National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

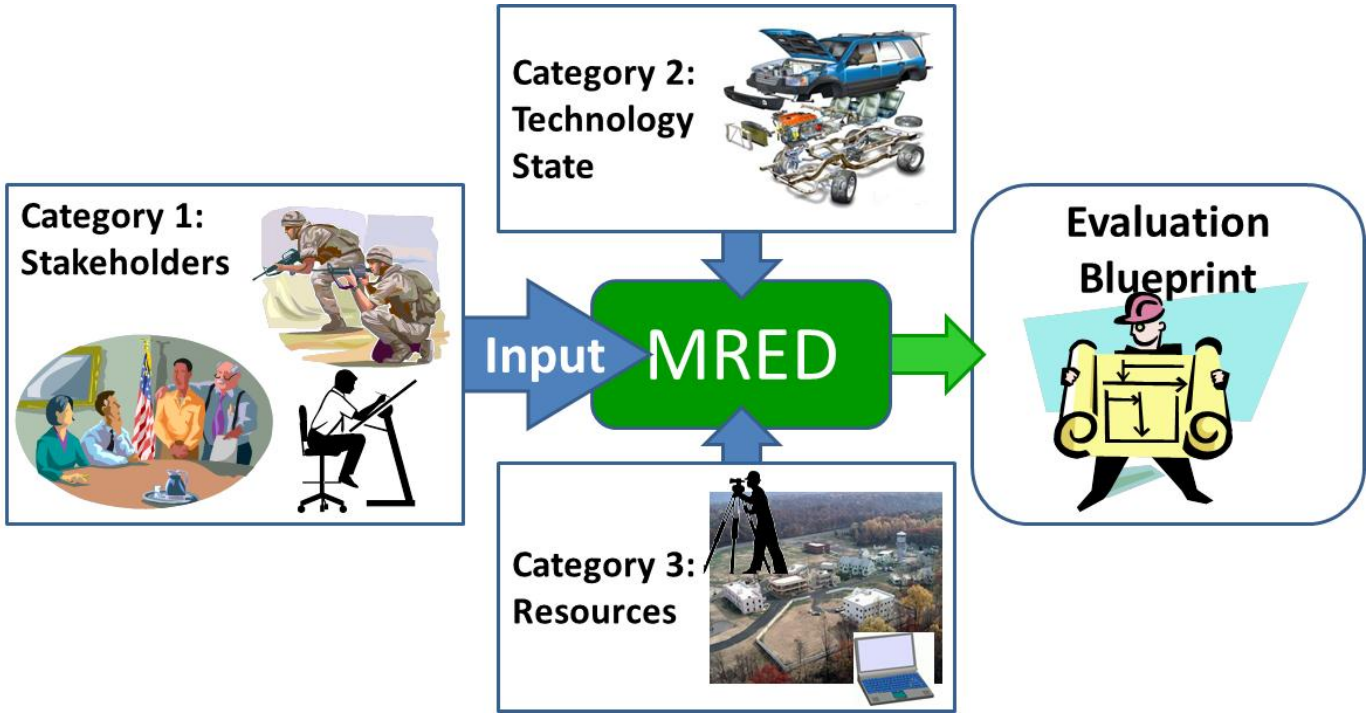
ERGING AND INTELLIGENT SYSTEMS

Brian A. Weiss
Michael L. Curry
Brian A. Weiss
Stephen Balakirsky
Miles Thompson
Jeffrey Schleher
Dan M. Davis

TECHNOLOGY for RAPID ACQUISITION AND TEST

Multi-Relationship Evaluation Design (MRED)

Formalizing Evaluation Design Input and Output Blueprint Elements for Testing Developing Intelligent Systems



Brian A. Weiss (NIST) and Linda Schmidt (University of Maryland)

Intelligent Systems Division, National Institute of Standards and Technology
100 Bureau Drive, Mailstop 8230, Gaithersburg, MD USA 20899

EMAIL: brian.weiss@nist.gov
PHONE: 301-975-4373

Today's Discussion

OUTLINE

- What is MRED and why it's needed
- State of MRED
- Established Inputs and Outputs
- Input Category – Technology State Factors
- Output Element – Technology Test Levels
- Input Influence on Output
- Conclusions and Future Work

NIST
National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

MRED

WHAT IS IT?

