



DCL™

Data Conversion
Laboratory Inc.

Using Automation for S1000D Conversion and QA

Naveh Greenberg,
Director, U.S. Defense Development,
Data Conversion Laboratory

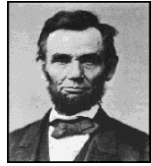
Agenda

- Is Fully Automated Solution Possible?
- Planning
- Project Start-up Methodology
- The process for achieving automation
- Q&A

Is Fully Automated Solution Possible?

- Depending on who you ask....
- Depending on what you consider is good output....
- WYSIWYG?
- Is it more reliable than hand-tagging?

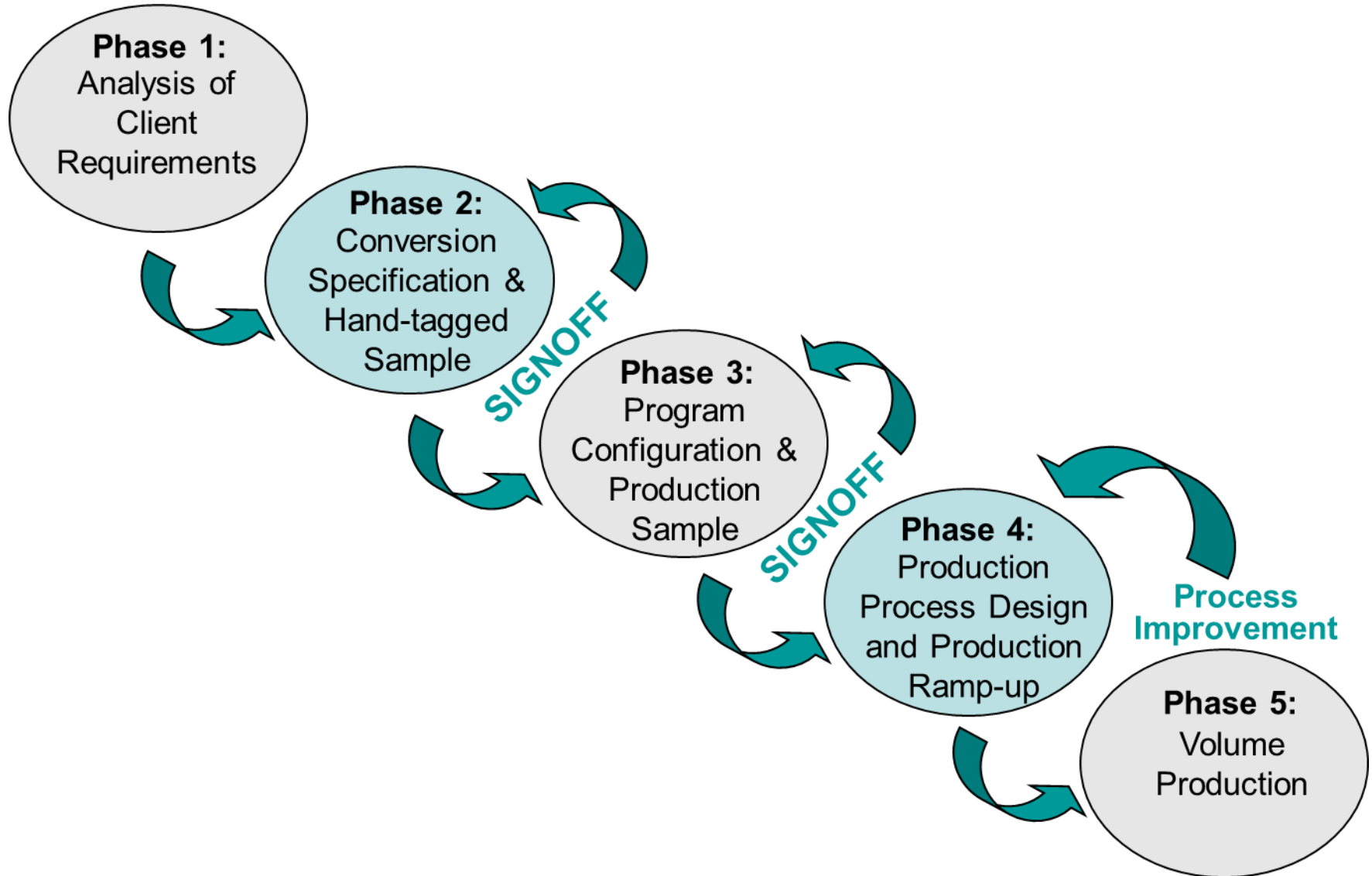
Planning



