



SHOAL

Aligning enterprise data quality with Defence needs

Luan Dinh and John Wharington

The information problem

- Military intelligence gathering in the modern information age faces both a **quantity** and **quality** problem
- Data fragments forming credible information come in all forms
 - OEM technical data
 - Marketing brochures
 - Press releases
 - Janes Database and publications

Passengers (Two-class seating)				Freighters			
	No. of Seats	Range (nm)	Range (km)		Revenue payload (Metric Tonnage)	Range (nm)	Range (km)
747-8	475	7,345	13,600	747-8F	137.7	4,100	7,600
777-9	400-425	7,800	14,075	777F	102.0	4,970	9,200
777-300ER	396	7,370	13,650	777-200ER BCF**	75-80	3,900-4,100	7,200-7,800
777-3	350-375	8,700	16,110	767-300F	52.5	3,255	6,025
787-10	330	6,430	11,910	767-300BCF	51.7	3,305	6,120
777-200LR	317	8,555	15,840	737-800BCF	22.7	1,995	3,700
787-9	290	7,635	14,140	737-700C	18.5	3,075	5,695
787-8	242	7,355	13,620				
737 MAX 9	178	3,515	6,510	* One-class seating ** Product development study BCF = Boeing Converted Freighter ER = Extended Range LR = Long Range			
737-900ER	180	2,850	5,470				
737 MAX 200*	200	2,700	5,000				
737 MAX 8	162	3,515	6,510				
737-800	162	2,935	5,440				
737 MAX 7	138	3,825	7,085				
737-700	126	3,010	5,570				

For more information on the Boeing family of airplanes, please visit our website boeing.com

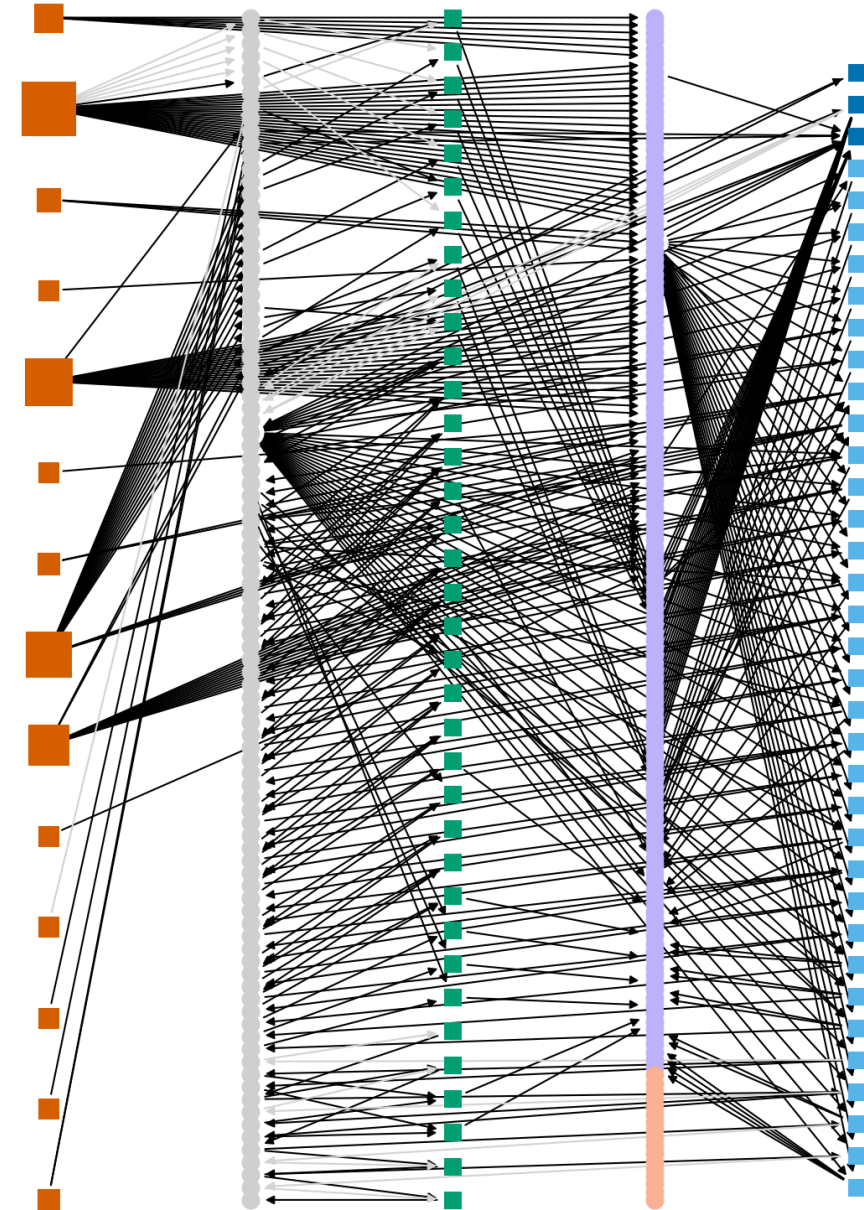


NOVEMBER 2016



Data trustworthiness

- Data fragment may:
 - Be duplicated
 - Have unit conversion or rounding errors
 - Erroneously transcribed
- How does one reconcile these disparate data fragments of varying *quality* and trustworthiness in a unified system?
- And do so in a **systematic, reproducible** and **auditable** way?
- Data **provenance** and **traceability** across handling steps, from ingestion to output



Enterprise Data Quality dimensions

Data Quality Dimension	Definition
Authority	Degree to which the data source is trustworthy and credible. Primary sources have higher authority than secondary or tertiary sources that derive data from others.
Completeness	Degree to which all data required for a particular use is present and available for use. Components to this dimension include the rules defining the conditions under which missing values are acceptable or unacceptable, similarly determining what data is critical and what is optional.
Consistency	Degree to which data sets and data models are consistent across instances within an organization, functional area, or community of interest. Values of data should not contradict other values representing the same entity, and should not contradict data in another data set.
Specificity	Degree to which data has been rounded or aggregated. Describes the level of detail or precision captured by the data.
Timeliness	Degree to which data values are up-to-date and current, or if there are latencies and lags in reporting.
Validity	Degree to which data conforms to predefined standards, business rules, or constraints.

