

## **Data-Driven Joint Force Design**

**Li Jiang, Thitima Pitinanondha**

*Force Design, Strategic Capability Analysis, JOAD Division, DST Group*

*Email: [li.jiang@dst.defence.gov.au](mailto:li.jiang@dst.defence.gov.au)*

**Abstract:** Force Design has long been considered as experience-based, ad-hoc and campaign-driven activities. As a result of the First Principles Review, there has been significant organisational restructure along with a review of Defence business process, particularly the Capability Life Cycle (CLC). Force Design as a component of the CLC requires considerable development to provide the robust and repeatable processes that are needed to meet Government and Defence expectations. In order to meet these expectations, the newly instituted Force Design Cycle (FDC) and Defence Capability Assessment Program (DCAP) were established for delivering an integrated Joint Force by Design that is capable, potent and agile in achieving the strategic Defence objectives directed by the Government of Australia. There are various aspects critical for the successful execution of FDC and DCAP processes. We argue that having valid and sufficient data is one of the crucial aspects. Even though using large data to support Force Design remains challenging, it is critical to establish a systematic and enduring approach in data capturing and analytic process that contributes to the Force Design. By learning from the experiences of Force Design over the last two decades, especially drawing upon the recent experience with the HEADLIGHT17 (the centre-led Joint table-top seminar conducted to gain a deeper understanding of Gaps and Opportunities in designing Joint Force) this paper aims to report our experience, and highlight key challenges in collating and synthesising the data to provide the evidence base for Force Design decisions. The paper also introduces and discusses a data analysis approach that could improve the Force Design process, decision quality, transparency and ultimately Defence outcomes.

**Keywords:** *Force design, data architecture, data analysis, information system theory*

## 1. INTRODUCTION

Considerable research has shown that evidence-based support is critical for making informed decisions (O'Dwyer 2004, Clancy and Cronin 2005). In addition, the use of large data to support force design and future force assessment remains a challenging issue. Our worldwide literature review (see (Jiang 2015, Jiang, Tay et al. 2016)) has shown that the lack of proven best practices and techniques as well as cultural impediments on using large data to support force design needs urgent attention and solutions. We acknowledge that the newly instituted Force Design Cycle (FDC) and Defence Capability Assessment Program (DCAP) were established with the aim of supporting evidence-based policy and capability investment decisions. However, our experience with the inaugural implementation of the DCAP process, especially on the HEADLIGHT17 activity, suggests that a detailed study on data-driven force design is required to maximise data support and generate an evidence-based and highly credible outcome. We argue that building a large database with a higher level of analytical capability to support the execution of FDC and DCAP processes is not only necessary, it is also technologically feasible despite it requiring proper investments in research, training and computer technologies. In this paper, we present two focus areas of our ongoing research:

- development of a sound data architecture to assist in building an effective database in support of force design; and
- development of an effective analytical approach for analysing data from the outcomes of force assessment and the relationships between both input and output data in the DCAP process.

In this paper, we will present the progress and key focus areas of our ongoing research initiative in “Data-Driven Joint Force Design”. The rest of the paper is organised as follows: Section 2 will present a brief review of the status quo. Section 3 will present the work done in the HEADLIGHT17 as well as the progress of our ongoing research. Section 4 concludes with the significance of the research and our planned future work.

## 2. THE STATUS QUO OF THE DATA SUPPORT FOR FORCE ASSESSMENT

Force design is a complex and complicated process in its nature due to the uncertainty of rapid advance of technologies and world geopolitical dynamics. To achieve an evidence-based force design outcome, a large volume of data is required in the force design process. The following is some examples of such data:

- (1) National Defence and Security strategy and policy and data related to strategic assessments;
- (2) Military strategy<sup>1</sup> based on (1);
- (3) Data related to the current force, objective force and their unbiased assessment based on (1) and (2);
- (4) Data related to the interdependency of the capabilities of the current force and objective force;
- (5) Data related to Whole-of-Force (WoF) Balance-of-Investment (BoI) analysis;
- (6) Data related to Australian Capability Context Scenarios (ACCS) and intelligence information about the force structure and capabilities of our strategic adversaries; and
- (7) Historical data related to the lessons we learned in the force design.