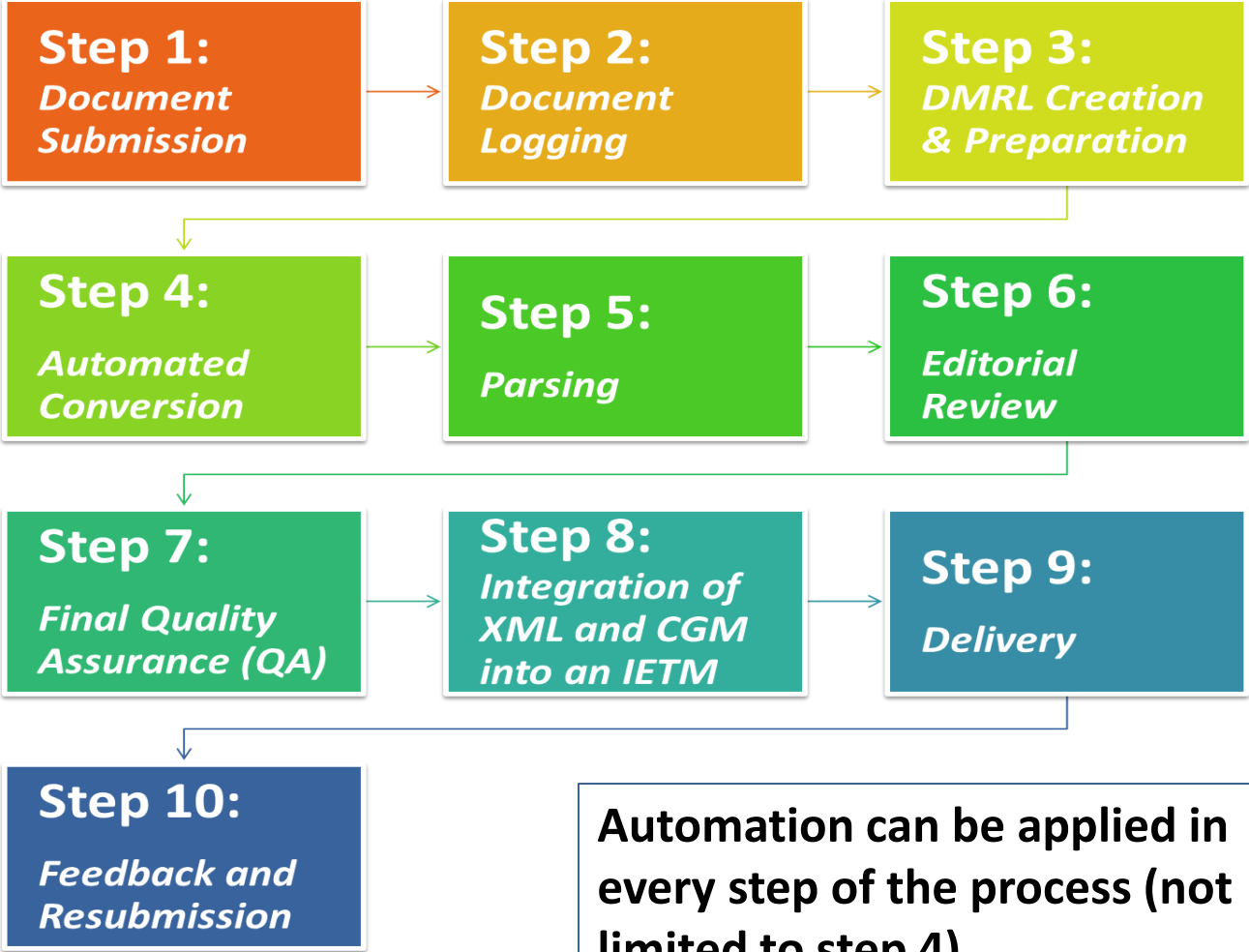
“If I had eight hours to chop down a tree, I'd spend six sharpening my ax.”

- **Things you must know:**
 - Who are the stakeholders? Who is the real client (they **have to** take part in the process)?
 - Budget.
 - Schedule (or at least know major milestones & completion date).
 - CSDB, IETM, sustainment tools.
- **Things you need:**
 - Legacy data, business rules, schemas, samples.
 - The “right people”. It’s a team work.
- **Things to do:**
 - QA plan
 - Agree on rules.
 - Have samples before moving forward.
 - Pilot and/or Limited Production Run.

