Meeting the Unmet Data Collection & Management Requirements of Big Data Analytics and AI for Military Decision Making

Collection and Management of Data for Analysis Support to Operations

System Analysis and Studies Panel
Research Task Group 111

31 May 2018

Agenda

1. **Study Background**

2. **The Military Data Requirement**

3. **DC&M Process**

4. **Mission Thread Approach**

Premise

- Digitisation of the battlefield will increase the volume, velocity and variety of data available
- Analysts supporting current operations struggle to find and manage data
- New methods and technologies for collecting, managing and analysing vast amounts of data are the challenge of our age
- 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 Methodology

Year 1 – What are the issues? (First Year Report – In Collaboration Portal)

Year 2 – Engagement – Solution paths

Year 3 – Recommendations

Participation:
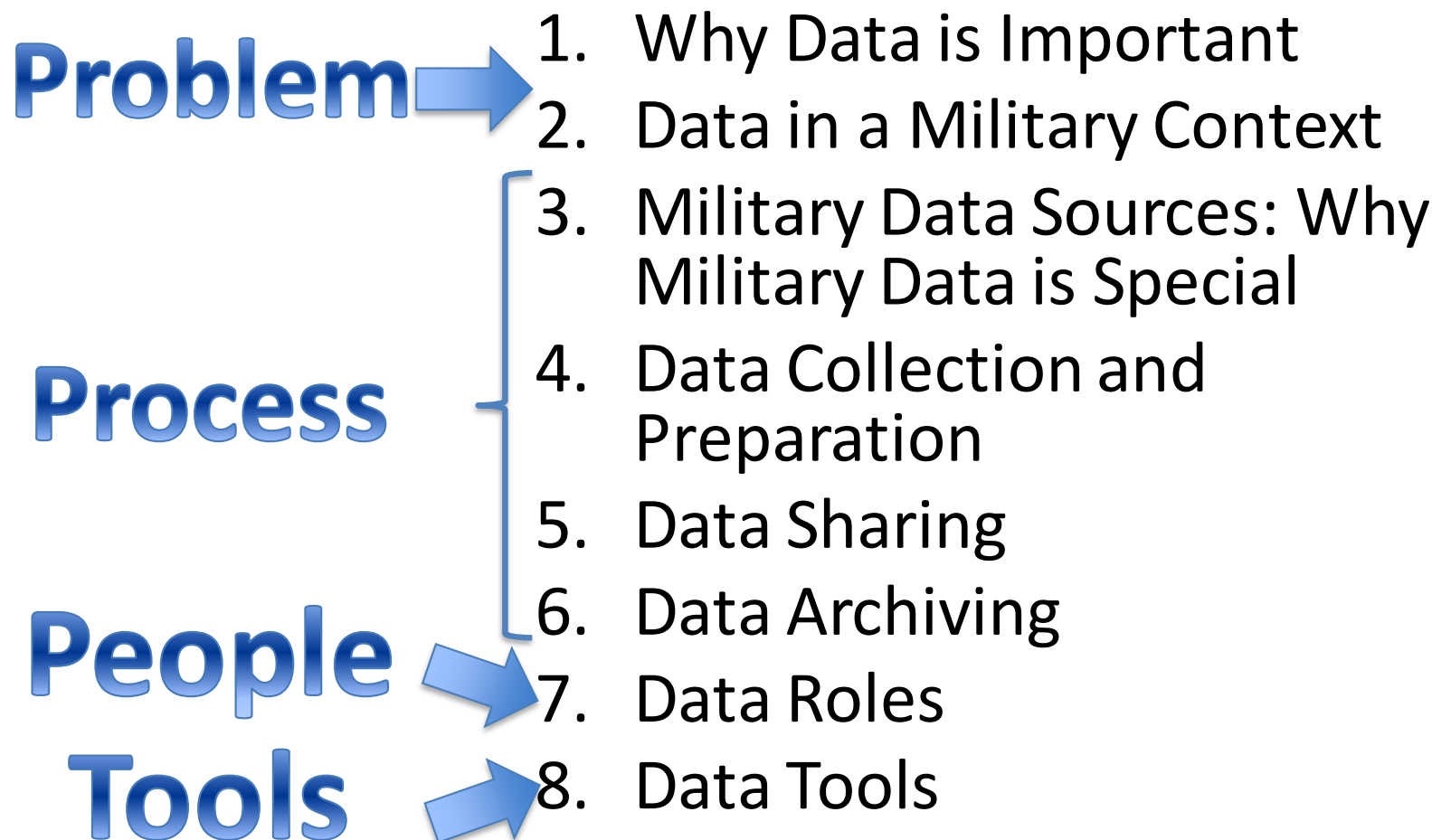
Co-Chairs: DEU, JALLC/OCS