However, based on literature review (see (Jiang 2015, Jiang, Tay et al. 2016)) and our experiences over the key force design processes (such as Force Appreciation, Force Structure Assessments, Force Structure Options and Choices (Thakur 2016)), we found there are a number of issues in using data in supporting and informing force design:

- (1) No effective database that can support data collection and data analytics over the entire FDC.
- (2) Historical data has not been used properly. Data obtained from previous force design and assessment cycle have never been used in the next round of FDC.
- (3) The extent of using data for force design and force structure assessment is piecemeal or limited due to the availability of data.
- (4) The assessment of the suitability of capability for the given scenario made by each participant is often from each single service perspective, lack of joint-focus approach which requires inclusive data across all services.
- (5) It is very difficult for people to have an adequate level of Integrated Investment Plan (IIP) knowledge required for the force design or force assessment due to a large number of the IIP projects in each service. This calls for a sound database and powerful tool that can support people to use IIP data whenever referencing these data is required during force assessment and force design

---

<sup>1</sup> This can include strategy for the employment of the current force and/or the future force

process. Timely data support during the force assessment process such as during the force assessment workshops is the key.

- (6) No or limited tool is available to support the entire process of elicitation, storage, design, and analysis of the qualitative data. This often leads to the inadequate credibility of the data obtained and very hard to validate the quality of the obtained qualitative data which are critical in qualitative research (Mays and Pope 2000, Hammersley 2007).
- (7) Force design and force assessment are often carried out within a short period of time. All SMEs involved from both military and civilian are often under extreme pressure; it is very difficult for these SMEs to generate a credible judgement without having a sound database system and analytic tool support.
- (8) It is difficult to generate an objective gap assessment of ADF's capability without having the data about opponents' force structure, behavior and their force options.

These problems call for a sound solution for all scientists in DST Group to work together to explore best solutions in data design and analytical space to support all analytical tasks in the FDC.

### **3. THE PROPOSED RESEARCH**

#### **3.1. Database for supporting force assessment**

In its nutshell, force design is a process that uses data/information as input from SMEs and produces a rational and sound judgement or decision about the force structure as an output. The output should be consistent, sound, and verifiable against the input data. We have discussed before that a large amount of data is needed in a force design process. For instance, we found that the following data were needed in the HEADLIGHT17:

- 1) Defence mission, which defines the strategic context for Defence capability, has a time-invariant property that links to essential parts of the White paper.
- 2) Military Strategic Objectives (MSO).
- 3) Australian Capability Context Scenarios (ACCS).
- 4) Joint War-Fighting Function ability statements (JWFAS).
- 5) The relationships between the elements in Defence Mission, MSO, ACCS, JWFAS,
- 6) Capability programs information in Integrated Investment Plan (IIP).
- 7) Capability projects information in IIP.
- 8) Existing capability information.
- 9) The information about adversaries' capability distribution which contains the information that helps formulate the likely adversaries' ORBAT that can be used in the force assessment process.
- 10) Interdependencies between existing capabilities. The information in this category can be very complicated as there are different types of interdependencies such as data, communication, schedule, functional (capability perspective), and structural dependencies. Additionally, the interdependencies are time-variant as the changes of capabilities over time. Interdependencies between programs in IIP are often required in the force assessment process.
- 11) Knowledge of the future technologies and capabilities for Defence.
- 12) Through-life costs include both the planned cost for the capabilities in IIP and sustainment cost estimation about the future capability.

In our previous research (Jiang, Tay et al. 2016), we have proposed an architecture of the force design. Fig. 1 shows a higher level categorisation of the architecture and the complexity of the data for force design. In the latest task on supporting the HEADLIGHT17, we have developed a database with the aim to examine the merits of using data from a rigorously designed database in supporting force design. We argue that a sound database structure is the first step toward obtaining a higher quality data, which is critical in supporting decision making in force assessment process, this is especially the case for the HEADLIGHT17. The database structure designed for the task is illustrated in Fig. 2. The design and employment of the database in the task are an experimentation of our approach in supporting the DCAP process. The database in Fig. 2 contains data from several categories from the proposed data architecture shown in Fig. 1: Force-in-being, Objective Force, and Force Concepts. A number of factors were considered in the design of the database structure including:

- the data structure should provide a unified and consistent representation of the data used in the force design process;
- the fundamental principles of the database design (Mittra 1990, Golfarelli and Rizzi 2009, Özsü and Valduriez 2011) should be considered such as reducing the redundancy of the data, improving the integrity of the data etc; and

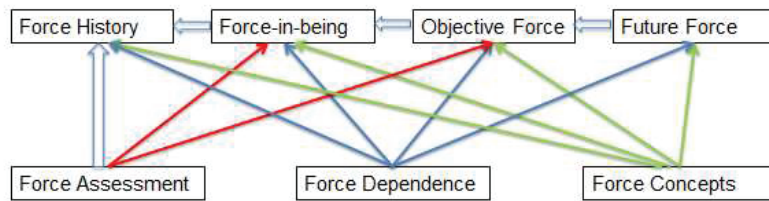


Figure 1. Higher Level Data Architecture for Supporting Force Design

Note: The double-line arrow indicate the data evolution, single-line arrow represents the relationships between different data categories. The colors of arrows illustrate the different data categories.

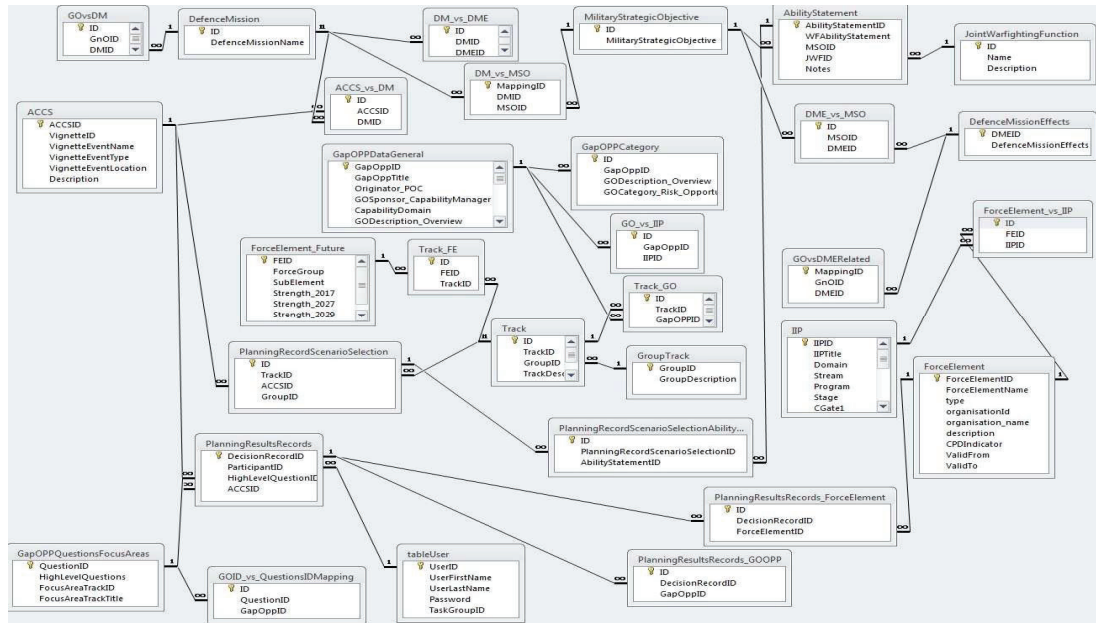


Figure 2. A Structure of Database Supporting Force Exploration Phase of the DCAP through HEADLIGHT17

