



# Data to Support Stabilization & Reconstruction (S&R) Operations

December 16, 2004



# Agenda



- Objectives, Sources
- Nature of the Problem
- A Strawman Initiative
- Key Issues



# Objectives



- Review key prior data initiatives (with emphasis on S&R operations) to derive major
  - Findings
  - Recommendations
- Formulate a plan of action to shape further efforts



# Key Sources of Information

- DoD Net-Centric Data Strategy (May 9, 2003)
- HIU Workshop on "Complex Emergencies: Collecting Data, Managing Information, Seeking Knowledge" (September 2003)
- MORS Workshop on "Improving Defense Analysis Through Better Data Practices" (Spring 2003)
- NDU Workshop on "Communications, Information Systems, and Networks for S&R Operations" (October 2004)
- DTIC briefings on DoD Metadata Registry
- SO/LIC, HIU Workshop on Civil Affairs (April 14, 2004)



# Agenda



- Objectives, Sources
- *Nature of the Problem*
- A Strawman Initiative
- Key Issues



# Background: The Data Problem is *Not New*



"The government are very keen on amassing statistics.

They collect them, add them, raise them to the n-the power, take the cube root and prepare wonderful diagrams.

But you must never forget that every one of these figures comes in the first instance from the village watchman, *who just puts down what he damn well pleases.*"

--Comment of an English judge on the subject of Indian statistics; Quoted in Sir Josiah Stamp in "Some Economic Matters in Modern Life" (1940)



# Key Terms -- Analogies to a Library



<i>Key Terms</i>	<i>Library Analogy</i>
Metadata Registry	Description of what information is to be filled out on each card in a library card catalog (e.g., structure, format, definitions of data)
Metadata Catalog	The actual filled out cards that describe each of the holdings in the library
Shared Space	The bookshelves in a library



# The S&R Data Problem

- Multi-sectoral (e.g., political, military, health/nutrition, shelter/logistics, agriculture)
- Multi-dimensional (e.g., affected area and population, vulnerable groups, assistance, resources)
- Contextual (e.g., historical, cultural, ethical)
- Multi-source (e.g., US Government, IO, NGOs, media)
- Non-standardized (e.g., formats, definitions, indicators, measurement indicators, methodologies)
- Twin dilemmas -- information overload *and* gaps

Source: Dennis King, HIU



# The Data Problems that the S&R Community Are Facing Include...



- Data sharing
- Data purity
- Metadata policy (e.g., standardization)
- Data shelf life
- Data naming conventions
- Data reconciliation
- Data maintenance
- Ontological development for intelligent searches
- Data acquisition
- Data conversion
- Data reuse
- Lack of good data dictionaries
- Lack of knowledge of original purpose
- Data protection
- Data provenance
- Data subrogation
- Data bloat

...



# Proposed Data Goals

- There is a need for data that are:
  - **(Rapidly) Available:** *visible, accessible, institutionalized*
  - **Usable:** *understandable, trusted, interoperable, responsive to user needs*

[Source: John Osterholz, OASD(NII), on Enterprise Data Strategy Goals]



# Barriers to Data Reuse

- Lack of knowledge about existence of legacy data
- Security restrictions
- Quality of metadata (e.g., failure to document conditions of collection)
- Varying
  - Definitions
  - Language
  - Measurement instruments
- Form of accessible data
- Rapid change of technical data
- Fear (e.g., misuse, misunderstanding, adverse consequences)



# Agenda



- Objectives, Sources
- Nature of the Problem
- *A Strawman Initiative*
- Key Issues



# A Strawman S&R Data Proposal



- Undertake a data initiative for S&R that systematically employs the DoD Net-Centric Data Strategy, suitably adapted to satisfy the needs of the S&R Community of Interest (COI)
- Employ an *evolutionary* approach that stresses the
  - Development of broad data *needs* for the S&R COI
  - Characterization of an *architecture* to guide the activity
  - Rapid development of a *core* capability to satisfy immediate needs
  - Periodic development of *increments* that
    - Reflect feedback from the users
    - Provide additional capabilities, functionalities
    - Incorporate emerging information technologies



