

# DATA COLLECTION & MANAGEMENT (DC&M) FOR ANALYSIS SUPPORT TO OPERATIONS

Ms Jacqueline Eaton, S&T Advisor, NATO  
NATO OR&A Conference, October 2018

NATO System Analysis and Studies Panel Research Task Group 111



# AGENDA

1. Introduction
2. DC&M Problem: Data in a Deployed HQ
3. DC&M Capability: Process, People, Tools
4. Key Findings

# AGENDA

- 1. Introduction**
2. DC&M Problem: Data in a Deployed HQ
3. DC&M Capability: Process, People, Tools
4. Key Findings

# THE PROBLEM

80% of analysts' time is spent  
on data collection & prep

- It takes ages to find the data in the first place
- Once it is discovered, the owner won't share
- When it is acquired, it is a complete disorganized mess
- When analysis is done, the data is discarded/deleted

# CONTEXT

- Digitisation of the battlefield is increasing the volume, velocity and variety of data available
- Analysts supporting current operations already struggle to find and manage existing data
- New methods and technologies can increasingly effectively automate traditionally manpower intensive data collection, preparation and management tasks

## RESEARCH QUESTION

**How can deployed NATO HQs enhance their ability to collect and manage the data required for analytical support to operations, including that for 'big data' analytics?**

# STUDY OUTLINE

- Year 1 – What are the issues? (First Year Report)
- Year 2 – What are the solutions?
- Year 3 – Guide to DC&M

# PARTICIPATION

- Co-Chairs – DEU, NATO JALLC/NATO STO-OCS
- Nations – FRA, GBR, TUR, USA
- NATO Orgs – HQ MARCOM, NCIA, (ACT)

# AGENDA

1. Introduction
- 2. DC&M Problem: Data in a Deployed HQ**
3. DC&M Capability: Process, People, Tools
4. Key Findings

# THE COMMANDER'S PROBLEM

- Bad data leads to poor decisions, good data enables good decisions
- Good decisions maximize effectiveness and minimize risk, bad decisions can cause injury, loss of life and mission failure