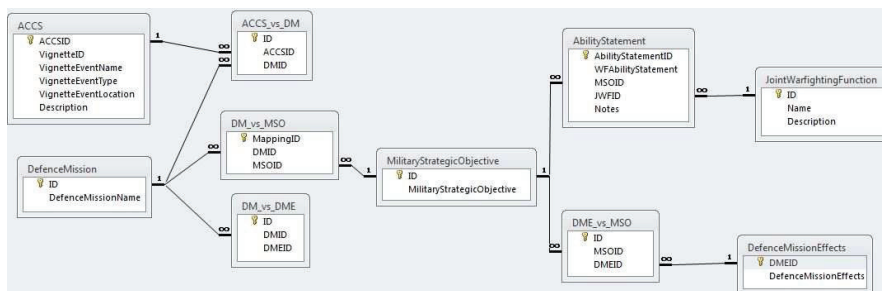
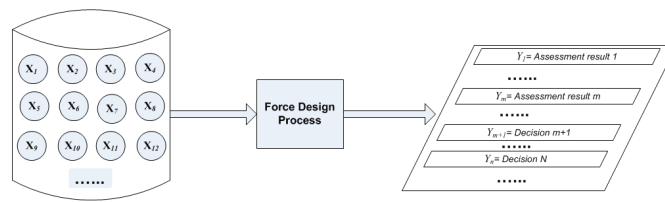


Figure 3. A Strategic Guidance Database

- the data in the different categories can be easily integrated and formulated into cohesive information required for force design.

Fig. 3 illustrates an extraction of the Strategic Guidance data structure designed for the tasks of supporting HEADLIGHT17. In the task with the data in the database, we were able to observe and examine how the data are used and frequency of the data used in the process. As can be seen from Fig. 2, the database is not a trivial one. It can provide effective evidence-based support if it is fully implemented and utilised. Due to the time constraint of the task, the data set in the database we designed is still relatively small. However, the database and application we developed did provide effective support for the HEADLIGHT17 despite it was still in its earlier stage of the development. The database that is able to support the entire FDC and DCAP process can be much more complicated, which requires dedicated and sound technical solutions and investments on both hardware and software. Despite the complexity and cost, we argue that it is the only way we can produce a sound solution for the force design and building the 21st century ADF.



**Figure 4.** Data input/output of force design process

Note: Each  $X_i$  represents a specific input dataset

### 3.2. Using Information Theory to Analyse Relationship between Data Used in the Force Design

#### 3.2.1 The problem

Reviewing the tasks supporting force structure assessment activities and force design processes over the last two decades, especially in the HEADLIGHT 17, we found that analytical capabilities played increasingly critical roles in helping the decision-making process. However, having a clear understanding of the linkage between the decisions being made and the source of the data that support generating these decisions are critical, yet remains a challenging issue. For instance, it is unclear what types of data actually provide critical information in the decision-making process; what data are redundant or not useful at all; which data are used more often than others; how to identify and link these data sets to the final decisions such that we can understand the rationale of the decision and validate the fidelity of the data from the output side.

In developing an analytical approach, we are utilising the Information Theory (MacKay 2003) and Entropy Theory Of Mind (Chen 2015) to investigate a method to improve the quality of the output of subject matter experts (SMEs) by linking the output to the input which is the evidence of the former. The idea of the research is to frame the force design process as an input and output process as illustrated in the Fig. 4. We contend that:

- the quality of output is correlated to the quality of the input data; and
- the larger, more structural, and less uncertainty the input data sets are, the more reliable and trustworthy of the output can be.

In Fig. 4, each bubble represents data in different categories in the force design database; each parallelogram represents the output of the force design process. In reality,  $X_i$  might represent a specific input data set such as Strategic Data, Existing Capabilities Data, and Capability Project Data in IIP, Capability Program Data in IIP, Capability Interdependency Data, and ACCS etc. We assume the database contains sufficient information that can help SMEs making credible decisions. More specifically, for each  $Y \in \mathbf{Y} = \{Y_1, \dots, Y_m\}$ , there are a set of data in  $X'$ , such that  $X' \subset X = \{X_1, \dots, X_n\}$  and  $X'$  provide sufficient information for SMEs to make sound decisions  $Y$  in the force structure assessment process. We argue that having this type of knowledge and information is critical for understanding the rationale and validity of the decision being made. The problem is that **how do we find such a data set?** We have been exploring a credible approach and a number of measures that helps to find such a data set  $X'$  over the last two years. In this research, we have proposed using information theory to help to partition the data being used in the database based on the frequency of the data being used in the force design and assessment process. Following two sections will present our current progress via a simple example.

