Converting Your Legacy Data to S1000D

Naveh Greenberg,
Director, US Defense Development
Data Conversion Laboratory (DCL)

Agenda

- A very quick introduction to S1000D conversions

- What the technical headaches are

- Whether to convert or rewrite

- Planning for a good conversion experience

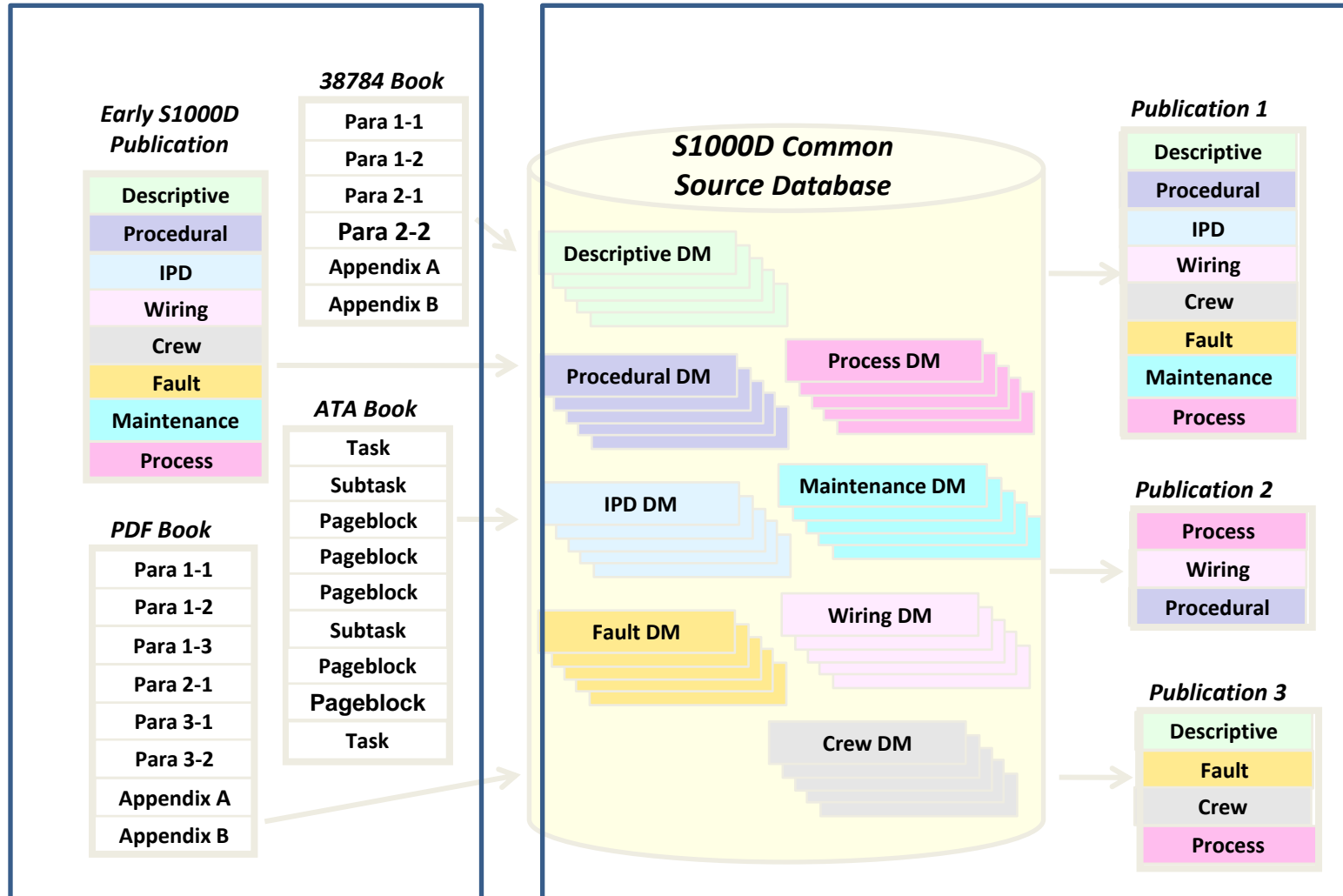- What the timeline looks like

- Some tools to help

The Story with S1000D

- Provides an international solution to standardize on a common spec across the services