# THE ANALYST'S PROBLEM

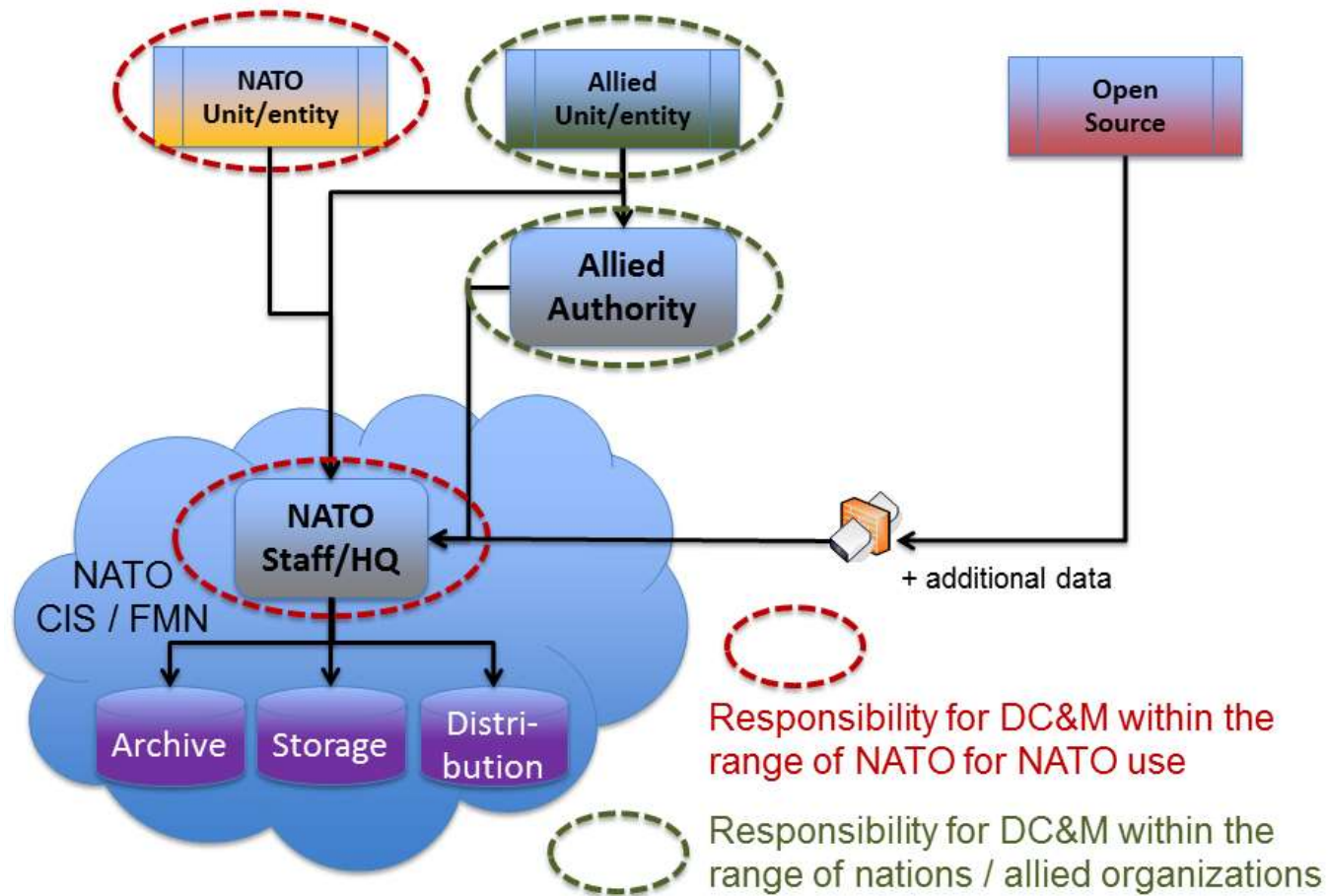
- Data collection planning is fragmented across the HQ
- Most data is collected on the fly and is therefore not ready for use in short-term analysis
- Limited or no resources are available to make data accessible and reusable, even within the HQ
- Data cannot be safely reused by anyone who does not understand the context in which it was collected



# MILITARY DATA IS SPECIAL

- Classification and ownership
- Specialized Military Data
  - date-time, mapping, formatted messages
- Data Standards
  - naming, taxonomy
- Accessing Military Data
- Availability of Military Data
  - digital vs non-digital
- Military CIS Systems
- Tools for Data Analysis

# DATA COMES FROM EVERYWHERE



# BIG DATA ELEPHANT IN THE ROOM



# AGENDA

1. Introduction
2. DC&M Problem: Data in a Deployed HQ
- 3. DC&M Capability: Process, People, Tools**
4. Key Findings

# DC&M GUIDE

1. Why Data is Important
2. Data in a Military Context
3. Military Data Sources: Military Data is Special
4. Data Collection and Preparation
5. Data Sharing
6. Data Archiving
7. Data Roles
8. Data Tools