Nations: FRA, GBR, TUR, USA

Organisations: MARCOM, NCIA, (ACT)

Guide to DC&M

- ## People
  - All staff are involved to some extent, some staff are experts

- ## Process
  - Initiation, Collection, Preparation, (Analysis), Archiving

- ## Tools
  - Data Clouds, Data Lakes, Bigger Data, Open Source, Data Selection, Data Mining, Programming
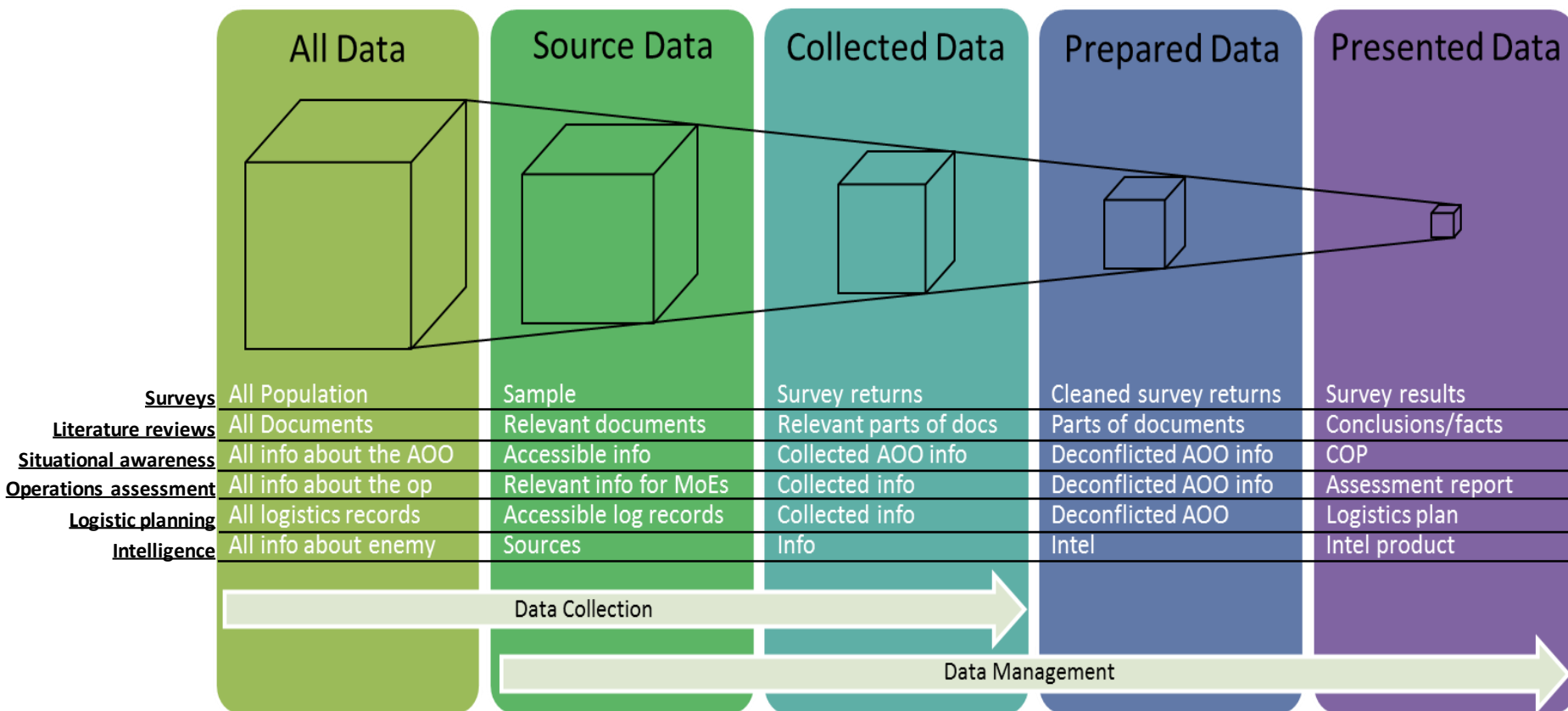
Guide Structure

