be offered.
  - A schedule of all milestones including a clearly charted description of the various risk mitigation strategies that will be undertaken to achieve program goals

Proposal Evaluation Criteria

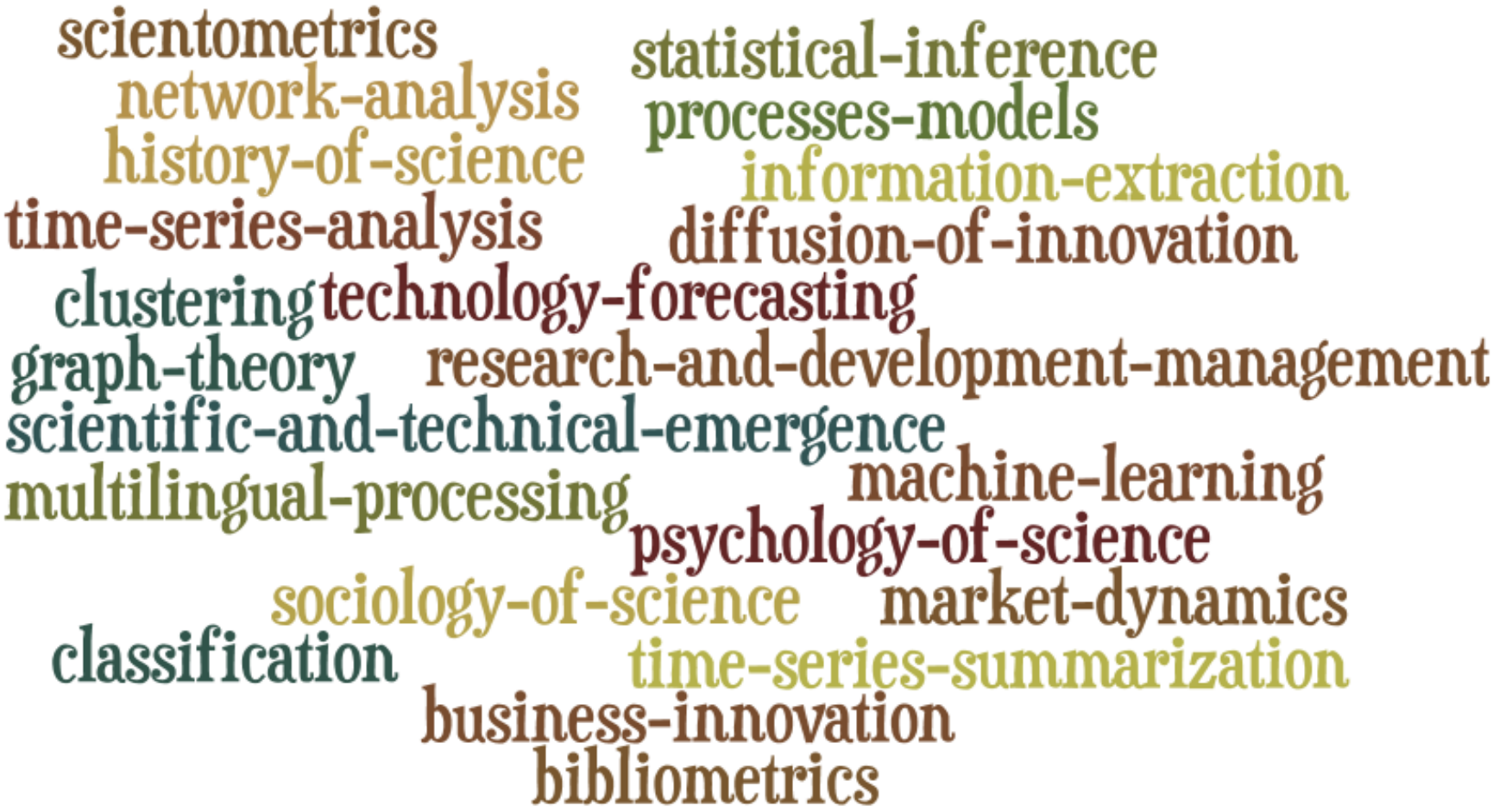
1. Overall Scientific and Technical Merit
2. Effectiveness of Proposed Work Plan
3. Relevance to IARPA Mission and FUSE
4. Relevant Experience and Expertise
5. Cost Realism

Evaluation criteria will appear in the BAA

Teaming

- Because of the many challenges presented by this program, both depth and diversity will benefit your team
  - Throughput : consider all that you will need to do, all the ideas you will need to test
    - Make sure you have enough people and expertise to do the job
    - Make sure you have sufficient resources to follow the critical path while still exploring alternatives
  - Completeness: teams should not lack any capability necessary for success, e.g. should not rely on enabling technology to be developed elsewhere
  - Tightly-knit teams
    - Clear, strong, management, single point of contact
    - No loose confederations
    - Each team member should be contributing significantly to the program goals. Explain why each member is important. If you didn't have them, what wouldn't get done?
    - No teaming for teaming's sake
- Remember, you may be very accomplished, but can you do it all?

Team Composition



Combination of technical challenges will require diverse expertise

Additional Information

- Email dni-iarpa-baa-10-06@ugov.gov with additional questions
- FUSE BAA will be posted on the FedBizOpps website (www.fedbizopps.gov)
- Q&As will appear after the BAA