**Problem**

**Process**

**People  
Tools**

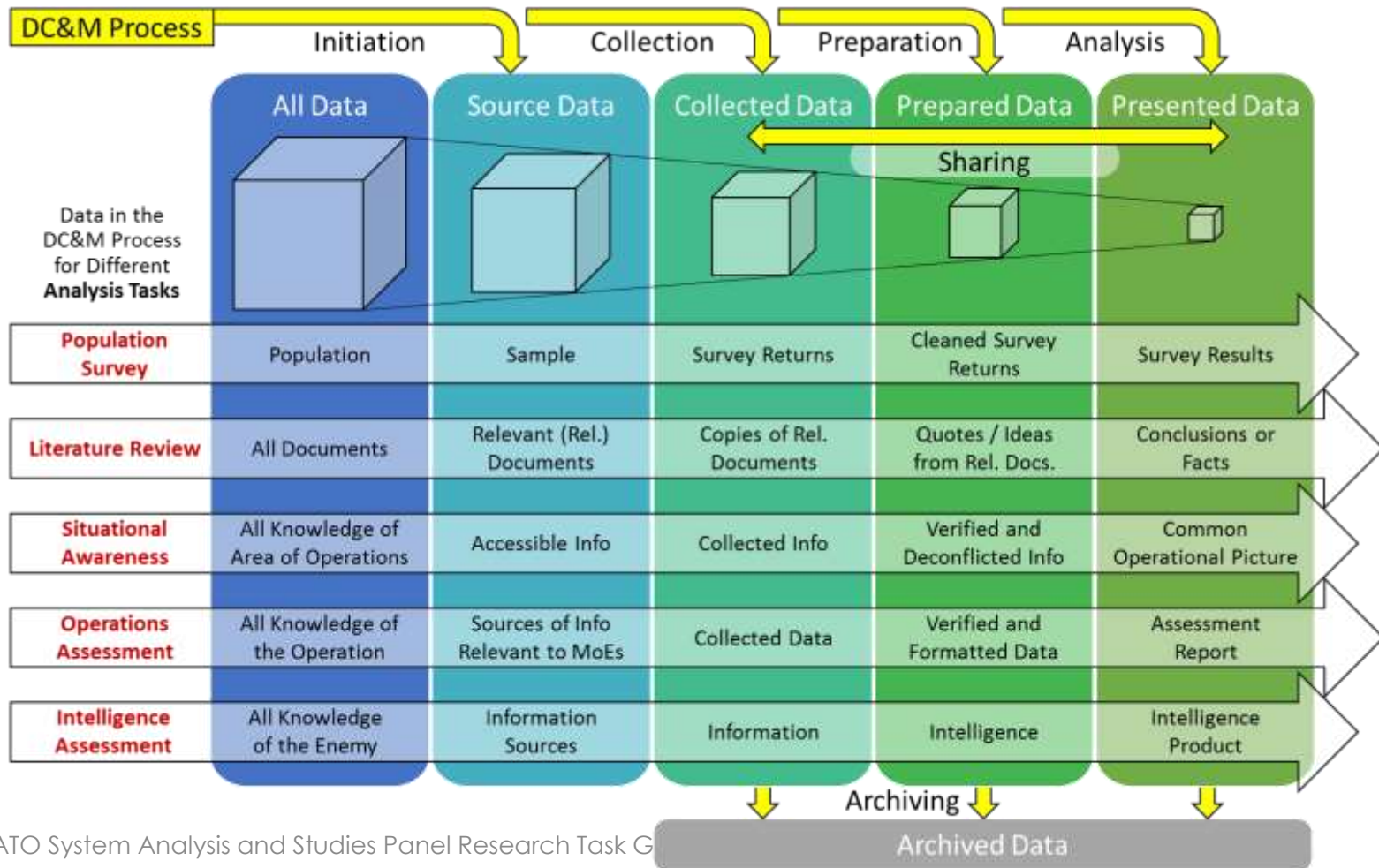
## AIM

**Provide advice on the DC&M capability (DOTMLPF-I) required for deployed HQs to achieve decision superiority**

(i.e. to free up analysts to focus on the clever stuff)

# GENERIC DC&M PROCESS

- Derived from Ops Assessment and Intelligence Processes



# DATA COLLECTION PLANNING IS KEY

- The most challenging data demands come from short-term analysis tasks
- A standard data collection plan for the routine collection and management of data is therefore essential
- However, requests for analysis support can rarely be predicted in detail
- Any data collection plan should be comprehensive and HQ-wide
- Big Data tools and techniques can help collect, prepare, analyse and manage some data but are only part of the solution