- Was developed from ATA Spec 100 to place greater emphasis on granularity and interchangeability of information

- Allows for modularization of content with Data Modules (DM), and for easy re-use in multiple outputs

- Standard and schemas are publicly available

- *Applicability* is available at a very fine level

- Standard format for external data exchange (Data Dispatch Notes)

- *Process DM* to allow interactive processing structures and *Learning DM* for training (4.0)

What Makes S1000D Conversion Difficult

- S1000D is a conceptual departure from linear information
  – and is difficult for many to get used to

- Turns the traditional book into a collection of DMs

- DMC & business rules.

- DMs can be thought of as interchangeable parts
  - to be reassembled in multiple ways
  - to be repurposed for multiple outputs
  - to be reused across multiple products

- …but your documents weren't likely to have been designed to do this.

Structuring a Book into Data Modules in S1000D

Early S1000D Publication

Descriptive
Procedural
IPD
Wiring
Crew
Fault
Maintenance
Process

38784 Book

Para 1-1
Para 1-2
Para 2-1
Para 2-2
Appendix A
Appendix B

ATA Book

Task
Subtask
Pageblock
Pageblock
Pageblock
Subtask
Pageblock
Pageblock
Task

PDF Book

Para 1-1
Para 1-2
Para 1-3
Para 2-1
Para 3-1
Para 3-2
Appendix A
Appendix B

S1000D Common Source Database

- Descriptive DM
- Procedural DM
- Process DM
- IPD DM
- Maintenance DM
- Wiring DM
- Fault DM
- Crew DM

Publication 1

Descriptive
Procedural
IPD
Wiring
Crew
Fault
Maintenance
Process

Publication 2

Process
Wiring
Procedural

Publication 3

Descriptive
Fault
Crew
Process

Further Complications in S1000D Conversion

- **There's the usual conversion issues**
  - Accuracy of the transferred text
  - Tables
  - Math
  - Special Characters

- **There's also the structuring issues**
  - Identifying DMs
  - Identifying reusable content
  - Identifying Applicability

- **And the people issues**
  - Deciding what needs re-authoring
  - Getting used to a new "document" paradigm
  - Getting rugged individualists to collaborate more

Some Examples of Converted Data

Bike Sample ≠ Your Legacy Data

Square Peg 1: Procedure Authored as a Table

Issue:

Tasks are done as tables rather than numbered lists. Automated conversion keeps the tables as tables, and steps are not tagged as steps. This is a big problem for Information Mapping users (standard layout for steps).

Step	Action
1	Check the phase sequence in the mains with the phase sequence indicator.
2	Check the direction of rotation of impeller. If a dry installation is made, check the direction of rotation through the inlet elbow access cover.

Note: If done consistently, tables can be deconstructed programmatically.

Square Peg 2: Multiple Steps in a Single Procedure

Replacing an XYZ Module
Use this procedure to replace an XYZ module

Remove XYZ Module
1. Loosen the screws.
2. Disengage the ejectors
3. Pull the module straight out

Insert Replacement XYZ Module
1. Align the module.
2. Insert the module, pressing in firmly
3. Engage the ejectors

Square Peg 3: Introductory Material That Applies to Numerous DMs

Example:

1-1 Introduction

The following paragraph discuss servicing…

Personnel Recommended:

One	1-2
Two	1-3

[[start]]

WARNING

Make sure XYZ module is not connected…..

1-2 Remove XYZ Module

1. Loosen the screws.

2. Disengage the ejectors

1-3 Insert Replacement XYZ Module

1. Align the module.

2. Insert the module, pressing in firmly

[[stop]]

Square Peg 4: Irrelevant Cross-References

Conversion to S1000D may make some source cross-references irrelevant. For example, assuming some chapter headings are dropped, a reference to a chapter is no longer valid.

