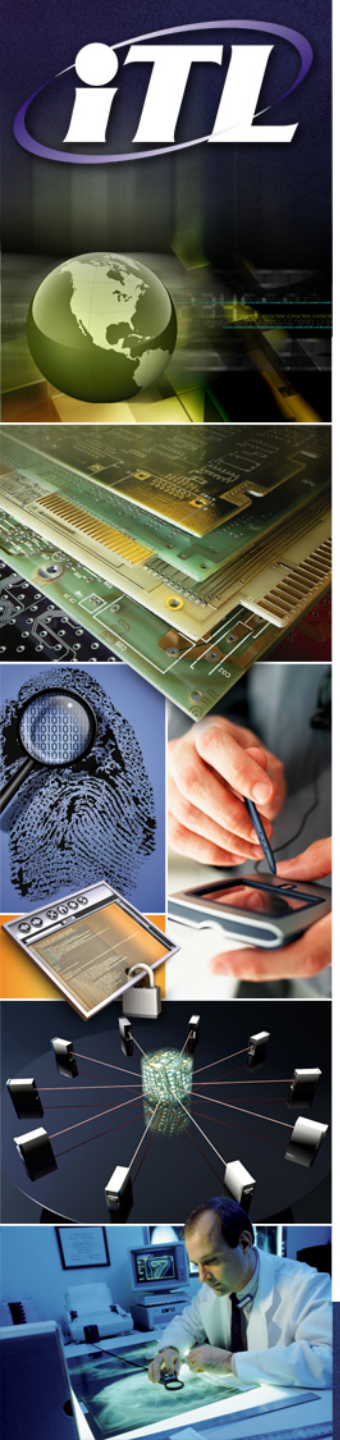


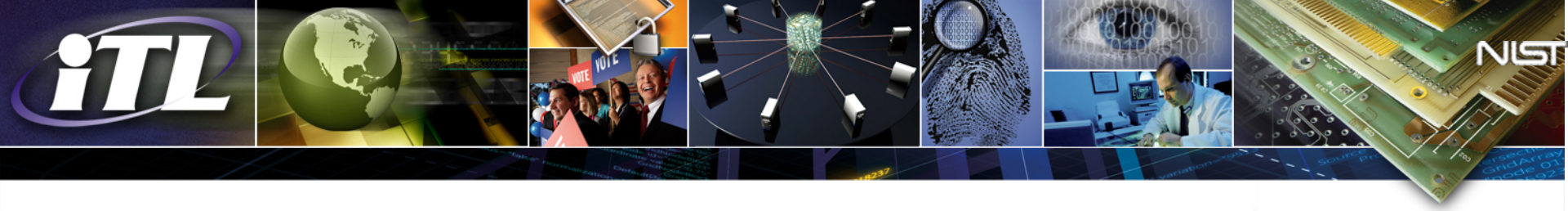
National Institute of Standards & Technology

DATA SCIENCE PROGRAM

Data Science Symposium

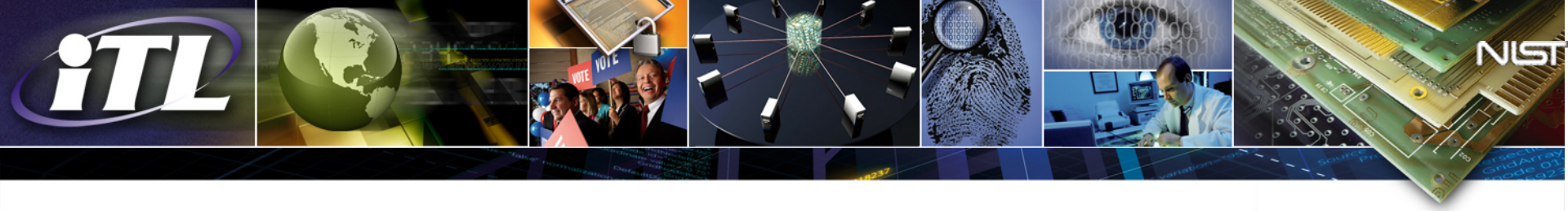
March 4th and 5th, 2014

Ashit Talukder
Chief, Information Access Division, ITL, NIST
Email: Ashit.Talukder@nist.gov



Data Science Measurement Challenge in the Information Age

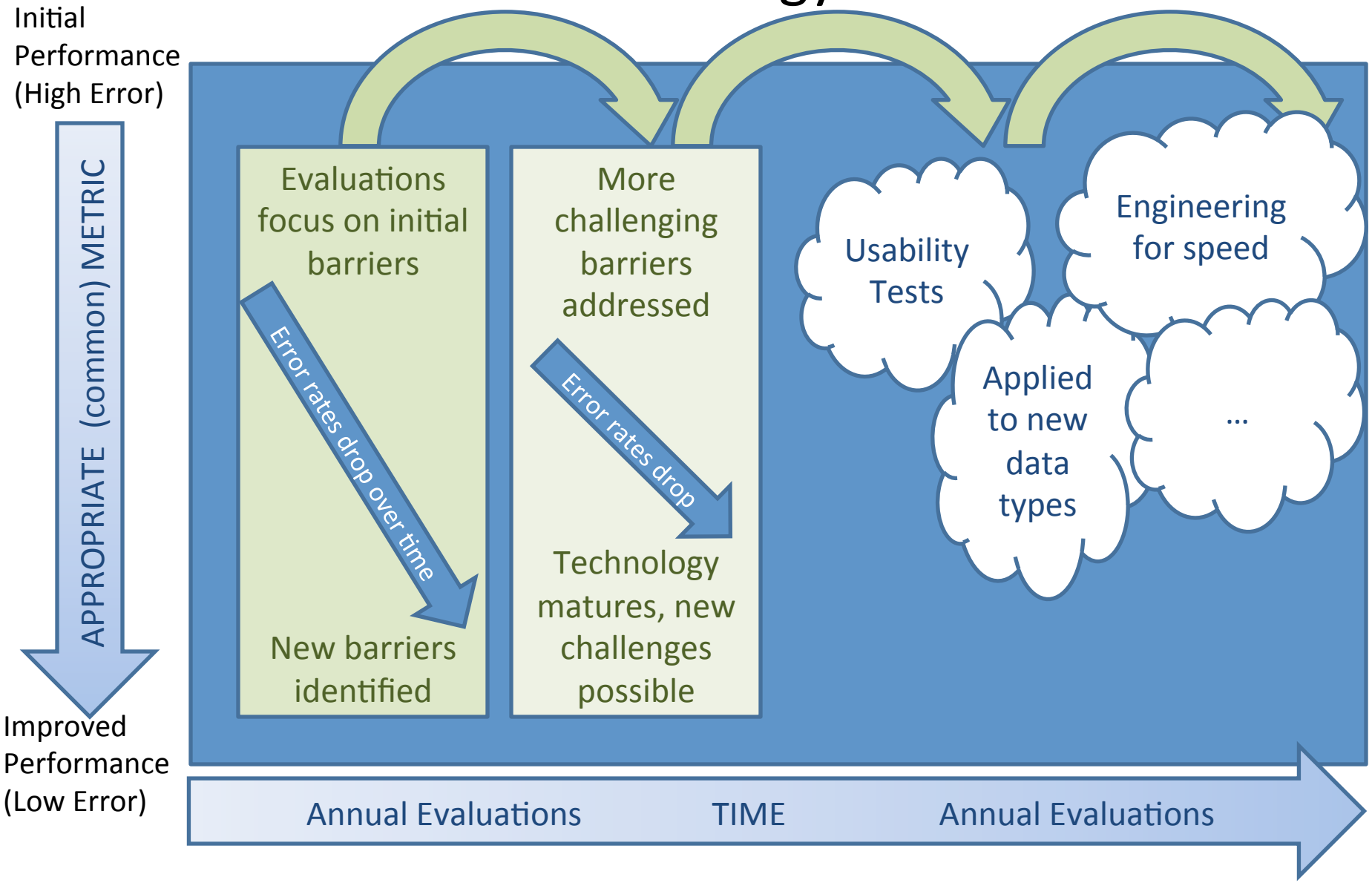
- Data-to-Knowledge Pathway (and associated workflows) is key component of decision-making and discovery in the information age
- Technologies to extract knowledge from data increasingly used in all segments of business, science, healthcare, and daily living
- Data Science enables better decisions, but approaches often lack the necessary measures for understanding:
  - the quality and context of the analyzed data
  - the rigor of the analytic process and tools employed
  - the impact of the human in the analytic process
  - the strength of the conclusions derived, questions answered, hypotheses tested, and discoveries made
- Since data science is driving crucial decision making, it is critical to understand the approaches, measure the performance of the underlying technologies, and to correctly interpret the final output