- OBJECTIVE: To Automatically Generate Evaluation Test Plans based upon Multiple Inputs.
- WHAT IT DOES:
  - Take information from three input categories and output one or more evaluation blueprints complete with their own specific test plan elements.
  - Characterize the relationships among inputs and the influences inputs have on outputs.
- STATUS:
  - Overall model has been developed
  - Input categories have been defined
  - Output evaluation elements have been specified
  - Relationships among inputs and between inputs and outputs are being explored

MRED

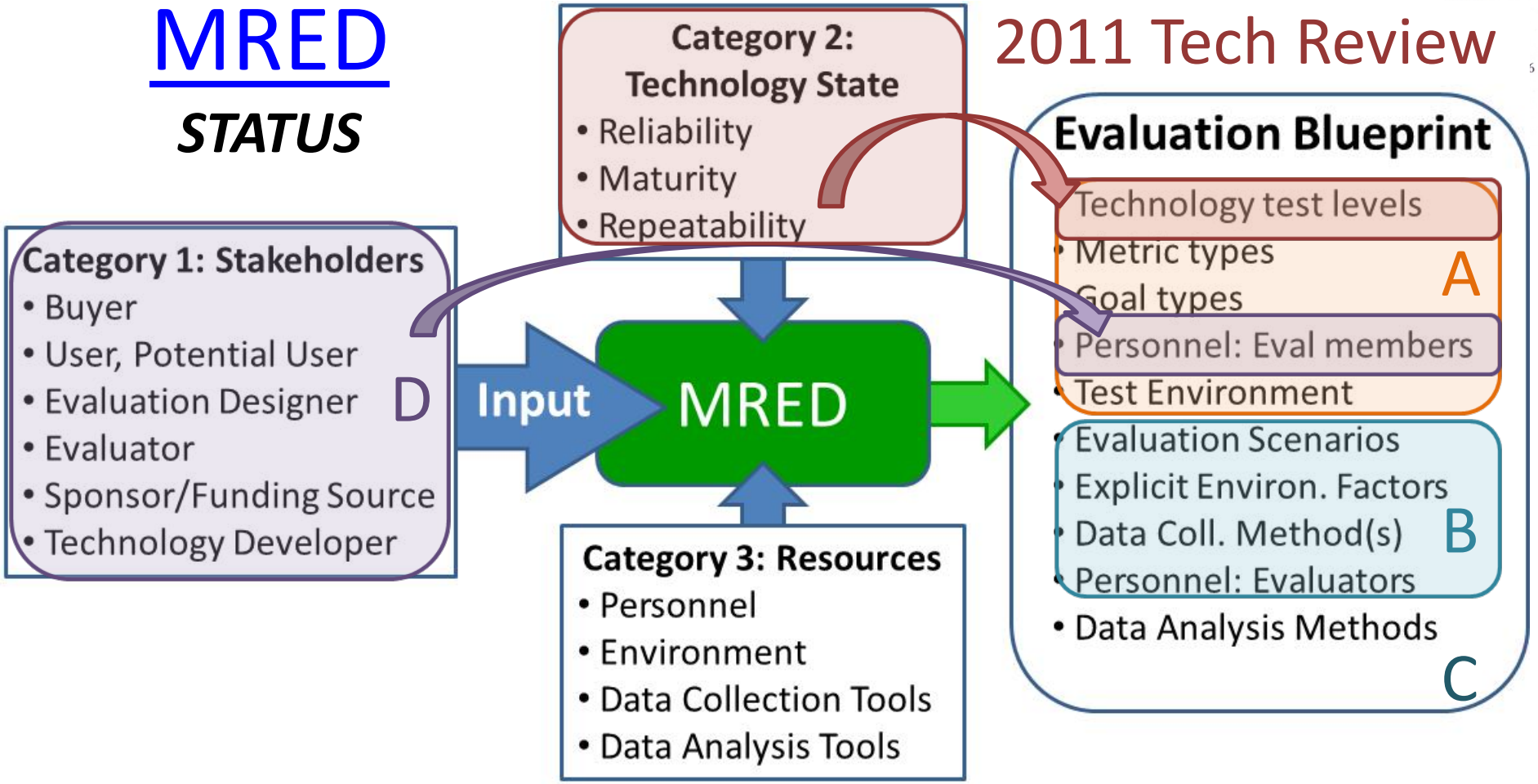
WHY IT'S NEEDED

- Tests of Advanced and Intelligent Systems tend to be Elaborate since the Technologies, themselves, are Complex
- Current Methods either Rely Heavily on Test Designer Knowledge and/or Experience OR do not Enable the Efficient Generation of Comprehensive Tests
- No Single Method has been Recognized as being Capable to Evaluate Quantitative and Qualitative Performance across a Range of Prototype and Physical Technologies Encompassing both Human-Controlled and Autonomous Capabilities