# Net-Centric Data Strategy



- Key attributes of the strategy include
  - Ensuring data are visible, available, and usable when needed and where needed to accelerate decision-making
  - “Tagging” of all data with metadata to enable discovery of data by users
  - Posting of all data to shared spaces to provide access to all users except when limited by security, policy, or regulations
  - Advancing the (COI) from defining interoperability through point-to-point interfaces to enabling the “many-to-many” exchanges typical of a net-centric data environment

Source: Adapted from Memo from DoD CIO, May 9, 2003



# Key Elements of the Data Vision



- *COIs* to address organization and maintenance of data
- *Metadata*, which provides a way to describe data assets and the use of registries, catalogs, and shared spaces, which are mechanisms to store data and information about data
- *Global Information Grid (GIG) Services* that enable data tagging, sharing, searching, and retrieving



# Net-Centric Data Strategy: Approaches to Achieve Data Goals (1 of 3)



- Goal 1: Make data visible
  - Post data to shared spaces
  - Associate discovery metadata with data assets
  - Create and maintain catalogs
  - Register metadata related to structure and definition
  - Inventory data assets
- Goal 2: Make data accessible
  - Create shared spaces and data access services
  - Associate security-related metadata



# Net-Centric Data Strategy: Approaches to Achieve Data Goals (2 of 3)



- Goal 3: Institutionalize data management
  - Govern data processes with sustained leadership
  - Incorporate data approaches into COI processes and practices
  - Advocate, train, and educate in data practices
  - Adopt metrics and incentives
- Goal 4: Enable data to be understandable
  - Define COI-specific ontologies
  - Associate content-related metadata
  - Associate format-related metadata
  - Define COI-specific content-related metadata



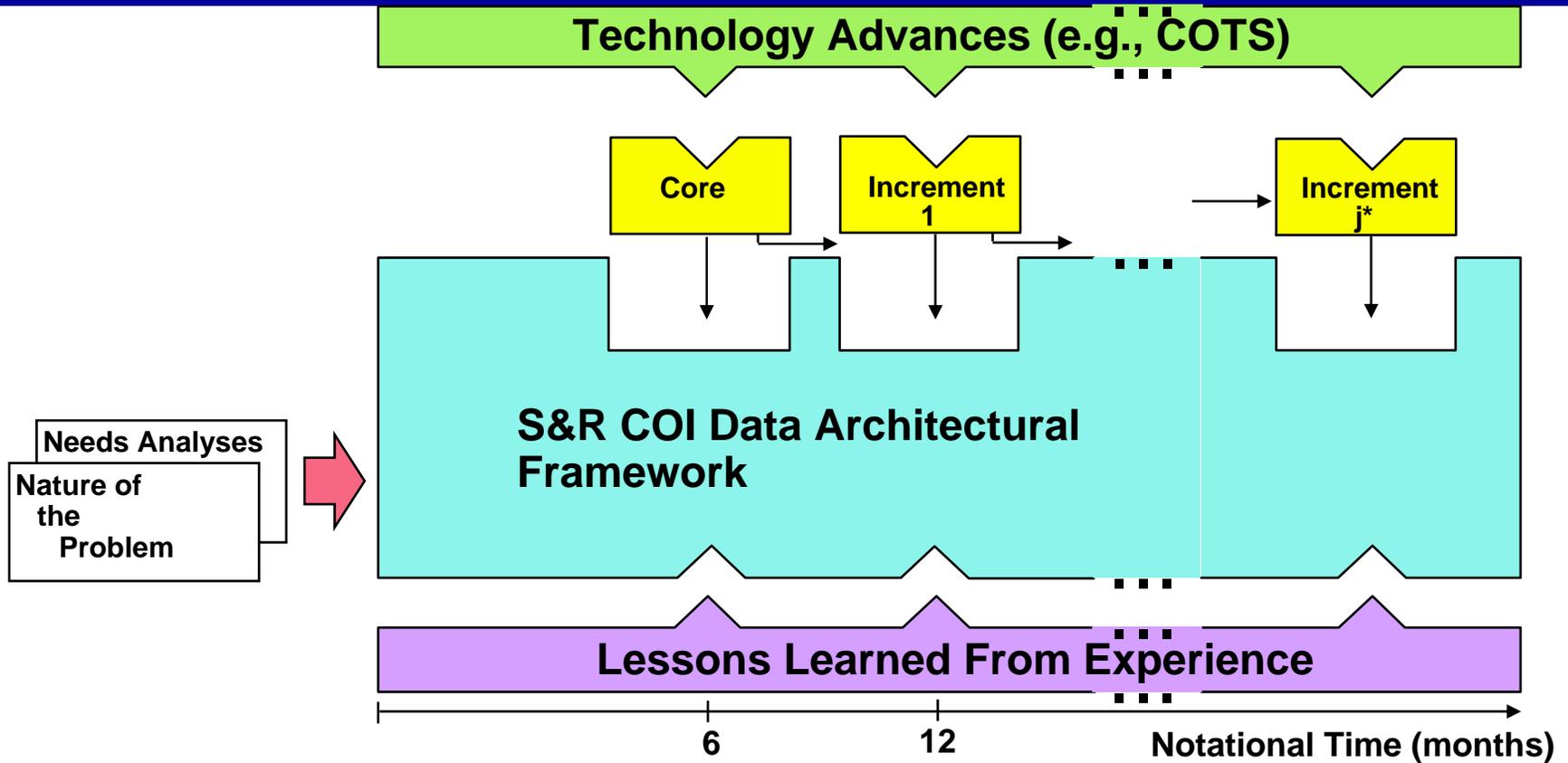
# Net-Centric Data Strategy: Approaches to Achieve Data Goals (3 of 3)