NIST Data Science Interests

- White House (OSTP) announced multiagency Big Data R&D Initiative
- NIST has strong history of fostering progress in analytics and data science for specific data modalities
  - NIST conducts open, community-based evaluations and benchmarking in specific data science domains involving structured and unstructured data by disseminating reference data, hosting challenge problems, and designing metrics
  - Identify gaps in specific data science applications through strong community participation, and overcome barriers to progress by having community solve common problems
    - Text search and information retrieval (TREC, TAC)
    - Speaker ID, Machine Translation,
    - Video Analytics (TRECVID), Image-based biometrics
    - *New Data Science Program*

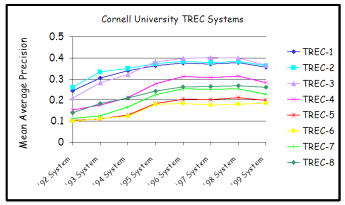
Challenge Problems to Advance Specific Technology



Driving Innovations in Data Science Through Measurements: Impacts

- Reference Data, metrics, and challenge tasks to advance information retrieval from text– TREC conferences since 1992

Improves the state of art



TREC data revitalized research on information retrieval. Having a standard, widely available, and carefully constructed data set laid the groundwork for further innovation in the field. The TREC conference fostered collaboration, innovation, and a measured dose of competition that led to better information retrieval.

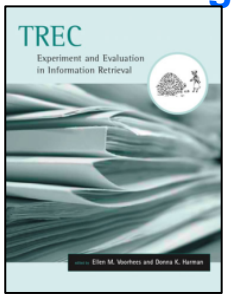
Hal Varian
Google Chief Economist
March 4, 2008

Forms/solidifies a research community

This project [the TREC Legal track] can be expected to identify both cost effective and reliable search and information retrieval methodologies and best practice recommendations, which, if adhered to, certainly would support an argument that the party employing them performed a reasonable ESI search, whether for privilege review or other purposes.

Magistrate Judge Paul Grimm
Victor Stanley v. Creative Pipe

Establishes the research methodology



TREC is an annual benchmarking exercise that has become a de facto standard in Information Retrieval evaluation.

Stephen Robertson
Microsoft
SIGIR 2007

Facilitates technology transfer

TREC has proven to be a valuable forum in which IBM Research has contributed to an improved understanding of search, while at the same time the insights obtained by participating in TREC have helped to improve IBM's products and services.