MRED

STATUS

2011 Tech Review



A. Weiss et al., 2010, “The Multi-Relationship Evaluation Design Framework: Designing Testing Plans to Comprehensively Assess Advanced and Intelligent Technologies,” ASME 2010 IDETC – 22ND International Conference on Design Theory and Methodology

B. Weiss & Schmidt, 2010, “The Multi-Relationship Evaluation Design Framework: Creating Evaluation Blueprints to Assess Advanced and Intelligent Technologies,” 2010 Performance Metrics for Intelligent Systems (PerMIS) Workshop.

C. Weiss & Schmidt, 2010, “The Multi-Relationship Evaluation Design Framework: Producing Evaluation Blueprints to Test Emerging, Advanced, and Intelligent Systems,” ITEA Journal, June, 2011.

D. Weiss & Schmidt, 2011, “Multi-Relationship Evaluation Design: Formalizing Test Plan Input and Output Elements for Evaluating Developing Intelligent Systems,” ASME 2011 IDETC)– 23RD International Conference on Design Theory and Methodology.

Key Definitions

OUTPUT ELEMENTS

- Technology Test Levels
  - *System:* Group of cooperative or interdependent *Components* forming an integrated whole to accomplish a specific goal.
  - *Component*: Essential part or feature of a *System* that contributes to the *System's* ability to accomplish a goal(s).
  - *Capability*: A specific ability of a technology. A *System* is made up of one or more *Capabilities*. A *Capability* is enabled by either a single *Component* or multiple *Components* working together.
- Metrics
  - *Technical Performance: Metrics* related to quantitative factors (such as accuracy, precision, time, distance, etc.).
  - *Utility Assessment: Metrics* related to the qualitative factors that express the condition or status of being useful and usable to the target user population.

Key Definitions – Goal Types

OUTPUT ELEMENTS

TECHNICAL PERFORMANCE
Quantitative Data Capture

1 - Component Level Testing
Breaks down a system into components in order to separate the subsystems that are essential for system functionality.

2 - Capability Level Testing
Requires the identification and isolation of specific capabilities from overall system behavior to the measure the individual capabilities' contribution to technical performance.

3 - System Level Testing
Targets a full system assessment where environmental variables can be isolated and manipulated to capture their impact on system performance.

UTILITY ASSESSMENT
Qualitative Data Capture

Users cannot interact with individual components, only capabilities and the full system.

4 - Capability Level Testing
Assesses the end-users' utility of a specific capability where the complete system's behavior is composed of multiple capabilities.

5 - System Level Testing
Evaluation type focuses on the end-users' utility of the entire system.

Technology State

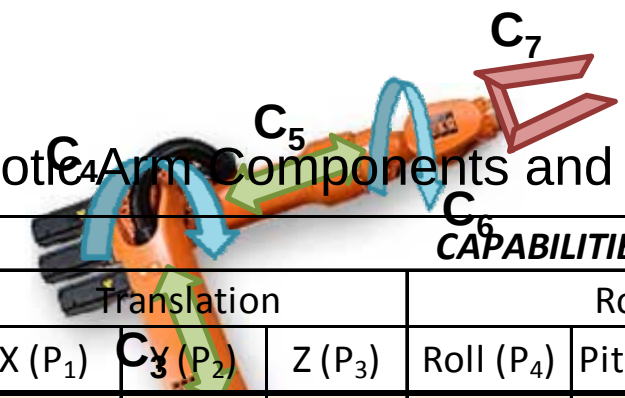
MRED INPUT CATEGORY

FACTORS	DEFINITION
Maturity	Technology's state or quality of being fully developed
Reliability	Technology's ability to perform a required function under stated conditions for a specified period of time
Repeatability	Technology's ability to yield the same or compatible results in previous test(s).