#### 3.2.2 Rationale of Data Partition

In this research, we propose using the Entropy analysis approach from information theory to help to partition data being used in the force design/assessment process. For instance, SMEs want to look at the information about the “Joint Strike Fighter (JSF)” and if “JSF” has “Sensor Fusion capability”. The data partition based on the criteria “JSF” and “Sensor Fusion capability” in a large database can be considered as an entropy reduction process, whereby the retrieved data items or data set based on the SMEs’ enquiry can be stored as an intermediate database created for the temporary usage over the FDC.

We can model SMEs retrieving the required data as a random event and the frequency of the using certain data item or a data set in a large database are modelled as a probability of using the data. More formally:

- Let  $a$  be a discrete random variable, let  $a \in A = \{a_1, a_2, \dots, a_n\}$ .  $a_i$  represents a random event such as retrieving/browsing a specific record from a database.
- $n$  is the number of variables in the set  $A$ .

- Each  $a_i$  with the probability of  $p_i \in P = \{p_1, p_2, \dots, p_n\}$  which represents the frequency of data item that is retrieved from the database.

We can use  $H(X) = \sum_{i=1}^n p_i \log(p_i)$  to represent the Entropy in each specific data set (such as a table) in a database at the given specific time. An illustration of using probability distribution of data (being used) to partition a table to reduce entropy to find the required information is shown in Section 3.2.2. With the idea discussed above, we propose the following approach to resolving the problem discussed in section 3.2.1:

Step 1: For the database system: Develop a mechanism that can record the usage of each data item in the database.

Step 2: Linking input data to the output data:

- 1) For each of the “output decision”  $Y$ , suppose  $S$  are the key words occurs in  $Y$ , find the data set  $Z \subset X$  that contains  $S$ . (the number of occurrences of the searched information in the data items is not considered for simplicity).
- 2) Partition the data according to the probability distribution of the data usage using Entropy as a measure in the partition process.
- 3) Derive the data set  $X' \subset Z$  which is the data set containing higher probability distribution of the data usage based on Step 1.  $X'$  is the likely dataset that supports making the decision  $Y$ .

We argue it is significant to study and develop an effective mechanism that helps to identify highly relevant data used for a force assessment process and identify the roles of the data in the FD decision-making process. Furthermore, statistical modelling can help to reveal the inherent linkage between the data that are used and the force assessment results, as well as FD decisions made in the process. This will, in turn, help provide evidence-based support in the FD process.

### 3.2.3 An Example

We assumed the decision  $Y$  contains information  $S$  about  $S$ =“JSF should have Sensor Fusion capability”. Using the proposed method discussed above, we can carry out following activities:

- 1) Search the data in the database that contains the information about “JSF” and/or “Sensor Fusion capability” and store the retrieve data set in an intermediate table as shown in the first column of Table 1 below.  $Z$ =data set in Table 1 (Using a1 to a5 and b1 to b5 to represent the data items that contain the searched information “JSF” and/or “Sensor Fusion capability” for the simplicity of presentation ).
- 2) Partition the table contained “JSF” and/or “Sensor Fusion capability” into two datasets based on Entropy measure as discussed in the Section 3.2.2. The result of the partition is shown in Table 2 and 3 respectively. We assumed this is the final partition (The partitions can continue until the Entropy value cannot be reduced.).
- 3) Find the data items that contains “JSF” and/or “Sensor Fusion capability” and were used with highest probability distribution of use by SMEs (assumed that they had searched the information about the “JSF” and “Sensor Fusion capability” during the force assessment process) and store the

**Table 1.** Original Data

Data Item	Number of times used by SMEs	Probability Distribution
a1	2	0.011
a2	3	0.016
b1	15	0.082
b2	10	0.055
a3	4	0.022
a4	5	0.027
b5	20	0.109
b3	18	0.098
b4	100	0.546
a5	6	0.033
Entropy		0.684

**Table 2.** Part Of Split Table 1 With Higher Probability Distribution

Data Item	Number of times used by SMEs	Probability Distribution
b1	15	0.092
b2	10	0.061
b3	18	0.110
b4	100	0.613
b5	20	0.123
Entropy		0.517

**Table 3.** Part of split table 1 with Lower Probability Distribution

Data Item	Number of times used by SMEs	Probability Distribution
a1	2	0.100
a2	3	0.150
a3	4	0.200
a4	5	0.250
a5	6	0.300
Entropy		0.671

The combined entropy of the two tables is: 0.594

**Notes:** Table 1 shows the original data set with Entropy 0684. Table 2 and Table 3 are derived from Table 1 with entropy of 0.517 and 0.671 respectively. The combined entropy of the two tables is 0.594. The entropy gain is:  $0.684-0.594=0.09$

data set which is shown in the second column of Table 1 below.  $X'$  = data set in Table 2.