Alan Marwick, et al.
IBM chapter of the TREC book
2005

Amortizes the costs of infrastructure

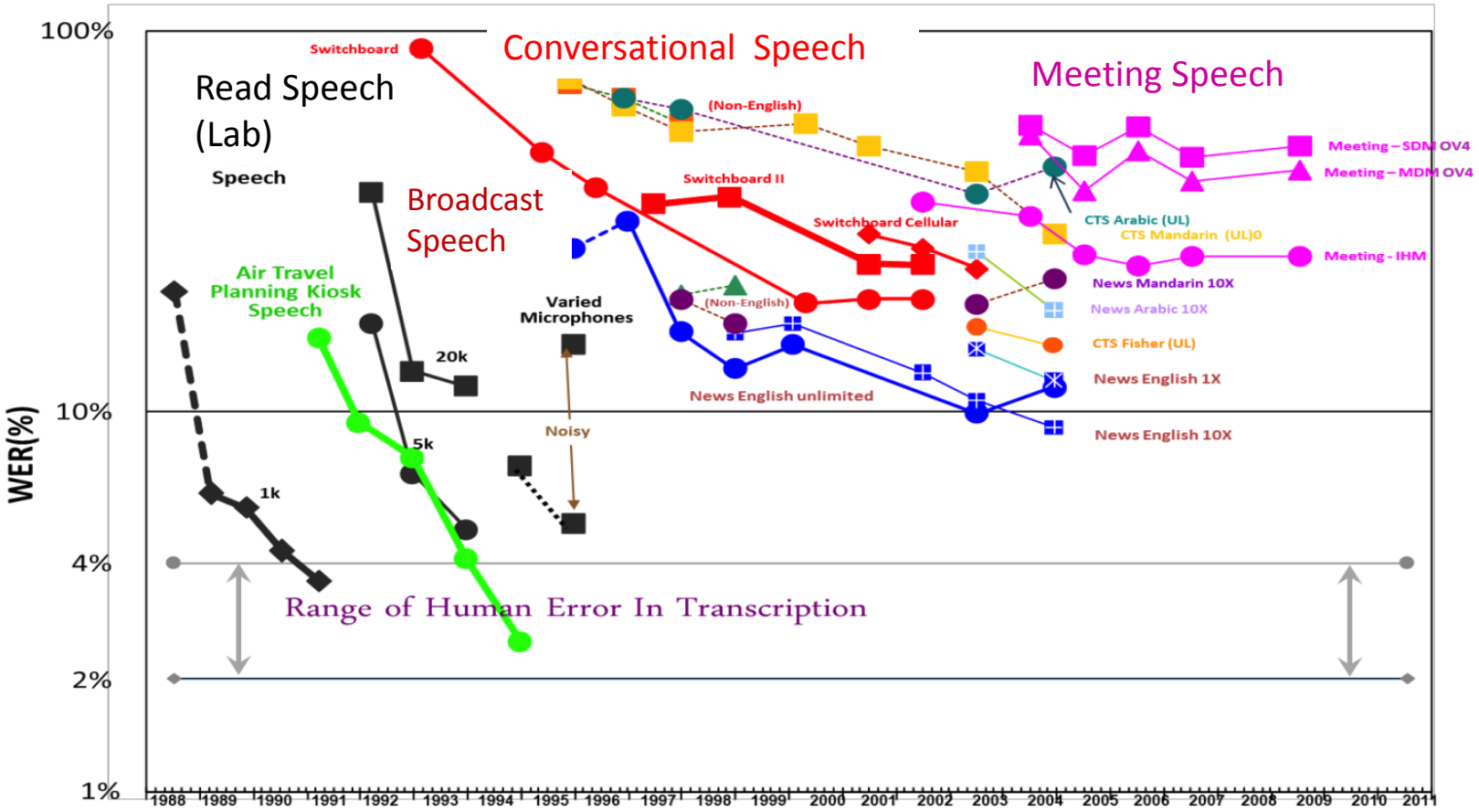
In other words, for every $1 NIST and its partners invested in TREC, at least $3.35 to $5.07 in benefits accrued to IR researchers…These responses suggest that the benefits of TREC to both private and academic organizations go well beyond those quantified by this study's economic benefits.

RTI International
Economic Impact Assessment of NIST's TREC Program
December 2010

Driving Innovations in Data Science Through Measurements: Impacts

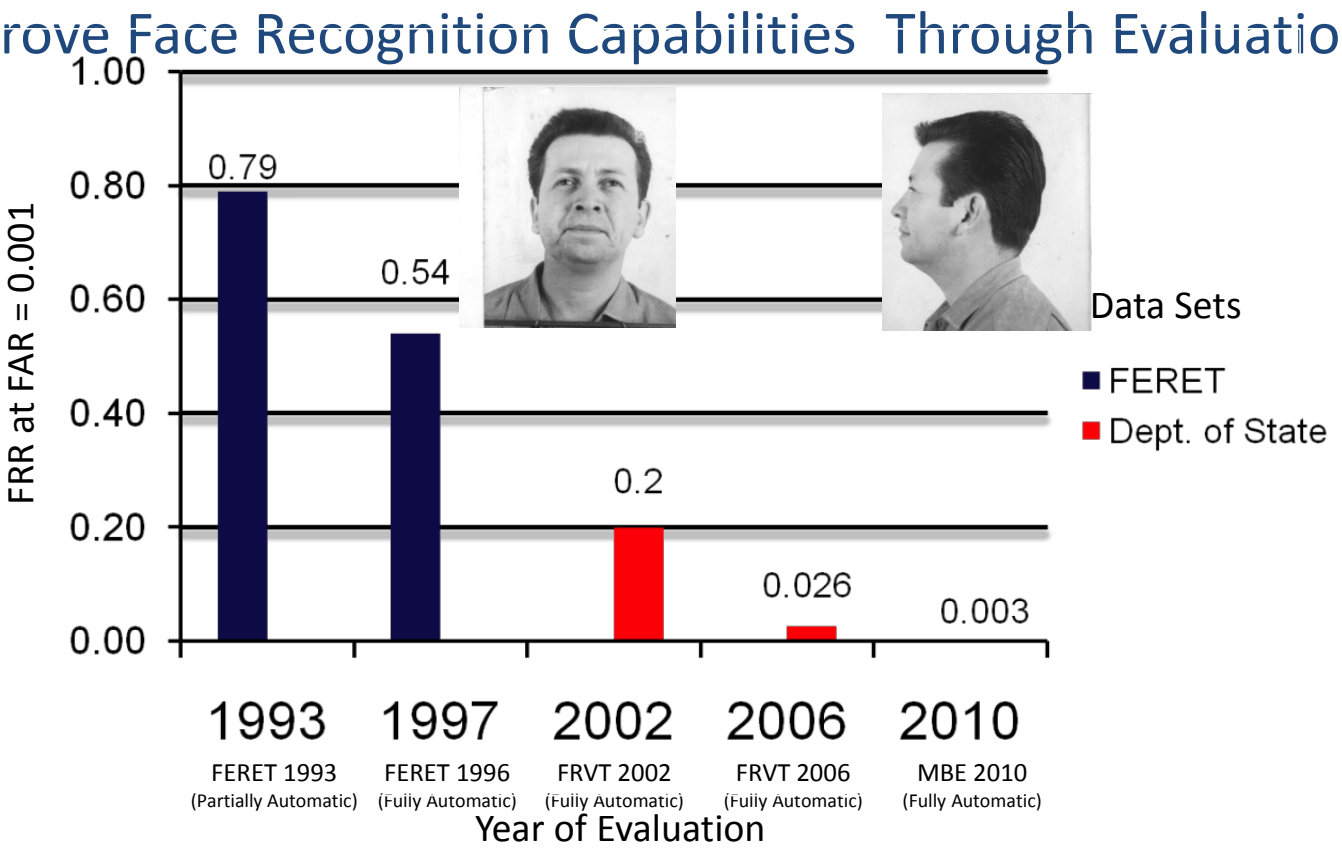
- Catalyze advances in speech technologies (speech recognition, translation, and voice biometrics) from infancy to mature prototypes through new metrics and challenge tasks

Progressive Advancement of Speech Recognition



Driving Innovations in Data Science Through Measurements: Impacts

- Comprehensive image-based NIST biometrics evaluations with strong stakeholder participation advanced biometrics field and led to biometrics standards, interoperability and mature biometrics products since 2001