- Maturity
  - Defined for the *System*, and each *Capability* and *Component* to be tested
  - Values include *Non-Functional*, *Functional*, and *Fully-Developed*
  - Based upon technology developer feedback
- Reliability, Repeatability
  - Defined for the *System*, and each *Capability* and *Component* to be tested
  - Values quantitatively range from 0% to 100%
  - Based upon previous test data
  - *Repeatability* is addressed in future work (author welcomes input with respect to its value)

Structure

ROBOTIC ARM EXAMPLE



Robotic Arm Components and Capabilities

	CAPABILITIES						
	Translation			Rotation			Grasping
COMPONENTS	X (P_1)	Y (P_2)	Z (P_3)	Roll (P_4)	Pitch (P_5)	Yaw (P_6)	(P_7)
Revolute 1 (C_1)	X	X				X	
Revolute 2 (C_2)		X	X		X		
Prismatic 1 (C_3)	X	X	X				
Revolute 3 (C_4)	X		X	X			
Prismatic 2 (C_5)	X	X	X				
Revolute 4 (C_6)				X	X	X	
Gripper (C_7)							X

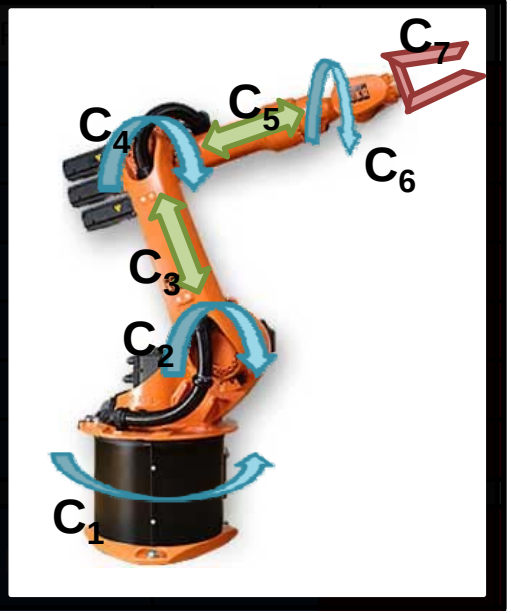
Note: The design of the physical arm shown is for example and does not necessarily reflect the actual *Component/Capability* relationships in the table.
Robot arm image courtesy of www.robots.com

Maturity

ROBOTIC ARM EXAMPLE

- Maturity Levels
  - *Non-Functional:* Testing cannot be performed on a *TTL* in this state
  - *Non-Functional to Functional:* Testing may be performed on a *TTL* in this state depending upon the specific technology, *Stakeholders'* discretion, etc.
  - *Functional:* Limited testing can be conducted on a *TTL* in this state
  - *Fully-Developed:* The *TTL's* state does not pose any restrictions on testing

		CAPABILITIES				
COMPONENT MATURITY	***COMPONENTS***	Translation			Rotation	Grasping
		X (P_1)	Y (P_2)	Z (P_3)	Roll (P_4)	
Fully-Developed (FD)	Revolute Joint 1 (C_1)	X	X			
Fully-Developed (FD)	Revolute Joint 2 (C_2)		X	X		
Functional (FN)	Prismatic Joint 1 (C_3)	X	X	X		
Functional (FN)	Revolute Joint 3 (C_4)	X		X	X	
Functional (FN)	Prismatic Joint 2 (C_5)	X	X	X		
Non-Functional (NF)	Revolute Joint 4 (C_6)				X	
Non-Functional (NF)	Gripper (C_7)					
	CAPABILITY MATURITY	FN	FN	FN	NF to FN	



Reliability

ROBOTIC ARM EXAMPLE

- Reliability
  - Below table assumes that *Capability Reliability* cannot be measured directly and *Capability Reliability* is product of *Component Reliability*
  - If *Capability Reliability* can be measured directly, then values could differ from the calculated values.