See Chapter 1, Introduction on page 2

Would be tagged as:

See <dmRef><dmRefIdent><dmCode modelIdentCode=".." systemDiffCode="A" . . . infoCodeVariant="A" itemLocationCode="D"/></dmRefIdent></dmRef>

NOTE: Hard-keyed page numbers are typically dropped from the cross-reference string since they are no longer relevant in S1000D. A project specific business rule will need to me made as to how to handle these links.

Square Peg 5: Writer Creativity in Source Material

<para>1. Clean the Engine.</para>

<step1><para>Clean the Engine.</para></step1>

<seqlist><item>Clean the Engine.</item></seqlist>

<entry>1.</entry><entry>Clean the Engine. </entry>

■ Install the system in an open rack whenever possible. If installation in an enclosed rack is unavoidable, ensure that the rack has adequate ventilation.

■ Maintain ambient airflow to ensure normal operation. If the airflow is blocked or restricted, or if the intake air is too warm, an over temperature condition can occur.

ITEM NO.	PART NUMBER	DESCRIPTION 1 2 3 4 5 6 7	UNITS PER ASSY	USE ON CODE	SM&R CODE
19	C811290-001	● ● FRONT PANEL, BLOWER	1		PAOZZ
20	C983586-001	● ● BLOWER, MODIFIED	2		PAODD
21	C983586-002	● ● FAN GUARD .... (ATTACHING PARTS FOR EACH OF ITEM NUMBERS 19, 20, 21)	2		PAOZZ
22	MZ53557-34	● ● SCREW, PAN HEAD, CROSS-	4		PAOZZ

Square Peg 5: Writer Creativity in Source Material

ITEM NO.	PART NUMBER	DESCRIPTION 1 2 3 4 5 6 7	UNITS PER ASSY	USE ON CODE	SM&R CODE
-11	M664688-C48	WASHER, FLAT, CRES, 0.219 ID X .... 0.438 OD X 0.049 T INCH	6		PADZZ
		• SCREW, FLAT HEAD, CROSS- .... RECESSED, CRES, 0.164-32 X 0.375 L INCH /NOTE 2/			
		OR			
	M661987-40	• SCREW, PAN HEAD, CROSS-RECESSED, ...	8		PADZZ

Specification	→	Min. Yld. Strength →	Min. Ten. Strength¶
SA-537, Cl. 2	→	60,000 psi → →	→ 80,000 psi¶
ASTM A 572, Gr. 60	→	60,000 psi	5,000 psi¶
ASTM A 572, Gr. 65	→	65,000 psi	→ 80,000 psi¶
ASTM A 633, Gr. E	→	60,000 psi	→ 80,000 psi¶

So… Maybe You Shouldn't Bother Converting Your Content?

- It seems like such a pain.

- Too many rules.

- There is always a need for some rewriting - few writers have the clairvoyance to author content with the intent that be converted in the future – might as well rewrite it all.

- My writers aren't very busy right now anyway.

- It's more fun and seems like less trouble to author anew.

Think Again….

- If you rewrite you still need to fulfil the standard requirements.

- Still need some XML knowledge.

- Can take a very long time. For example a 5000 page project:
  - 15 minutes per page to rewrite.
  - 75,000 minutes => 1250 hours
  - At 7 hours per day (no breaks) => **179 days**
- But worse …
  - Inconsistencies.
  - Writer creativity.
  - IETM/CMS may not properly function
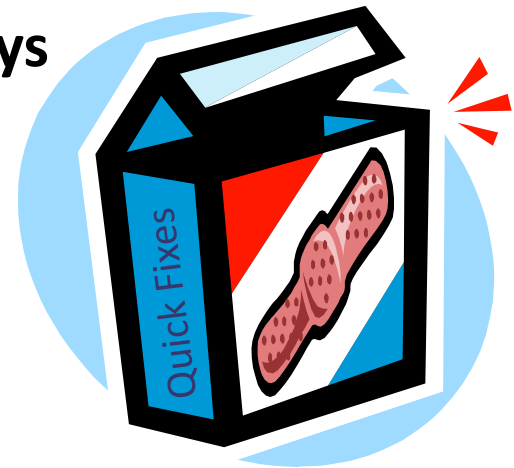
In Reality… Converting Your Content is Worth the Bother