NIST Data Science Program Vision and Goals

Vision: A interdisciplinary multi-sector Data Science program focused on enabling the understanding and evaluation of data science approaches, identifying gaps, and fostering advancements through benchmarking and rigorous measurement science, reference datasets, and evaluations

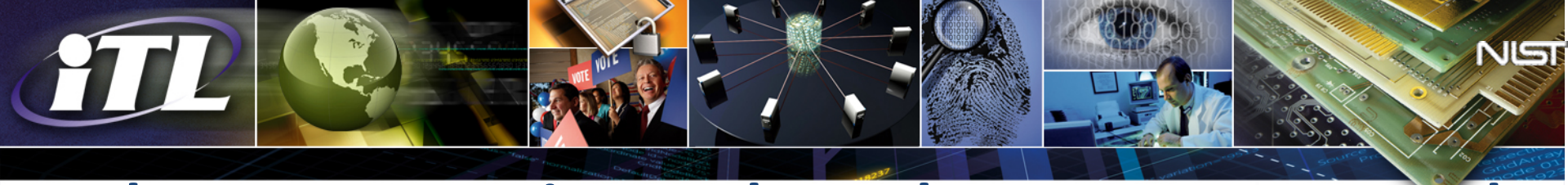
Goals:

- ❑ Work with Data Science community to address technology and measurement barriers and characterize uncertainties critical for data-to-knowledge workflows
- ❑ Advance Data Science capabilities and address R&D to resolve gaps through:
  1) advancement of **rigorous measurement techniques**
  2) development of **reference frameworks**
  3) development of **open challenge problems** on **use cases addressing compelling classes of technology challenges**
  4) community collaboration (***engaging* stakeholders from ALL sectors**)
- ❑ Help all stakeholders understand the state-of-the-art and facilitate collaboration, interchange, direct comparison of approaches on common problems and reference data with standard metrics

Data Science Program Expected Impacts

- **End Users** will have an objective resource to understand the technology space and implementation considerations
- **Industry** will have a means for engaging a broad R&D community to understand the state-of-the-art and best practices for measuring and integrating technology
- **Academia, Researchers & System Developers** will have resources and methods for objectively comparing approaches, performing scalable research, and understanding performance from both a component- and system-level perspective

➢ **Accelerated R&D transition, overcome barriers to progress, reduced transition risk, better more reliable analytic technology**



Advance Data Science Through Measurement and Evaluation

Data Science Technologies

Data Science Stakeholders

- Expertise
- Data Science Use Cases
- Data Sharing
- Formal and Informal Collaborations
- Outreach Events
- Evaluations
- Evaluation Workshops

Measurement Science

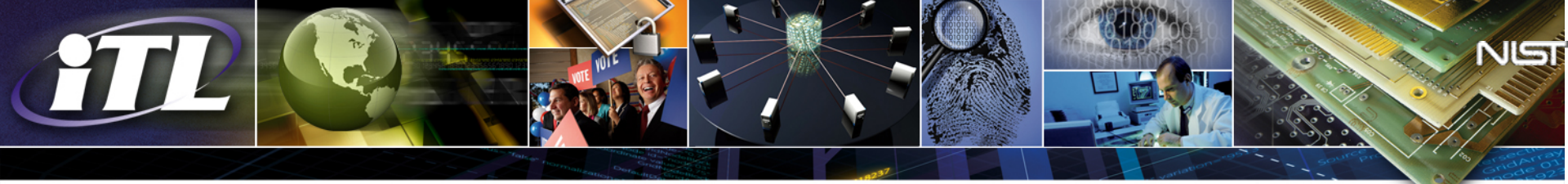
- Metrics
- Methods
- Data and Ground Truth
- Reference Frameworks

Measurements for Data Science
Hardening metrics, Identifying critical gaps, and advancing state-of-the-art

- Novel Metrics
- Massive annotated data collections
- Novel multidisciplinary/ multi-sector communities of interest
- Forward-thinking analytic challenge problems and open evaluations
  - generating critical R&D mass
  - accelerating development of advanced analytic technologies
- Reference evaluation and R&D architectures and algorithms
- Best practices, consensus standards, generally recognized de facto standard performance testing methods and data, and formal standards
- Analytic Technology interoperability and well-understood state-of-the-art
- Advance capabilities in NIST and stakeholder Data Science use-cases

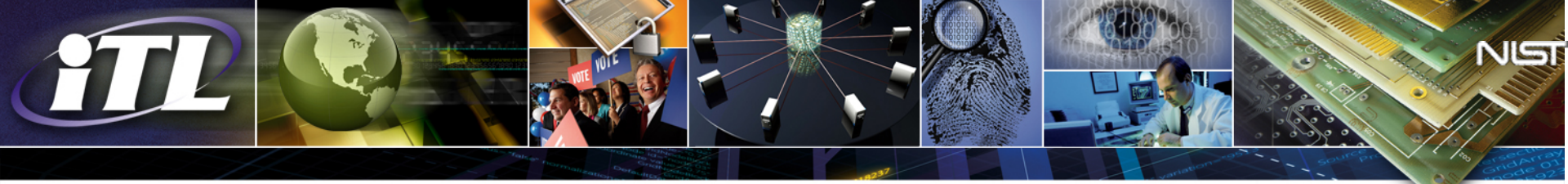
NIST Data Science Program Phased Approach

- **Common Framework** for data science workflows
  - Create common taxonomies and categorization for data analytics
  - Allows stakeholders across multiple domains to speak common language
  - Enable multi-sector experts to identify common solutions & approaches
  - Assist in identification of critical gaps and hurdles
- **Adoption of reference datasets** that will be reusable and applicable across multiple domains
- **Foster cross-sector collaborations** to catalyze advances by identifying solutions in one sector reusable in other domains
- **Multi-year open evaluations** involving communities of experts using a phased approach and consensus process



NIST Data Science Program Phased Approach

- **Common Framework** for data science workflows
  - Create common taxonomies and categorization for data analytics
  - Allows stakeholders across multiple domains to speak common language
  - Enable multi-sector experts to identify common solutions & approaches
  - Assist in identification of critical gaps and hurdles
- **Adoption of reference datasets** that will be reusable and applicable across multiple domains
- **Foster cross-sector collaborations** to catalyze advances by identifying solutions in one sector reusable in other domains
- **Multi-year open evaluations** involving communities of experts using a phased approach and consensus process