Problem →
1. Why Data is Important
2. Data in a Military Context

Process
3. Military Data Sources: Why Military Data is Special
4. Data Collection and Preparation
5. Data Sharing
6. Data Archiving

People →
7. Data Roles

Tools →
8. Data Tools

Agenda

1. **Study Background**

2. **The Military Data Requirement**

3. **DC&M Process**

4. **Mission Thread Approach**

From Data to Knowledge

	All Data	Source Data	Collected Data	Prepared Data	Presented Data
Surveys	All Population	Sample	Survey returns	Cleaned survey returns	Survey results
Literature reviews	All Documents	Relevant documents	Relevant parts of docs	Parts of documents	Conclusions/facts
Situational awareness	All info about the AOO	Accessible info	Collected AOO info	Deconflicted AOO info	COP
Operations assessment	All info about the op	Relevant info for MoEs	Collected info	Deconflicted AOO info	Assessment report
Logistic planning	All logistics records	Accessible log records	Collected info	Deconflicted AOO	Logistics plan
Intelligence	All info about enemy	Sources	Info	Intel	Intel product

Data Collection

Data Management

NATO
OTAN

NORTH ATLANTIC TREATY ORGANIZATION
SCIENCE AND TECHNOLOGY ORGANIZATION

S&T
organization

Data to Knowledge - Problem Statement

- Military commanders wish to reduce losses and minimize risk by better informed decisions

- Current use cases show that data is still often collected on the fly.

- The importance of its reusability is <u>not</u> 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.

- However, access to large amounts of data in a timely manner has proved challenging

Data Led Analysis Requirements

Timelines

Staff/HQ tasks	Short-term analysis	Long-term analysis	Historic analysis
Conflict system analysis		X	X
Operations assessment	X		
Planning assistance	X	X	
JOA analysis		X	X
What-if		X	
Training		X	X
War gaming, simulation	X	X	
Capability development	X	X	
Force development and composition		X	
Concept development		X	
Review of staff/HQ structure and composition		X	X
CIS-Planning	X	X	

NATO
OTAN
NORTH ATLANTIC TREATY ORGANIZATION
SCIENCE AND TECHNOLOGY ORGANIZATION
S&T
organization

Data Creating Organisational Relationships

NATO
OTAN
NORTH ATLANTIC TREATY ORGANIZATION
SCIENCE AND TECHNOLOGY ORGANIZATION
S&T
organization
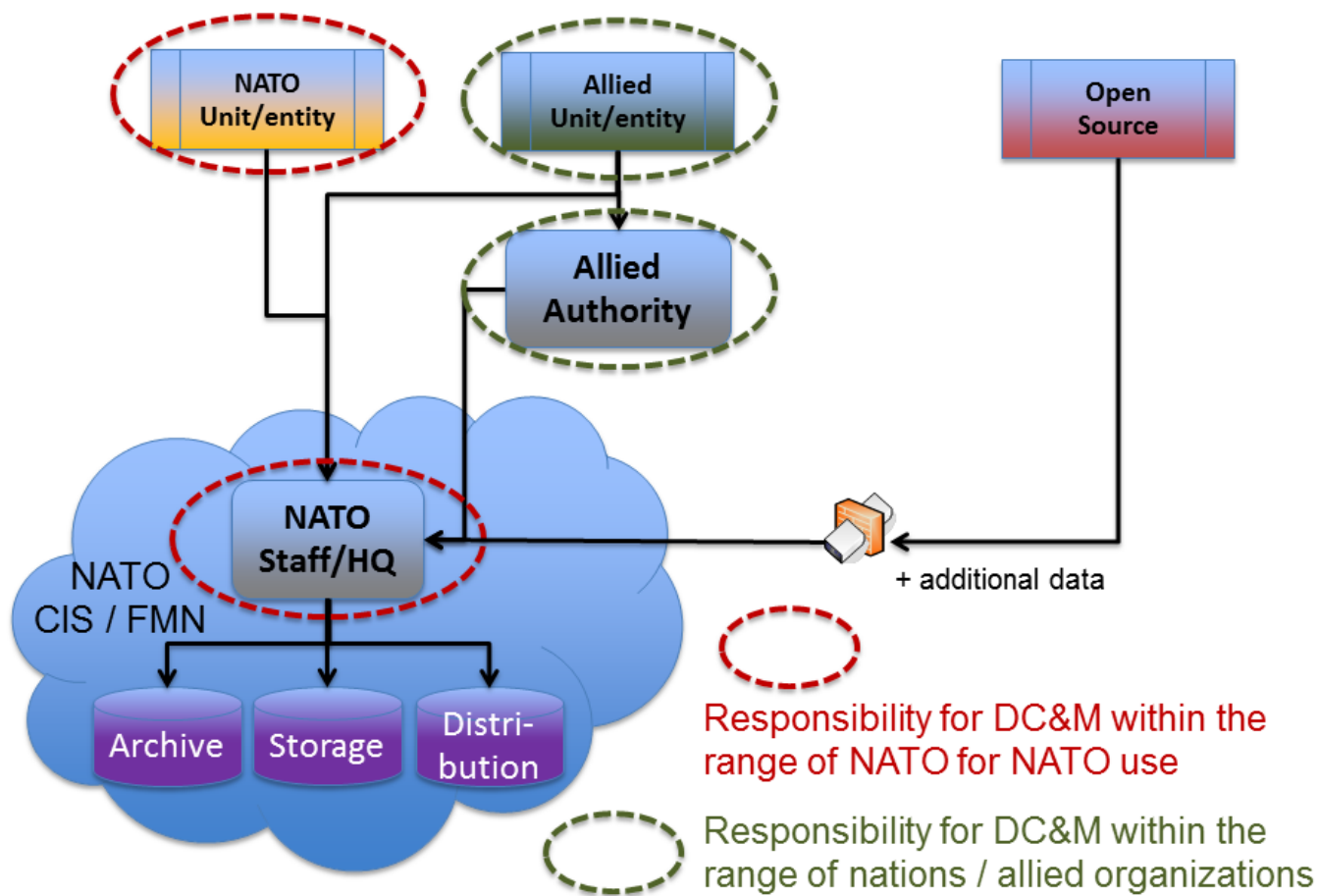
Why Military Data is Special