NATO & Defence Data Quality dimensions

NATO (2025, August 29). Data Quality Framework for the Alliance

Data Quality Dimension	Definition and Measure
Accuracy	Degree to which data values are sufficiently representative of the real-world objects they intend to describe.
Completeness	Degree to which all data required for a particular use is present and available for use. Components to this dimension include the rules defining the conditions under which missing values are acceptable or unacceptable, similarly determining what data is critical and what is optional.
Uniqueness	Degree to which data sets and data models are consistent across instances within an organization, functional area, or community of interest. Values of data should not contradict other values representing the same entity, and should not contradict data in another data set.
Timeliness	Degree to which data values are up-to-date and current, or if there are latencies and lags in reporting.
Validity	Degree to which data conforms to predefined standards, business rules, or constraints.
Granularity	Degree to which data has been rounder or aggregated. Describes the level of detail or precision captured by the data.

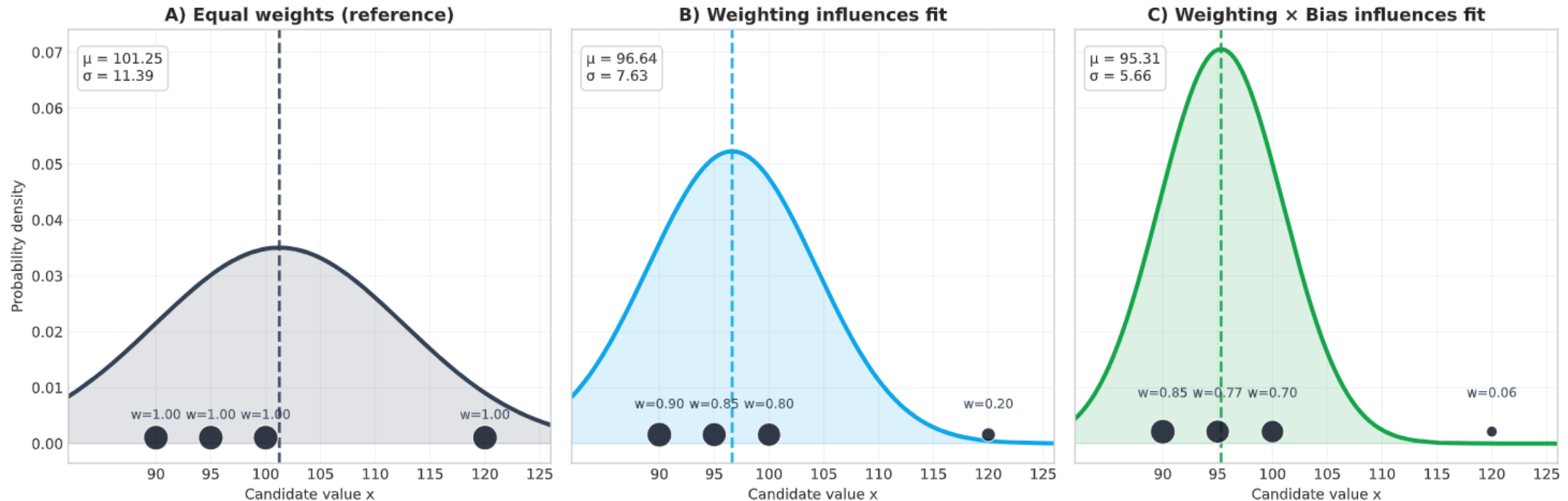
How data quality dimensions shape confidence

- The Data Quality dimensions capture two key concepts - **source credibility** vs **data value reliability**

$$B = \sum_{i \in \{C,T,A\}} w_i s_i$$

$$W = \sum_{i \in \{S,V,K\}} w_i f_i$$

Confidence = Bias + Weighting



Analytic Hierarchy Process

- Developed in the 1970s by Operations Researcher Thomas Saaty
- A structured methodology for eliciting expert judgement from communities of interest
- Converting qualitative preferences into defensible quantitative weights

	C1	C2	C3	C4	C5	C6
C1		1/3	1/5	3	1/2	4
C2	3		1/2	5	2	6
C3	5	2		4	3	7
C4	1/3	1/5	1/4		1/2	3
C5	2	1/2	1/3	2		5
C6	1/4	1/6	1/7	1/3	1/5	

Summary and forward-looking

Mission-Driven

- Quantitative system for automated data handling of open-source information.
- Address the data integrity, traceability and provenance for assured decision making

Next-steps

- AI-supported data source verification and scoring
- Share and workshop with other community of interests

SHOAL™



107 WRIGHT STREET, ADELAIDE SA 5000

ADELAIDE | BRISBANE | CANBERRA | SYDNEY | AUCKLAND



+61 2 6239 4288



support@shoalgroup.com



shoalgroup.com



Shoal Group Pty Ltd