Example Common Framework for Data Science Workflows

Data

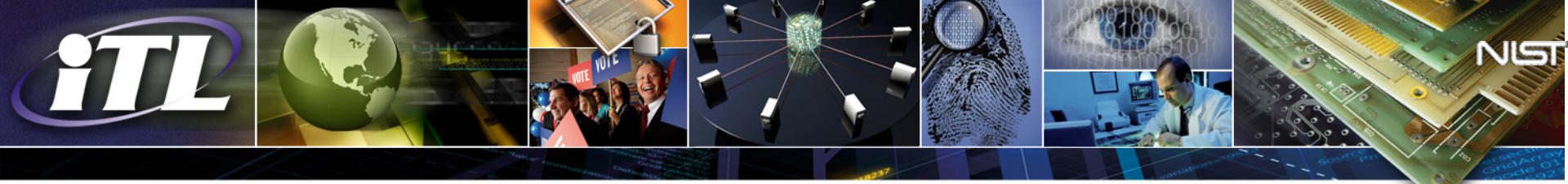
Sources

- Observational
- Transactional
- Model output
- Metadata
- Experimental

…

Representations

- Records
- Sequences
- Graphs
- Spatio-temporal
- Natural language text
- …



Example Common Framework for Data Science Workflows

Analytic Components

Objectives

- Detection
- Modeling
- Transformation
- Inference
- Prediction
- …

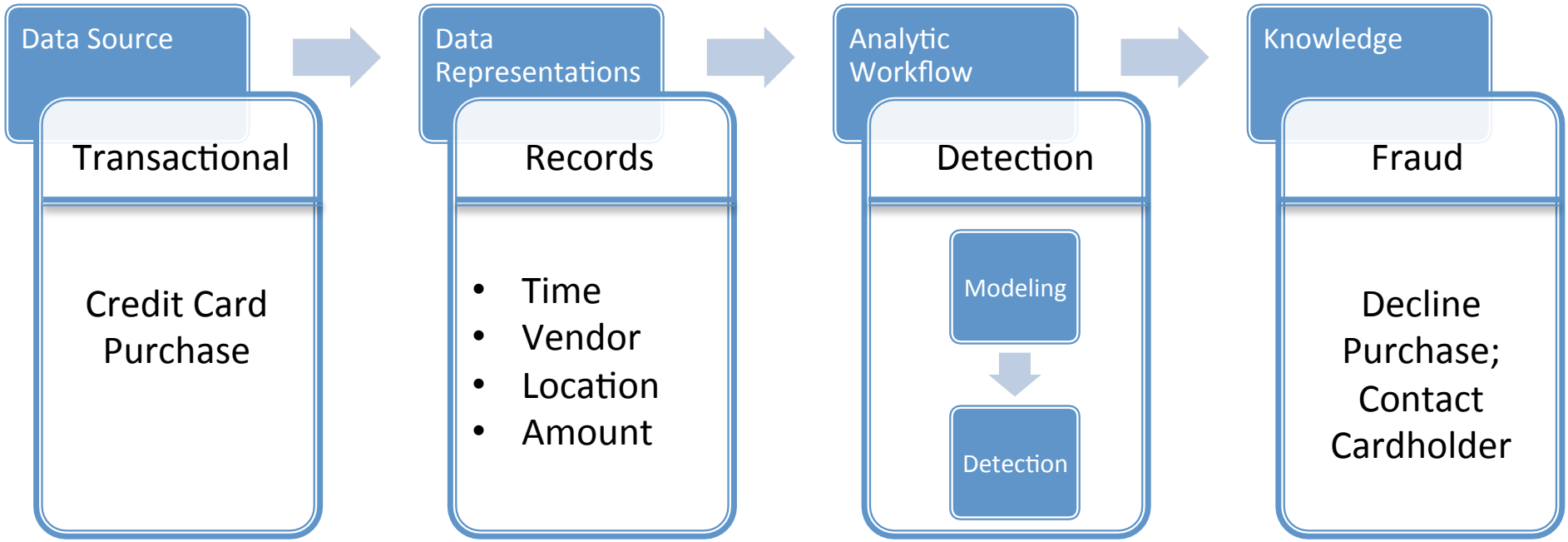
Approaches

- *Regression*
- *Association rule learning*
- *…*

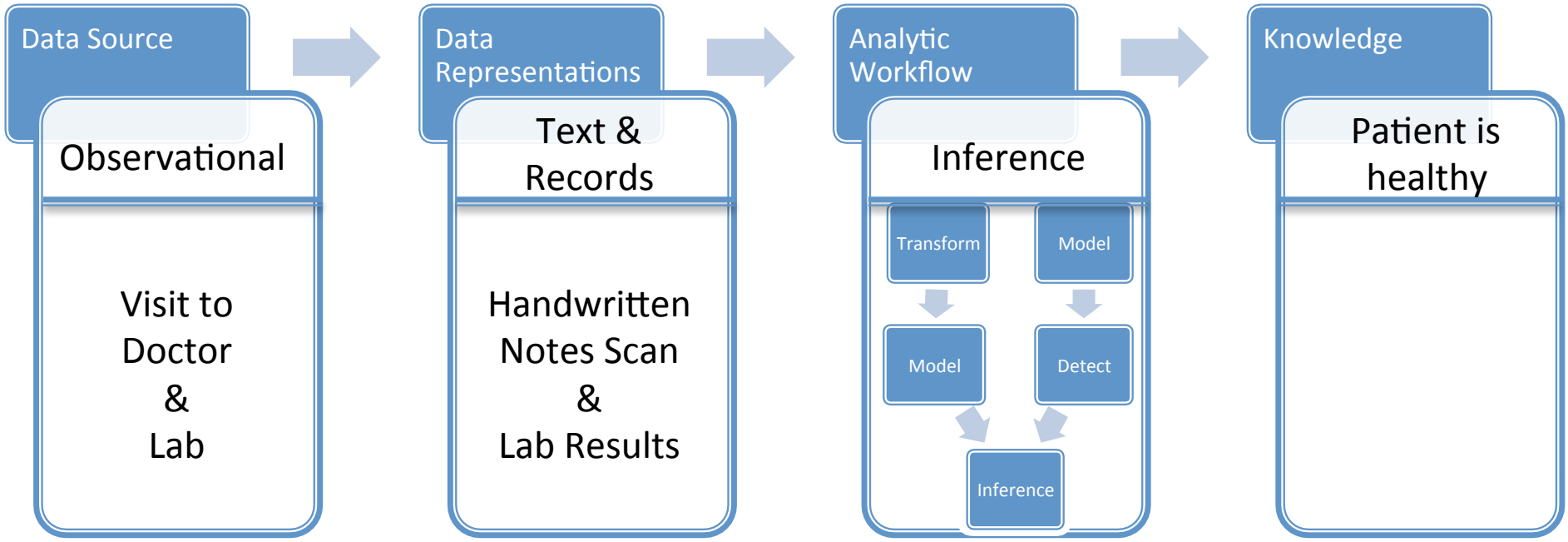
Techniques

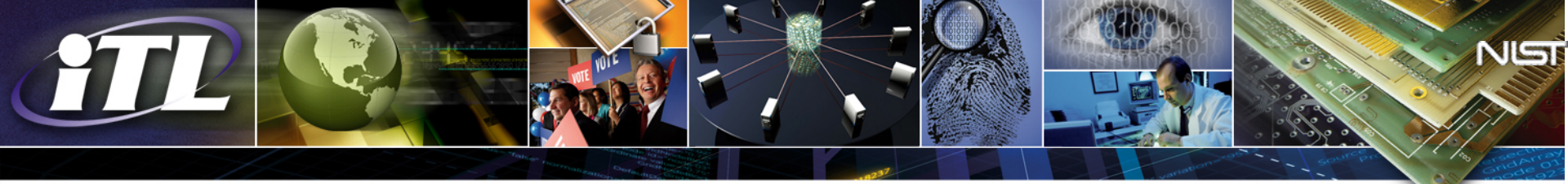
- *Support vector machines*
- *Neural networks*
- *…*