- Classification and ownership

- Specialised 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

NATO
OTAN
NORTH ATLANTIC TREATY ORGANIZATION
SCIENCE AND TECHNOLOGY ORGANIZATION
S&T
organization

Data to Knowledge in a Military Environment

- Requests for analysis support can rarely be predicted in detail
    - Any data collection plan should be comprehensive
- The most challenging demands in terms of data evolve mainly from short-term analysis tasks
- A standard data collection plan for the routine collection and management of data is essential
- Big Data tools and techniques to collect, prepare, analyse and manage some data but it is not the sole solution for NATO's bigger data problem.

Big Data Elephant in the room
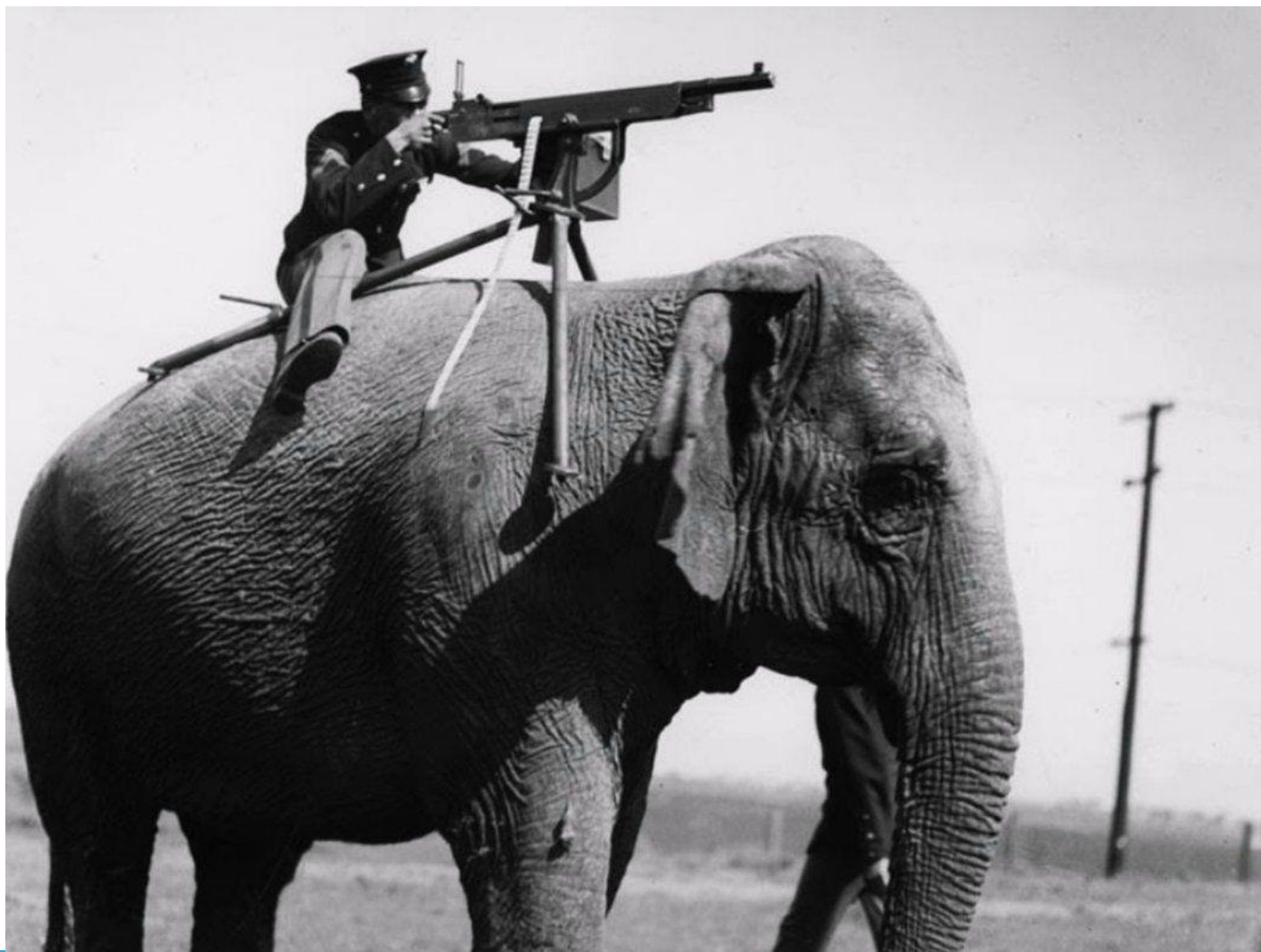
Agenda

1. **Study Background**

2. **The Military Data Requirement**

3. **DC&M Process**

4. **Mission Thread Approach**

Process

Steps	Process	Sub-process	Product	
RFI/Commander's Information Requirements **Operations Assessment Question etc.**				
Pre data-life	**Initiation**	- Identify data requirements - Select sources	Source data	
In use	**Collection**	- Create data environment - Populate data environment	Collected data	Data sharing
	Preparation	- Cleansing - Verification - Validation	Prepared data	
	Analysis		Analysis product	
Post data-life	**Archiving**	- Archiving - Preserve long term	Archived data/ operational records	
Data available for post operation analysis and future requests (e.g. historic analysis etc.)				

Initiation

- Accessibility of the data
  - Bandwidth
  - Classification
- Knowledge about the data source
  - credibility and reliability must be evaluated;
- Data Structure
  - Already structured or can be formatted and structure automatically
- Audit trail
  - Ultimately from source data to prepared data
- Timeliness
  - To understand how current any analysis will be

Data Acquisition

- The act of acquiring data from another entity, either free of charge or paid
- Acquired data, free or paid, can rarely be used without considering licensing or legal constraints
- The acquisition process is time consuming, so should be started early
  - Free data often needs sophisticated algorithms to extract and prepare
  - Proprietary data requires memorandums of understanding
  - Purchased data requires a business case
- Quality data is worth paying for, so there should be budget allocated for data acquisition