Project Start-up Methodology



Workflow With Automated Conversion



Automation During Analysis – If Legacy in ASCII

- 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](#)

[253P6310 line: 90](#)

This is the safety alert symbol. When you see this symbol on your pump or in this manual, look for one of the following signal words and be alert to the potential for personal injury.

This is the **safety {safety-alert.} alert symbol**. When you see this symbol on your pump or in this manual, look for one of the following signal words and be alert to the potential for personal **injury. {injury; injury!}**

This is the safety alert symbol. When you see this symbol on your pump or in this manual, look for one of the following signal words and be alert to the potential for personal **injury-injury!**

[S502 line: 69](#)

[S506 line: 56](#)

[S509 line: 54](#)

- 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](#)

Paper Legacy - Zoning

Image

 Edit Image
 Analyze
 Text
 Picture
 Table

VOL. 34, 1961 VACCINATION AND IMMUNITY TO TULAREMIA 709

however, were recovered from all of the i.p. vaccinated rats through day 4. The RAW fluids from all SPA-vaccinated rats yielded LVS organisms for 14 days at concentrations that did not exceed $3 \log_{10}$. The LVS also was recovered from the RAW fluids from one-third of the i.p. vaccinated rats between days 3 through 10. Viable LVS bacteria were not detected in the RAW fluids of any members of the i.n. vaccinated group.

Serum and RAW fluids were assayed for LVS agglutinin titers (Table 2). For all vaccinated groups, serum titers were evident by day 7. Generally, the titers peaked by day 10 and subsided to their lowest levels ($1-1,200$) by day 42. The highest serum titers were produced by i.p. vaccination. Low-titered RAW agglutinins were

detected for about two-thirds of the SPA-vaccinated rats through day 10. Serologically positive RAW fluids were detected only on day 7 for one-third of the i.p. vaccinated rats and for one-third of the i.n. vaccinated animals only on day 10.

Changes in concentrations of the different classes of immunoglobulins were detected only for IgM (Fig. 2). Serum IgM levels in both aerosol and i.p. vaccinated rats were significantly higher than those of the nonvaccinated control rats from days 7 through 14 (analysis of variance, $P \leq 0.05$). The serum IgM concentration for the i.n. vaccinated rats was increased significantly only at day 14. IgM also was detected in the RAW fluids of some of the SPA-vaccinated rats on days 7 and 10 and in one i.p. vaccinated rat on day 7. The increase in IgM levels in both serum and RAW fluids paralleled the increase and abatement of the LVS agglutinin titers in these fluids (Table 2).

Response of LVS-vaccinated animals to aerosol challenge. At 42 days after LVS vaccination, the sham-vaccinated control rats and the five groups of vaccinated rats were challenged with a mean aerosol dose of $5.3 \log_{10}$ cells of virulent *F. tularensis* SCHU 54. The data shown in Fig. 3, however, are for the control, aerosol-, i.n., and i.p. vaccinated groups of rats only. All of the control rats appeared ill 2 days after aerosol challenge. By day 7 after challenge, concentrations of challenge organisms ranged from $7 \log_{10}$ in the lungs to about $5 \log_{10}$ in the spleen and lymph nodes of the control rats, despite the presence of a high serum antibody titer (mean, $1:1,178$). Of these rats, 34% died

TABLE 2. Serum and RAW agglutinin titers in LVS-vaccinated rats

Vaccination and	"Mean titer (range)" with time by day		
Group	7	14	42
a. Serum			
SPA	100 (10-20)	100 (10-70)	100 (10-20)
RAW	47 (2-4)	47 (2-4)	11
b. Serum			
SPA	73 (27-9)	78 (45-2)	100 (10-60)
RAW	11	41 (1-6)	11
c. Serum			
SPA	100 (10-20)	100 (10-70)	100 (10-20)
RAW	11	11	11

* Geometric mean log₁₀ concentration of agglutinin per unit of serum, 0.1-0.5 ml.
 † Percentage of positive animals.

FIG. 2. Relative levels of IgM in serum and RAW fluids as measured by radial immunodiffusion and based on diameter (millimeters) of precipitin ring.

FIG. 3. Relative levels of IgM in serum and RAW fluids as measured by radial immunodiffusion and based on diameter (millimeters) of precipitin ring.