- Goal 5: Enable data to be trusted
  - Associate data pedigree and security metadata
  - Identify authoritative sources
- Goal 6: Support data interoperability
  - Register metadata
  - Associate format-related metadata
  - Identify key interfaces between systems
  - Comply with net-centric interface standards
- Goal 7: Be responsive to user needs
  - Involve users in COIs
  - Establish a process to enable user feedback



# Depiction of the Evolutionary Approach for S&R Data





# Agenda



- Objectives, Sources
- Nature of the Problem
- A Strawman Initiative
- *Key Issues*



# Issues to Address (1 of 2)



- Is it appropriate to adopt (and adapt) the DoD Net-Centric Data Strategy?
- Should the effort focus on data associated with existing S&R activities (e.g., Afghanistan or Iraq) or should it address potential future activities (e.g., HA/DR; future S&R operations)?
- What segment of the problem might be used to create the initial core activity?
- Is it possible to exploit on-going activities to support the proposed effort (e.g., build on initiatives begun in the HIU Workshop on “Complex Emergencies: Collecting Data, Managing Information, Seeking Knowledge” (September 2003))?



# Issues to Address (2 of 2)



- Who would constitute the COI (e.g., Core? Augmentees?)
- Where would the resources come from to support the effort?
- Many of the findings of the HIU workshop identified *training* as a key issue (e.g., with respect to strategies and systems for the management of data, information, and knowledge; use of emerging tools, processes); should we
  - Identify a range of training initiatives to support all of the participants in S&R operations?
  - Convene a workshop to explore options to implement these training initiatives?



# Back-up Slide





# Background: The Data Problem is *Not New*



- "It is a capital mistake to theorize before one has data"
  - *Sherlock Holmes*
- "Data! Data! Data!" he cried impatiently. "I can't make bricks without clay."
  - *Sherlock Holmes*
- "*Metadata! Metadata! Metadata!...*"
  - *Simone Youngblood, MORS September 2003*
- "Theory without data = philosophy; data without theory = noise"
  - *Anonymous*
- "Researchers are as willing to use other people's data models as they are to wear other people's underwear"
  - *Richard Morris, NIH, MORS September 2003*