Collection and Preparation Findings

- 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**

NATO
OTAN

NORTH ATLANTIC TREATY ORGANIZATION
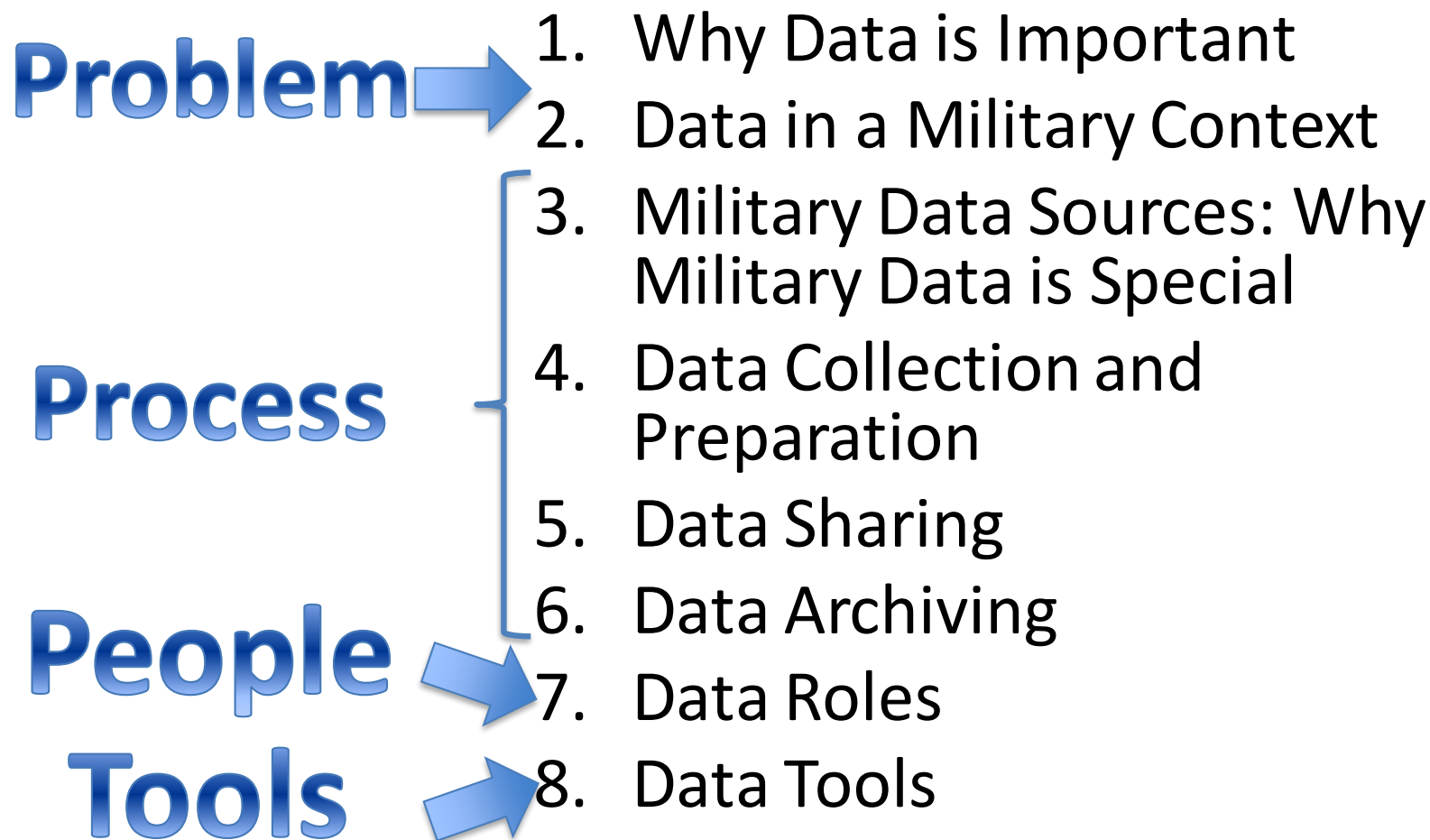SCIENCE AND TECHNOLOGY ORGANIZATION

S&T
organization

Data Sharing – The Common Thread

- Preparing data for sharing depends on who it will be shared with:
  - Non-NATO Entities and Partners not connected to NATO CIS
  - NATO Entities, Operational Units or Partners connected to NATO CIS
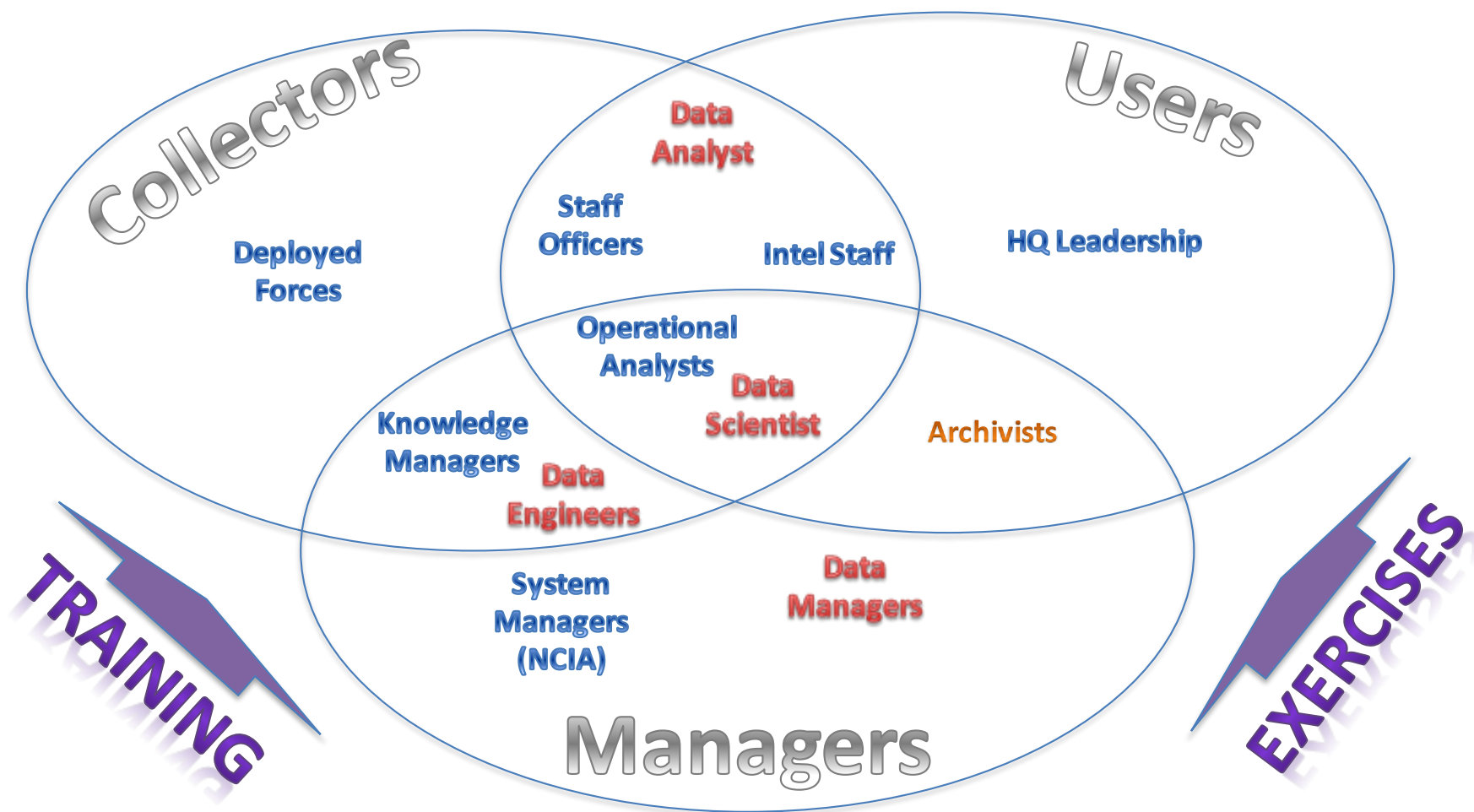 - Own HQ
  - Future Users

From Data Collection to 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.

Final Report Structure

Problem ➡

1. Why Data is Important
2. Data in a Military Context

Process

3. Military Data Sources: Why Military Data is Special
4. Data Collection and Preparation
5. Data Sharing
6. Data Archiving

People ➡
7. Data Roles

Tools ➡
8. Data Tools

Roles

Tools

Data Collection

Amazon Web Services

Google Cloud

Microsoft Azure

Data Preparation

DataPreparator

Carrot2

Elastisearch

WEKA

R

NLTK

PYTHON

RAPIDMINER

ORANGE

PANDAS

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

Agenda

1. Study Background

2. The Military Data Requirement

3. DC&M Process

4. **Mission Thread Approach**

SAS-111 - National Study "Development of a Data Collection Environment based on a Mission Thread Approach"

Heger, UniBw München

Problem Statement

- Current 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.