We can further examine the data items in Table 2 to determine which data items provide sufficient information for the decision  $Y'$ . There are a number of merits of this type of data partition in supporting force design. For example, it can:

- help to understand what data are the most valuable data in the decision making process;
- help to link the data to the decision making process as well as the decision, assessment results or conclusion that are derived from the analytical process;
- provide rich information about database restructuring in a large database system to improve the performance of the database as the queries made into the restructured database can be much faster and efficient; and
- help to analyse the linkage between input and output more effectively if the semantic level analysis for the textual data function is available. The support will be very effective and efficient for validating the outcome of the force assessment and force design. Therefore, it can help to produce a sound decision for the ADF in the complicated force design process.

#### 4. CONCLUSION

In this paper, we have presented our research and experience on using data to support the DCAP process. Developing rigorous data architecture and building advanced analytic capability is critical for obtaining sound results in the FDC. Our approach is still subjected to further research and validation in the future tasks in the FDC and the DCAP process. Our future work will continue research and implement our proposed approaches discussed in the paper. More specifically, we will focus on:

- (1) Implementing a prototype of the proposed data model that enables the collation, maintenance and management of Force Design data in a format that is suitable for near-term migration into the Force Design Body of Knowledge
- (2) Developing a set of measures and techniques to analyse the data obtained in the force design process. We will continue using apply Entropy Analysis Techniques (MacKay 2003) and methods in the Entropy Theory of Mind (Chen 2015) to analyse the collective entropy on different data sets from both input and output data.
- (3) Comparing data or computing correlations between the data sets from input and output
- (4) Studying an effective mechanism that can help preparing sufficient and effective data on the input side; which, in turn, can help reduce the decision bias in the force design process.
- (5) Once a set of results have been collected and analysed over several FDCs, the data obtained from the analysis can be used to build a simulation model. If successful, the simulation model can be used to identify which data sets provide better support for force design; thus helping to improve our understanding of force design process.

#### REFERENCES

- Chen, J. (2015). *The unity of science and economics: A new foundation of economic theory*, Springer.
- Clancy, C. M. and K. Cronin (2005). "Evidence-based decision making: global evidence, local decisions." *Health affairs* **24**(1): 151-162.
- Golfarelli, M. and S. Rizzi (2009). *Data warehouse design: Modern principles and methodologies*, McGraw-Hill, Inc.
- Hammersley, M. (2007). "The issue of quality in qualitative research." *International Journal of Research & Method in Education* **30**(3): 287-305.
- Jiang, L., N. Tay and G. Bulluss (2016). Exploring A Feasible Data Architecture For Supporting Force Design. *Defence Operations Research Symposium (DORS 2016)*. Canberra.
- Jiang, L. T., Nicholas; Seif Zadeh, Hossein; Bulluss, Gary (2015). *Towards Defence Strategic Data Planning*. 21st International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, Gold Coast, Queensland, Australia.
- MacKay, D. J. (2003). *Information theory, inference and learning algorithms*, Cambridge university press.
- Mays, N. and C. Pope (2000). "Qualitative research in health care: Assessing quality in qualitative research." *BMJ: British Medical Journal* **320**(7226): 50.
- Mittra, S. S. (1990). *Principles of Relational Database Design*, Prentice Hall PTR.
- O'Dwyer, L. (2004). "A critical review of evidence-based policy making." *Australian Housing and Urban Research Institute* **31**: 32.
- Özsu, M. T. and P. Valduriez (2011). *Principles of distributed database systems*, Springer Science & Business Media.
- Thakur, N. (2016). *Scoping an Enduring Force Design Capability*, Draft Technical Report. Defence Science and Technology Group. Department of Defence.