- **Throwing it out and starting over is an expensive option**
  - Rewriting at $25-$45+/page vs. converting to S1000D at $6-$12/page
  - The hidden costs of redoing index entries, links and other features you've built in
  - The hidden cost of reviewing and recertifying it all

- **It's usually easier to use what you have as a base, and convert over**
  - Needs planning
  - Needs time

- **Planning for a good conversion experience**
  - Which content will you need?
  - Which content is worth converting?
  - Which content is suitable for re-use in multiple places?
  - What tools are available?
  - How to specify the conversion to get it right?
  - When do you start all this planning?

Most Importantly – Plan!!!

"If I had eight hours to chop down a tree, I'd spend six sharpening my ax."
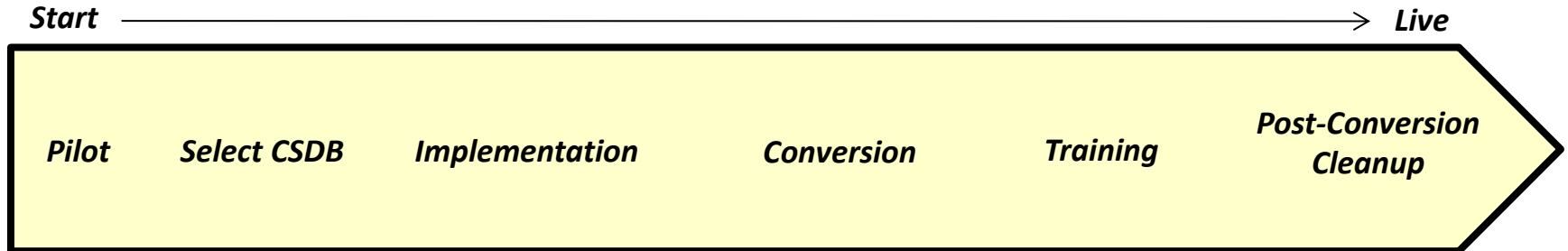
- **Ask the important initial questions**
  - Who are the stakeholders. Who is the final client/user?
  - What is the estimated volume and deadline?
  - What version of S1000D?
  - Do we know what CMS or rendering tools will be used?
  - Source format. Not all source data are created equal.
  - Budget?
- **Ask around or join discussion groups.**
- **Prepare for the next step**
  - Get your hands on the source data, business rules, and schemas.
  - Begin looking for the right people. You don't need to be a S1000D savvy but you do at a minimum understand the concept.
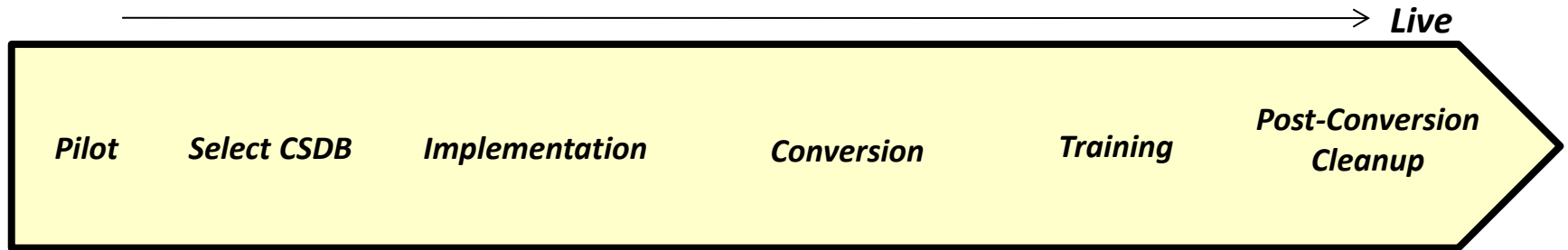
What to Convert, and in What Order