Example Common Framework for Data Science Workflows Illustration



Example Common Framework for Data Science Workflows Illustration



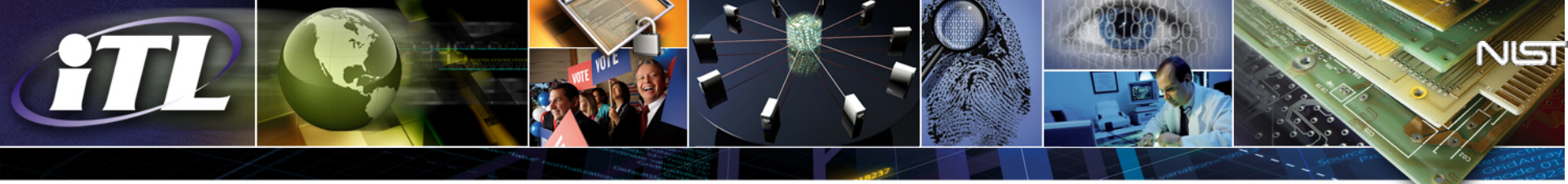


Example Common Framework for Data Science Workflows Illustration

- Operations
- Experiments
- Transactions
- Observations
- Model outputs
- Metadata, Ontology
- …

- *Machine Generated*
- *Human Generated*

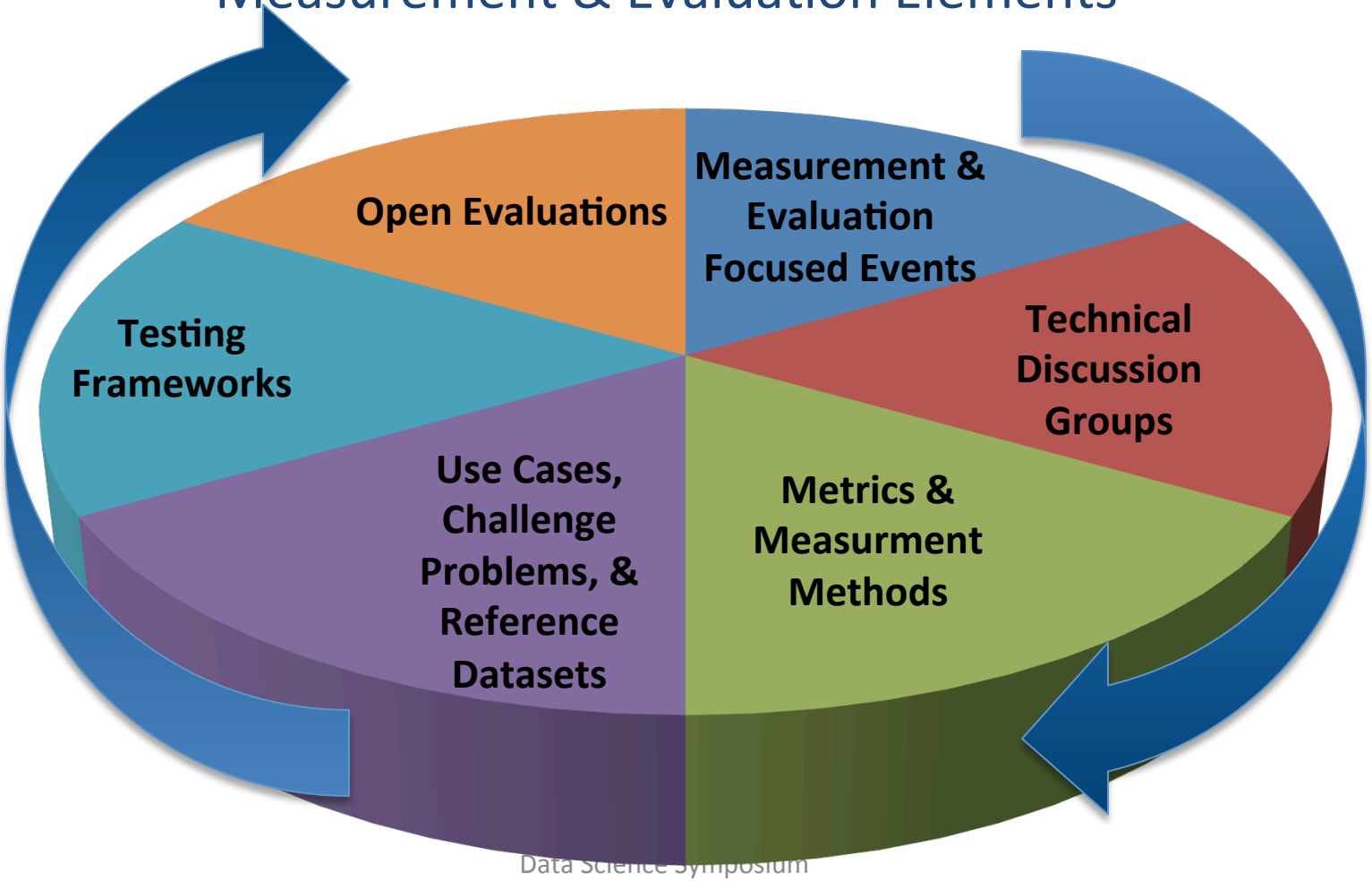
	Detection	Grouping	Modelling	Prediction	Inference	Transform
Records	Medical Diagnosis Support		Medical Diagnosis Support		Medical Diagnosis Support	
Sequences						
Graphs						
Spatial/ Image						
Temporal						
Spatio-temporal / Video						
Text			Medical Diagnosis Support		Medical Diagnosis Support	Medical Diagnosis Support