# DATA ACQUISITION IS COSTLY/COMPLEX

- Data acquired from another entity, either free of charge or paid, can rarely be used without considering licensing or legal constraints, of which military staffs are generally unaware
- The acquisition process is time consuming, so should be started early
  - Free data from the internet often needs sophisticated algorithms to extract and prepare
  - Proprietary data requires memorandums of understanding
  - Purchased data requires funds to be released/allocated
- Quality data is worth paying for, so all military HQs should have budget allocated for data acquisition



# AUTOMATED DATA COLLECTION & PREPARATION IS PREFERRED

- Manual input is extremely time and effort consuming
- Proper data processing and Big Data tools can be effective in reducing analyst's workload
  - Analyst's focus will shift to data analysis, leading to better informed decisions
  - Analytical results will rely on data of higher quality
  - Analyst will be able to answer more questions more timely and more accurate for enhanced decision making

# DATA SHARING/REUSE SHOULD BE THE NORM NOT THE EXCEPTION

- Preparing data for sharing depends on who it will be shared with:
  - Own HQ
  - Non-NATO Entities and Partners not connected to NATO CIS
  - NATO Entities, Operational Units or Partners connected to NATO CIS
  - Future Users including your future self
- The do minimum action is to add metadata in a universally readable format to encourage appropriate reuse