- **Categorizing**
  - Active documents in good shape
  - Active documents that need a lot of work
  - Somewhat inactive document that will likely be retired
  - Archival materials

- **Prioritizing**
  - Documents that are most used
  - Documents that are customer favorites
  - Documents with longest product life
  - Start with most recent documents and go back

- **Identifying the process**
  - Can be converted as is
  - Can be converted with some work
  - Needs to be rewritten
  - Don't convert – just keep archival copies

The Classic Conversion Process

Start →→→→→→→→→→→→→→→→→→→→→→→→→→→→→→→ *Live*

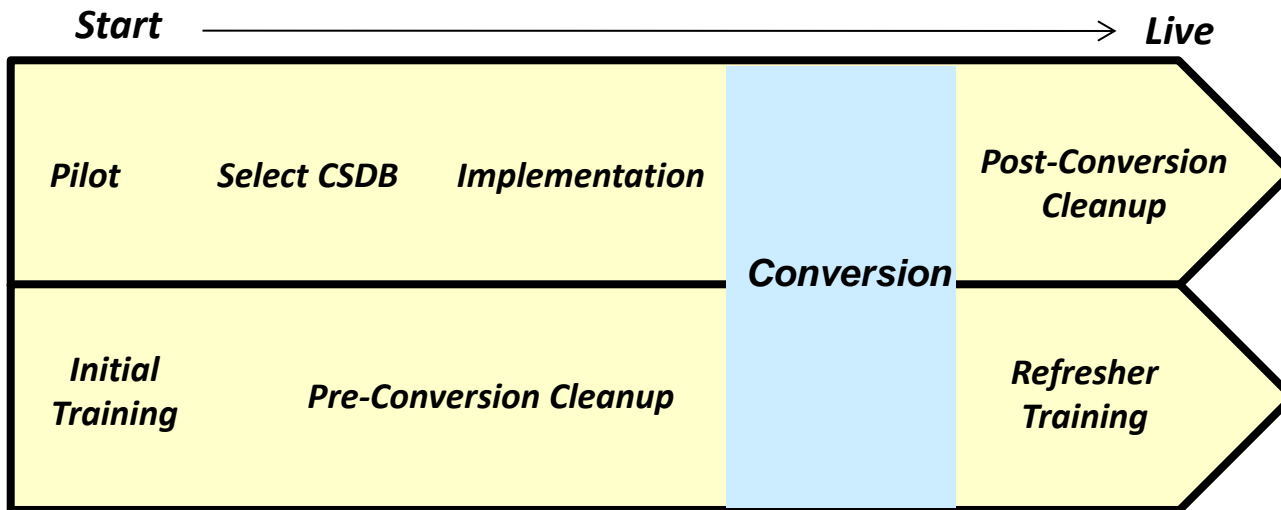| *Pilot* | *Select CSDB* | *Implementation* | *Conversion* | *Training* | *Post-Conversion Cleanup* |

- Conversion of legacy data is usually an afterthought – a lot of scampering after the system is up and running

- 3 years has been typical for DITA implementations, with another 1-2 years to really get it going; S1000D is longer

The Classic Conversion Process

Live →

| Pilot | Select CSDB | Implementation | Conversion | Training | Post-Conversion Cleanup |

Doing Cleanup and Implementation in Parallel Tracks:

Start → Live

| Pilot | Select CSDB | Implementation | | Post-Conversion Cleanup |

Conversion

| Initial Training | Pre-Conversion Cleanup | | Refresher Training |

Some Tools for a Good Conversion Experience

- ***Conversion specifications*** – a formal process to specify the conversion so that all stakeholders can agree on a proper process

- ***DMRL (Data Module Requirements List*)** - to verify that Data Modules have been properly identified

- ***Content reuse identification*** – software to located potentially reusable content

- ***Best practice scanners*** – to scan code to verify best practices have been followed

- ***Test conversions into S1000D*** – to make sure the process works all the way through