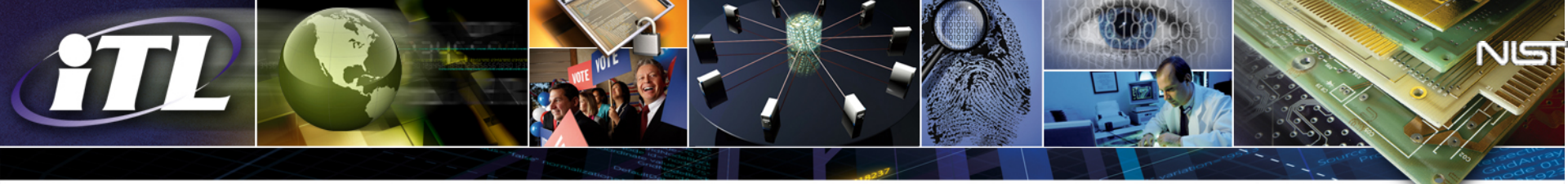


NIST Data Science Program Phased Approach

- **Common Framework** for data science workflows
  - Create common taxonomies and categorization for data analytics
  - Allows stakeholders across multiple domains to speak common language
  - Enable multi-sector experts to identify common solutions & approaches
  - Assist in identification of critical gaps and hurdles
- **Adoption of reference datasets** that will be reusable and applicable across multiple domains
- **Foster cross-sector collaborations** to catalyze advances by identifying solutions in one sector reusable in other domains
- **Multi-year open evaluations** involving communities of experts using a phased approach and consensus process

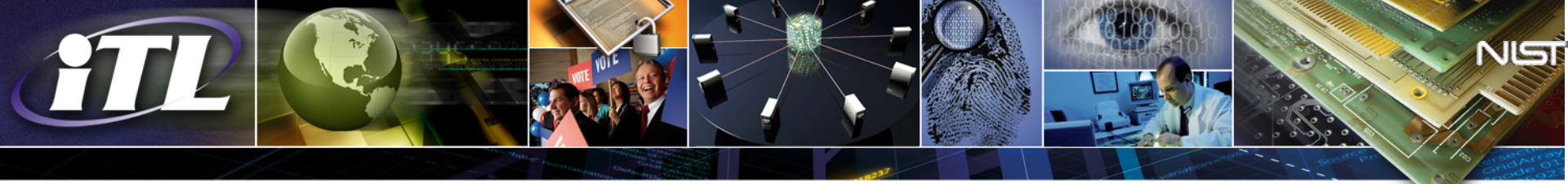
Data Science Program
Measurement & Evaluation Elements





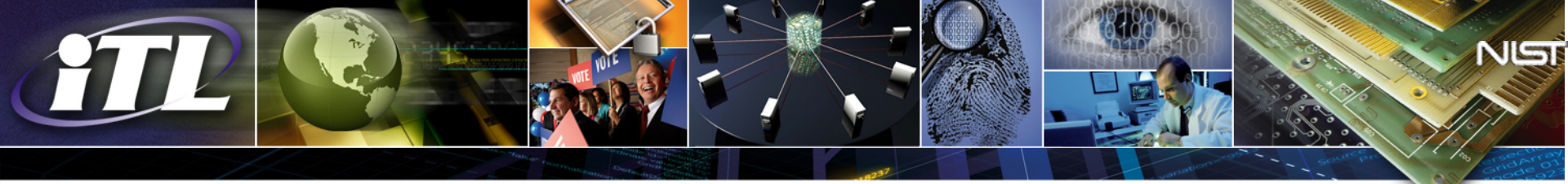
Metrics Development

- Create general, extensible measures of:
  - **Scalability, resource utilization , speed, and accuracy** of analytic components, end-to-end systems with and without humans in the loop
  - **Propagation of error and uncertainty** through the system
- Develop evaluation strategies that explore
  - **Transferable methods and complex data representations** (heterogeneous, streams, etc.)
  - **A variety of analytic objectives** (prediction, clustering, tracking, search, filtering, etc.)
  - **User interfaces and usability**



Use Case and Reference Data Development

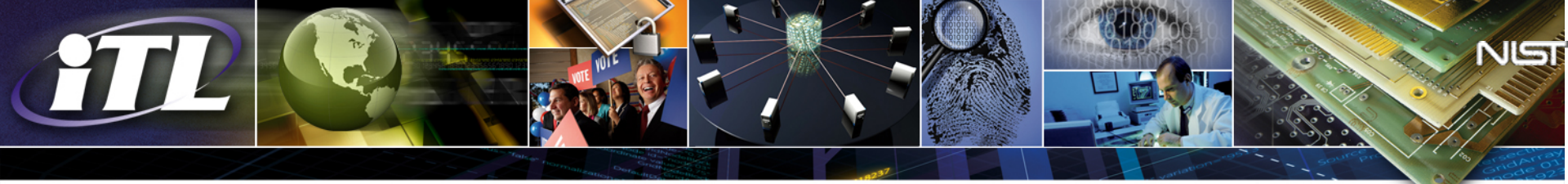
- Utilize common framework for data science workflows to **build consensus across use cases** around abstract data and analytic categories
- **Identify candidate use cases** for challenge problems, metrics development, R&D, and evaluations
  - Staged approach, starting with low-hanging fruit, overcoming barriers, addressing new challenges
- **Make available reference datasets** for use in research and evaluation that will be reusable and applicable across multiple domains



Testing Frameworks to Support Research, Measurement, and Evaluation

Testing framework to:

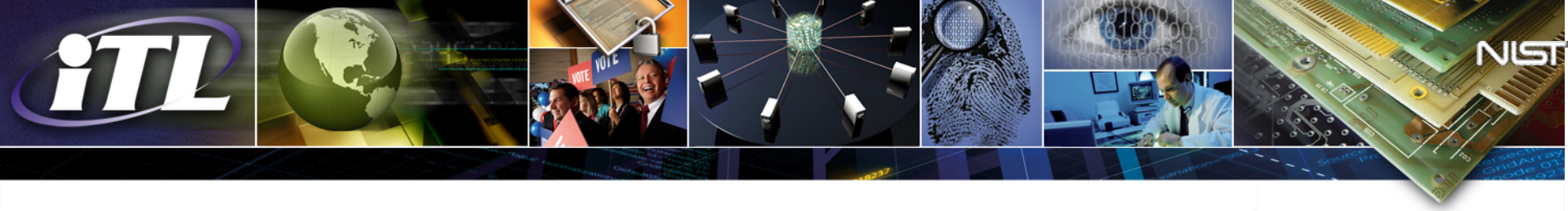
- develop and implement data science evaluations using a **common approach**
- foster **community-contributed** open source tools, libraries, data, computing resources, and experiments
- assist both data science practitioners and application developers **understand performance characteristics** of different implementations



Measurement and Evaluation Focused Events

Data Science Symposium Series will be an annual event, with the following objectives:

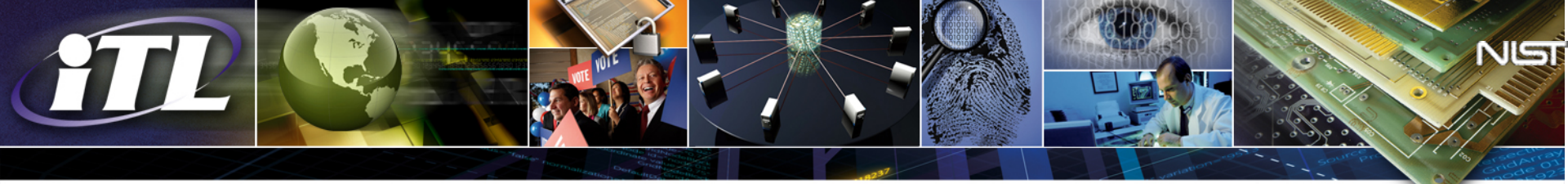
- Convene a **broad community of experts** to discuss the data science landscape, directions, and measurement and collaboration needs
- Foster **collaboration and resource sharing**
- Report on data science **evaluation results**
- Discuss the **state of the art**
- Expose **technical gaps and challenges**
- Plan and prioritize **future challenge evaluations** to focus research



Data Science Evaluation Series

- Open challenge **evaluation series with several analytic challenge tracks** addressing a variety of research challenges
  - Tracks chosen to **address common problems** across domains
  - Coordination of tracks to have components in common, allowing **cross track research**
  - **Representative data** to be tested against, and **objective metrics**
  - Evaluations to include **human interaction** with data as applicable

- Evaluations **results and research reports** will be presented during an annual symposium series by both the track leaders and participants.

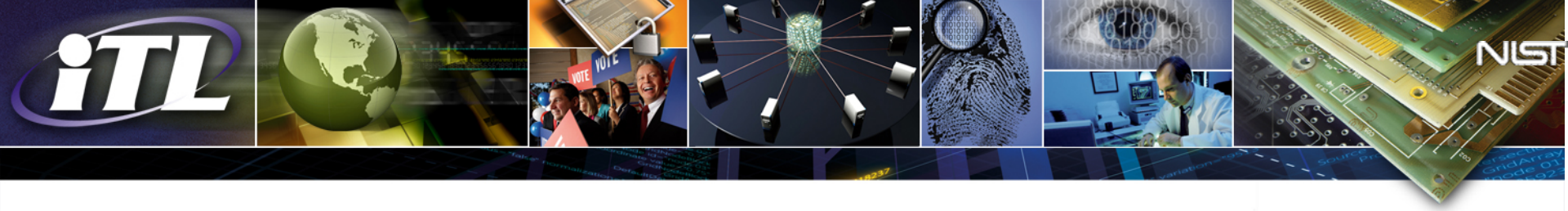


Data Science Evaluation Series Outcomes

- **Establish evaluation framework**, enabling research coordination for mutual benefit by providing common
  - Objectives
  - Metrics
  - Datasets
  - Measurement Methods
- **Measure progress** in the state of the art
- **Identify technical hurdles** and engage community to overcome hurdles
- **Harden metrics and methods** of measurement

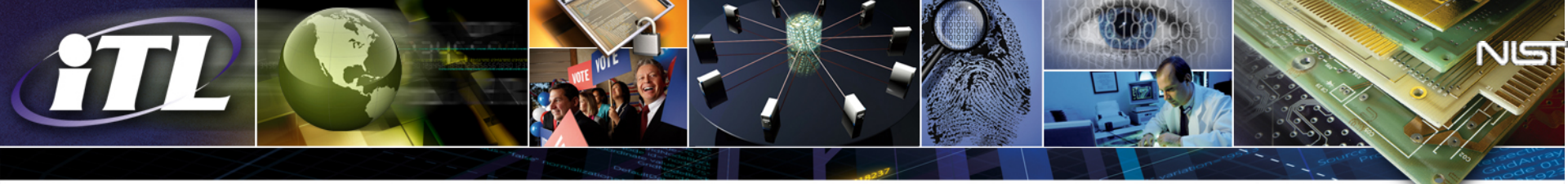
Data Science Evaluation Series
Proposed Timeline & Outputs

Form Discussion Groups	Run Pilot Evaluation (1-3 Tracks)	Coordinate Community Driven Evaluation	Testing Framework Available
Now	2014	2015/ 2016	2016/ 2017
• Common Framework for Data Sci. Workflows • Community Consensus Plans	• Reference Datasets • Establish Baseline Performance	• Coordinated Research Efforts • Measures of the state of the art	• Resource for Research and Measurement • Common Framework Enabled Approaches



Way Forward

1) Form Data Science **evaluation planning group**
2) Solicit **track proposals** and champions
   - Prioritize based on desiderata, evaluation balance, and risk
3) Solicit initial **expression of intent** for candidate tracks
4) Form **track development** groups and track implementation plan
   - Develop implementation schedule
   - Identify/create track data resources
   - Develop evaluation tools
   - Define track evaluation plan (tasks, data, metrics, methods)
5) Deploy track development **resources and tools**
6) Conduct **dry run evaluations**, testing the pipeline
7) Obtain **commitments to participate**
8) Conduct **formal evaluation**
9) **Report results** at annual symposium



How to Contribute

Join in symposium activities

- Help create the common framework for data science workflows
- Attend the working lunches and contribute to the breakout group discussions
- Share your ideas

Engage in evaluation planning

- Become a member of the evaluation planning group
- Suggest datasets, challenge problems, metrics
- Propose or champion an analytic research track

Participate in evaluations