National Guide

- Especially 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 of that document 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

② Meta-Data Checklist

- Preserve meaning of data and context in which is 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
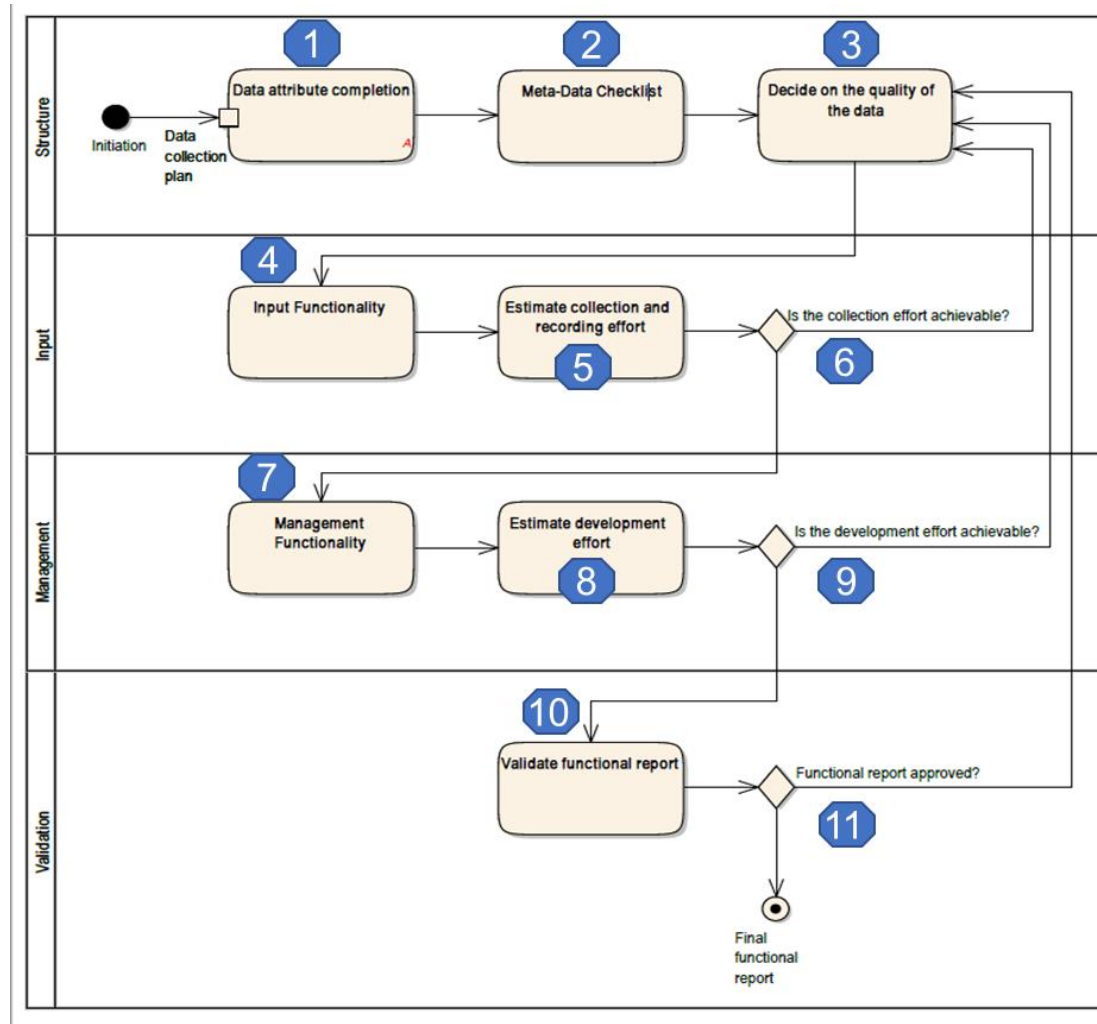
Conclusion

- 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

NATO
OTAN

NORTH ATLANTIC TREATY ORGANIZATION
SCIENCE AND TECHNOLOGY ORGANIZATION

S&T
organization

12th ANNUAL OPERATIONS RESEARCH & ANALYSIS CONFERENCE

NATO OR&A Conference
15-16 October 2018
Zagreb, Croatia

S&T
organization

Tackling Complexity and Confusion in NATO Operations in Contested and Degraded Environments

SAVE THE DATE

This year presentations from the NATO OR&A community will be built around the streams *Understanding and Operating in the* **Cyber** *Domain, Challenges and Opportunities of Rapid Reinforcement & Force Mobility and its* **Logistics** *Aspects,* and *Sense Making Through* **Analytics**. A challenging programme is being developed to include training and enlightening sessions, as well as keynotes from senior NATO military and S&T leaders.