Data Module Requirement List (DMRL)

Para No	Original Title	DMC	Type	Modelic
	FOREWORD	SAMPLESYSTEM-A-00-00-00-00A-018A-D	Desc	SAMPLESYSTEM
1	INTRODUCTION.	-------		
2	MANUAL ORGANIZATION.	-------		
CHAPTER 3	PREPARATION FOR USE AND RESHIPMENT	NA		
SECTION I	PREPARATION FOR USE	SAMPLESYSTEM-A-00-00-00-00A-006A-A	Desc	SAMPLESYSTEM
3.1	INTRODUCTION.	-------		
3.1.1	Nomenclature Setup and Initialization Instructions.	SAMPLESYSTEM-A-00-00-00-00A-007A-A	proc	SAMPLESYSTEM
SECTION II	PREPARATION FOR RESHIPMENT	SAMPLESYSTEM-A-00-00-00-00A-530A-A	Desc	SAMPLESYSTEM
3.2	INTRODUCTION.	-------		
3.3	CONDITIONS AND METHODS OF RESHIPMENT.	-------		
CHAPTER 8	ILLUSTRATED PARTS BREAKDOWN	SAMPLESYSTEM-A-00-00-00-00A-018B-D	Desc	SAMPLESYSTEM
SECTION I	INTRODUCTION	NA		
8.1	INTRODUCTION.	-------		
8.2	MAINTENANCE PARTS LIST.	-------		
SECTION II	MAINTENANCE PARTS LIST	SAMPLESYSTEM-A-00-00-00-00A-075A-D	IPB	SAMPLESYSTEM
SECTION III	NUMERICAL INDEX	N/A		
SECTION IV	REFERENCE DESIGNATOR INDEX	N/A		
	GLOSSARY	SAMPLESYSTEM-A-00-00-00-00A-006A-D	Desc	SAMPLESYSTEM

Divide and Conquer for Effective Content Reuse

- Content reuse frequently showed as a reason for moving to S1000D

- However few had yet implemented; it was just too daunting a task

- Divide and conquer the identification of reusable content

- Use tools that help to identify reusable content

- Create an iterative process to manage content reuse

Content Reuse Analysis Reports

- Finding exact or similar text will help you when mapping to Data Modules

A17-1_2010_U002_025 line: 2054

engineering test : a test carried out by or witnessed by a registered or licensed professional engineer, testing laboratory, or certifying organization to ensure conformance to Code requirements.

\<Exact matches\>

A17-1_CSA_B44_Edtn_2013_000505_U100_2013_015 line: 2164

A17-1_Handbook_Edtn_2010_032114_000501_u004_2010_033 line: 19514

A17-3_Edtn_2011_000502_U001_2011_045 line: 7406

- It will also help to detect applicability and inconsistencies

AG_1b_U004_054 line: 14401

FD-5220 OTHER WELDS Fillet welds, butt welds, and seal welds shall be visually inspected in accordance with AA-6331.

FD-5220 {FH-5220} OTHER WELDS Fillet welds, butt welds, and seal welds shall be visually inspected in accordance with AA-6331.

FD-5220 FH-5220 OTHER WELDS Fillet welds, butt welds, and seal welds shall be visually inspected in accordance with AA-6331.

AG_1b_U004_054 line: 22357

Summary

- **Conversion of legacy data to S1000D is not trivial**
  - Significantly faster, safer and less expensive that rewriting
  - There are special S1000D consideration to be taken into account
  - Needs planning

- **Starting the conversion process earlier makes life easier**
  - Shifts the critical path; speeds the process
  - More of the work can be done by the content owners
  - Eases the training and change acceptance burdens

- **Content re-use also needs planning**
  - Starting early allows time for a more through review
  - Setting up collaborative teams sets the tone and allows one to "divide and conquer"

Q&A

Naveh Greenberg
Director, US Defense Development
Data Conversion Laboratory (DCL)
(718) 307-5758
ngreenberg@dclab.com
@dclaboratory