The mean diameter of the precipitin ring was calculated on the basis of three replicate samples per rat ($n = 6$). The asterisk above a bar for the serum samples indicates an IgM concentration significantly higher (analysis of variance, $P < 0.05$) than that of the broth control. The numbers over the bars for the RAW samples indicate the proportion of the rats with IgM concentrations increased over baseline values.

DCL
Data Conversion Laboratory Inc.

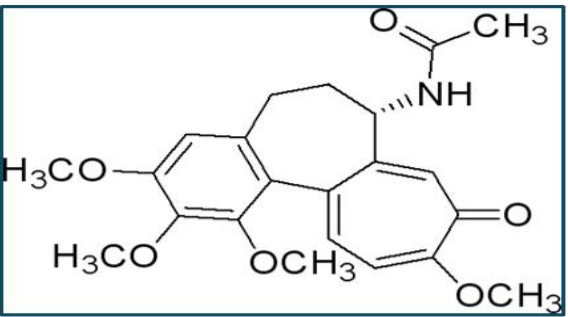
www.dclab.com

Confidential & Proprietary

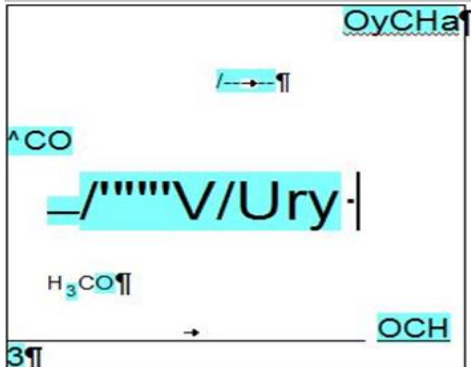
8

Proofreading – Extracted Data

Sample Document Text



OCR Output

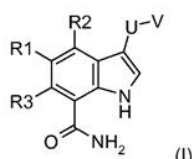


Sample Page

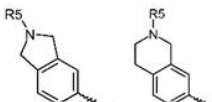
DETAILED DESCRIPTION OF THE INVENTION

The invention is directed to compounds according to formula (I):

15



(I)



where R1 is the group -XYZ or or ;
 X is phenyl, heteroaryl, 1,2,3,4-tetrahydronaphthalenyl, or 2,3-dihydro-1H-indenyl,

Sample Page After Artifact Extraction

DETAILED DESCRIPTION OF THE INVENTION

The invention is directed to compounds according to formula (I):

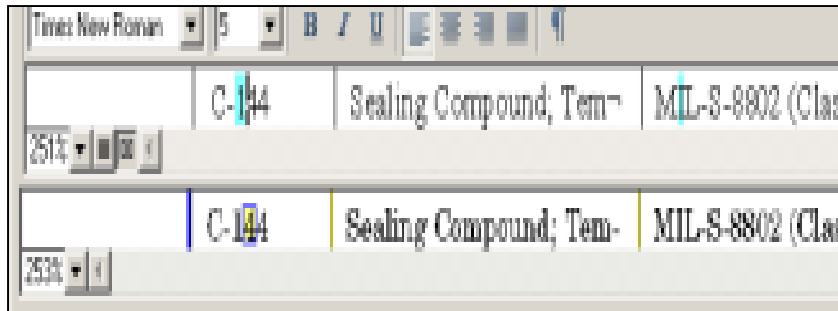
15

(I)

where R1 is the group -XYZ or or ;
 X is phenyl, heteroaryl, 1,2,3,4-tetrahydronaphthalenyl, or 2,3-dihydro-1H-indenyl,

Proofreading – Visual & Automated

C-1440ⁿ Sealingⁿ 0-Ring



and TP2 for measurement of the SH2 to test points TP3 and TP4 ved at J1-L. To test the overload of R2 by OVERLOAD switch AUX 1 output is +5 V dc. The short ORT CIRCUIT switch S21. tput mode of the UUT.

to test points TP1 and TP2 for meas across output shunt SH2 to test poin AUX 1 output is received at J1-L. To calibrated resistance of R2 by OVE amperes when the AUX 1 output is ground through SH0RT CIRCUIT s

----- Probable ERRORS -----			
String	Char	Count	Line #

115-Vac	#1	1	117
125psi	#1	2	554,788
1070T10	#0	3	3049,3437
1070T10-7	#0	1	3083
1070TO-5	#0	1	3095
0-ring	#0	1	7434
40F76Y56-72014	0	1	8179

VARIANTS -----			
Not symmetrical space -----			
A01 -HR	# 1		
1 -Reference	# 1		
1- 223645	# 1		

Space before punctuation mark -----			
removed .	# 2		
21 .	# 1		
of .	# 2		

Preparing for conversion

(1) Connect shovel attachment (5) into socket (4) on end of ax (1).