COMPONENT RELIABILITY	COMPONENTS	CAPABILITIES						
		Translation			Rotation			Grasping
		X (P_1)	Y (P_2)	Z (P_3)	Roll (P_4)	Pitch (P_5)	Yaw (P_6)	(P_7)
99%	Revolute Joint 1 (C_1)						X	
98%	Revolute Joint 2 (C_2)				X			
72%	Prismatic Joint 1 (C_3)	X	X	X				
[illegible]	[illegible]				X			
[illegible]	[illegible]					X	X	
[illegible]	[illegible]							X
	CAPABILITY RELIABILITY	23.6%	35.6%	23.4%	1.95%	2.94%	2.97%	No Data

Data Corresponds to Fully-Developed Components

Data Corresponds to Functional Components

Data Corresponds to Non-Functional Components

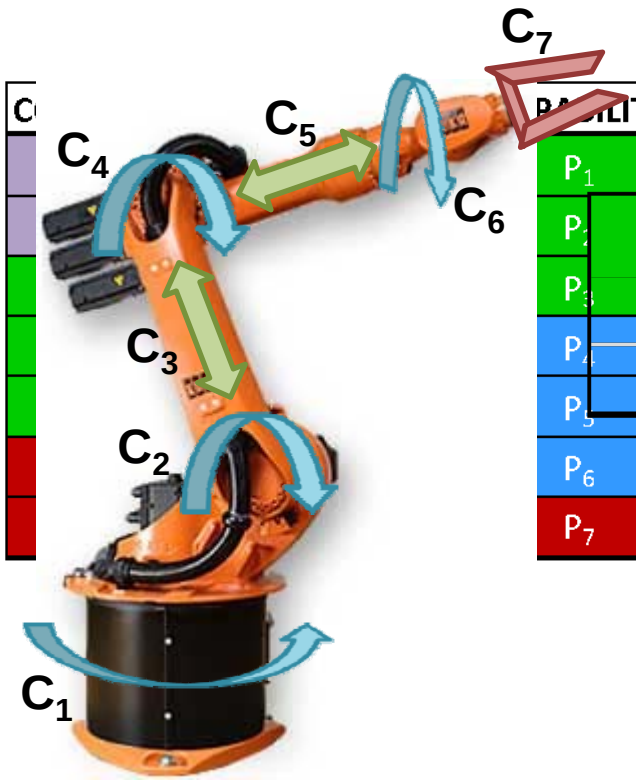
Data Corresponds to Non-Functional to Functional Capabilities

Data Corresponds to Functional Capabilities

Technology State Factor Influence

ROBOTIC ARM EXAMPLE

Available Technology Test Levels



C_1	C_2
Fully-Developed	
FULL TESTING AVAILABLE	

LEGEND
Fully-Developed (FD) element
Non-functional to Functional (NF to FN) element
Non-functional (NF) element

C_3	C_4	C_5	P_1	P_2	P_3
LIMITED TESTING AVAILABLE					

P_4	P_5	P_6	S_1
Non-Functional to Functional			
NO to LIMITED TESTING AVAILABLE			

C_6	C_7	P_7
Non-Functional		
No TESTING AVAILABLE		

Further Influence

ROBOTIC ARM EXAMPLE

- *Technology Test Levels*
  - Example highlights influence by *Maturity*
  - *Reliability* and *Repeatability* can have further influence
  - Will be influenced by other inputs (*Stakeholders* and *Available Resources*) in addition to *Technology State Factors*

- *Goal Types*
  - Since *Technology State Factors* dictate which *Technology Test Levels* are available, they therefore influence the available *Goal Types* for testing

- Further Influence – *Technology State Factors* will influence other input categories and additional output blueprint elements including…
  - Sponsor/Funding Source (Input - *Stakeholder*)
  - Technology Developer (Input - *Stakeholder*)
  - Test Environment (Output - Blueprint Element)

Future Work

WRAPPING IT UP

- Simple Robot Arm Example Illustrates MRED's Broad Potential to be Applied to the Evaluation Design of Complex Systems
- Upcoming Efforts will Formalize Relationships among the Input Categories and Output Blueprint Elements
- Expansion of the MRED Model will Lead to Mathematical Formalization
- Robot Arm Example will be used to Explore *Resources* Input Category Including the Outputs it Influences
- Examples Involving more Complex Technologies will be Used

NIST
National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

UNIVERSITY OF MARYLAND 1856

For more information, please contact:

Brian A. Weiss

brian.weiss@nist.gov

301-975-4373

Dr. Linda C. Schmidt

lschmidt@umd.edu

301-405-0417