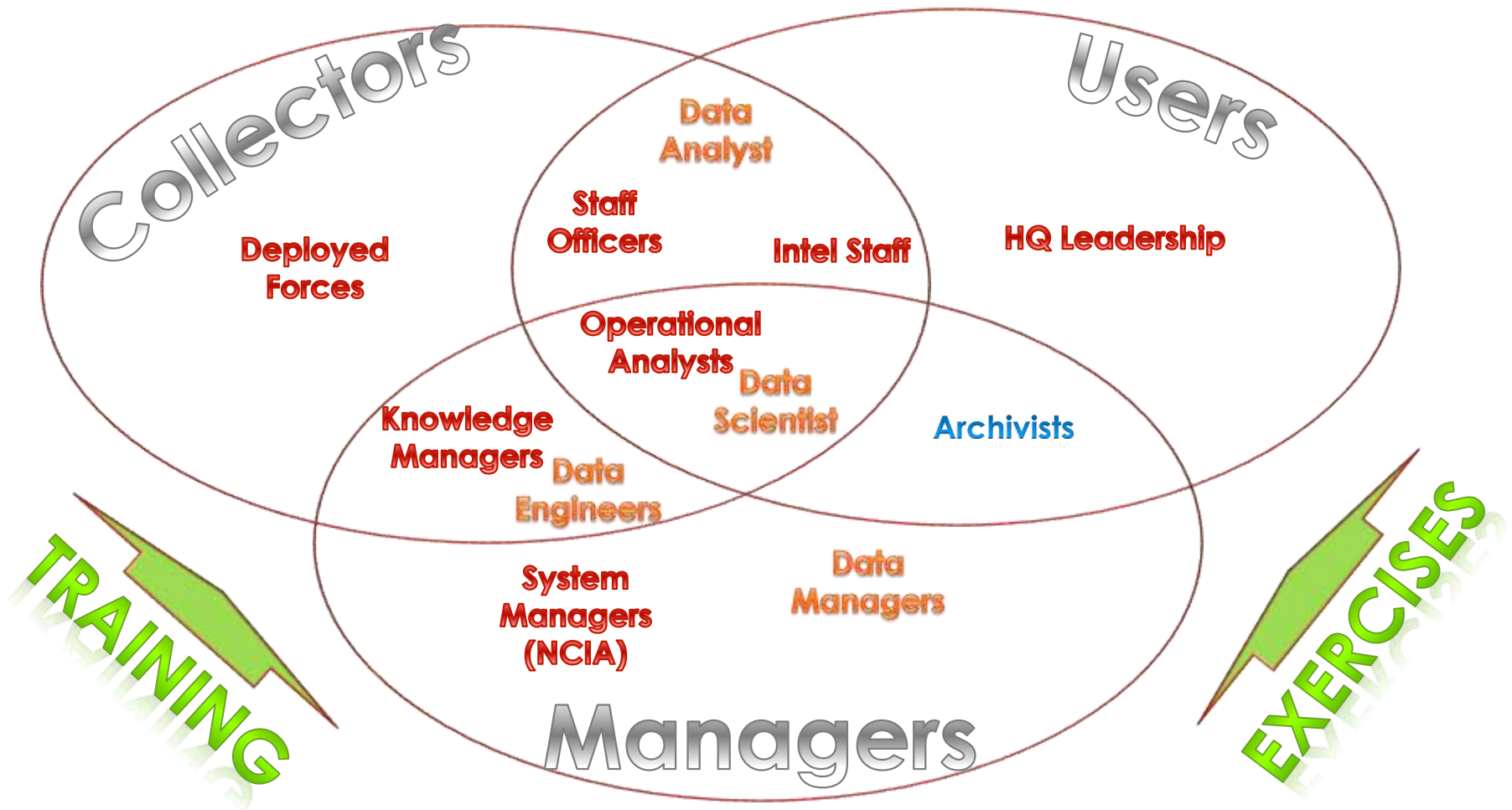
# DON'T FORGET ARCHIVING

- Data Archives are both critical and valuable assets
- In a 'Big' or 'Bigger' Data future, no data may be considered obsolete
- Archive solutions need to be secure, controllable, and recoverable
- NATO military data archives must also align with legal regulatory compliance
- Archiving requirements should be formulated at the start of every NATO operation and not be something considered in the aftermath

# (SUB-)GUIDE TO CREATING A DATA ENVIRONMENT

- Written to guide people through the requirement design process of the development of a data collection environment
- Describes this process in a standardized manner based on a Mission Thread Approach
- The aim is to give recommendations on data environments for:
  - Any type of data format
  - Any type of data source
  - Any type of data base
  - Any type of NATO information system
- The guide includes a Metadata checklist

# DATA ROLES



# DATA TOOLS

## Data Collection

Amazon Web Services

Google Cloud

Microsoft Azure

DataPreparator

Carrot2

Elastisearch

WEKA

R

NLTK

PYTHON

RAPIDMINER

ORANGE

PANDAS

## Data Preparation

# KEY FINDINGS / TAKEAWAYS

- Access to more comprehensive data leads to more insightful decisions
- Bigger data, but perhaps not Big Data
- SAS-111 DC&M Process matches
  - Intelligence Data Collection (AJP2)
  - Operations Assessment (NOAH)
- Specific data needs are hard to predict, so data collection needs to be routine
- The required skills and tools are not commonplace in military HQs
  - Analysts typically do 80% data collection, 20% analysis
- Data archives are a critical and valuable asset

# BACK UP SLIDES

NATO System Analysis and Studies Panel Research Task Group 111



# SAS-111 - NATIONAL STUDY

## DEVELOPMENT OF A DATA COLLECTION ENVIRONMENT BASED ON A MISSION THREAD APPROACH

Heger, UniBw München

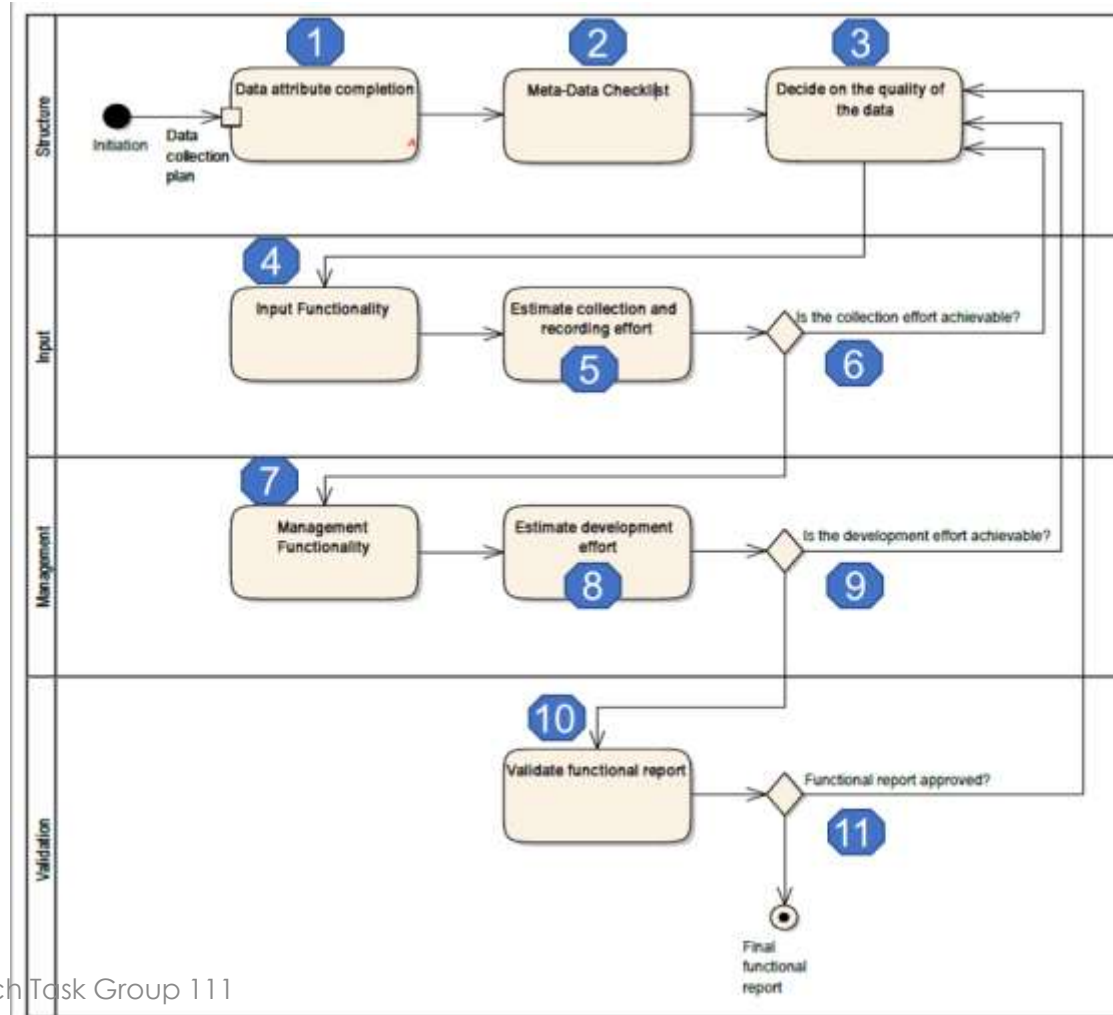


# PROBLEM STATEMENT

- Use cases show that data is still often collected on the fly
- The importance of its reusability is not recognized
- Partners that were not present during the collection process are likely to have less knowledge on the context in which the data was gathered
- Accurate data is necessary to evaluate performance measures (MOPs/MOEs)
- Analysis based on data is necessary for well informed decision making

# META-DATA CHECKLIST

- Preserve meaning of data and context in which it was collected  
 → different aspects of data to be described and stored along with dataset = metadata
- Metadata reduces problems of misunderstanding the meaning and context of data



# CONCLUSION

- The national study “Development of a Data Collection Environment based on a Mission Thread Approach” uses NATO architecture diagrams such as Activity Charts to give a structured overview and guideline of the process
- As a result, the people responsible for designing the requirements for the data collection environment are aided in that process in order to find a Pareto optimal solution that ensures
  - required quality of the collected data
  - the required ease and speed of insertion
  - while keeping the development costs proportionate