(2) Insert safety locking pin (6) into hole. (3) Grease safety locking pin (6)

(4) To remove shovel attachment (5), remove safety locking pin (8) from

attachment (5) and ax

d. Tool Installation

(1) Replace attachments

(2) Install max tool kit c
footwell (8) of vehicle

24.2[[pgbr]]

2-23. SLAVE ST

2-3. GENERAL

A permanent record of the services, repair must be recorded. (Army) See DA Pam 738 required and how to complete them. (Marin

2-4. CLEANING INSTRUCTIO

a. Cleaning is an after operation service

Automated Conversion – Analyze & Normalize

2-1. COMMON TOOLS AND EQUIPMENT.

For authorized common tools and equipment refer to Modified Table of Organization and Equipment (MTOE) applicable to your unit, Navy, Air Force and Coast Guard users are authorized to use the following common tools:

Task	Paragraph	Tool
Maintenance of 9mm Pistol	2-11	None required
Maintenance of Receiver Assembly	2-12	Screwdriver
Maintenance of Magazine Catch Assembly	2-13	Screwdriver

2-2. SPECIAL TOOLS AND SUPPORT EQUIPMENT.

There are no special tools for this item. Tools and test equipment are listed in appendix B of this manual. There is no Test, Measurement, and Diagnostic Equipment (TMDE) for this item.

2-3. REPAIR PARTS. Repair parts are listed and illustrated in appendix C of this manual.

2-13. MAINTENANCE OF MAGAZINE CATCH ASSEMBLY.

This task covers removal, reversing and installation.

INITIAL SETUP

Tools and Special Tools

Tool Set, Small Arms (SC 5180-95-CL-A07)

Equipment Condition

Pistol grips removed

WARNING

Make sure weapon is clear and there are no obstructions in the barrel or chamber.

REMOVAL

Remove the magazine catch assembly (1) by pushing in and to the rear with the fingertip, on the side opposite the magazine catch assembly button (2). The magazine catch assembly will then drop out.

REVERSING AND INSTALLATION

NOTE

To reverse the magazine catch assembly, install the button on the opposite side.

DMRL will map to DM but, when possible, automation is needed to restructure data

Collect Titles to a “Pre analysis Spreadsheet” or a DMRL

Frame Maker:

Chapter 1: XML
 Working with XML files
 About XML
 XML is a language designed to mark content so it can be output in multiple ways. Like an HTML file, an XML file includes tags that surround and organize content. However, unlike HTML, XML tags do not span

WORD:

008-02
 Organizational, Intermediate, and Depot-Level Maintenance

Support Equipment Required

Quantity	
1	Wrench, 7/16 in

8-20. GENERAL

8-21. This section contains the procedures for inspection, testing

Upon completion of any maintenance action (e.g.

8-22. Procedural steps outlined in this section are listed under 8-23. INSPECTION

8-24. ACCEPTANCE/TURNAROUND/DAILY/Required Visual Inspections of aircraft (acceptance, turnaround

SGML/XML:

```
<chapter><title>
<section><title>
<para0><title>
<proc><title>
```

PDF:

1. General
 - A. This procedure has these tasks:
 - (1) TASK 31-31-11-000-802, Flight Data Recorder
 - (2) TASK 31-31-11-400-802, Flight Data Recorder
 - B. [SF] The flight data recorder is installed between STA compartment.

TASK 31-31-11-000-802

Flight Data Recorder Removal
2. References
 - A. References

Indesign:

Keeping control over the design process
 When design and flexibility is required to create an output there is a designer to do what they do best. IdXML helps designers work without compromising their creativity.

What is IdXML?
 IdXML is a plugin for the DITA Open Toolkit that creates a simplified structured document ready to import into Adobe InDesign.

What can it do?
 Adobe InDesign can be used to create more than just printed output. It is also a powerful tool to create reflowable and fixed-layout ePubs, interactive PDFs, integrate with Adobe Digital Publication Solution and much more.

Customers are using IdXML to create a variety of outputs from content within the DITA CCMS, including:

- datasheet
- specification
- pharmaceutical package insert
- product user guide
- fixed layout ePub

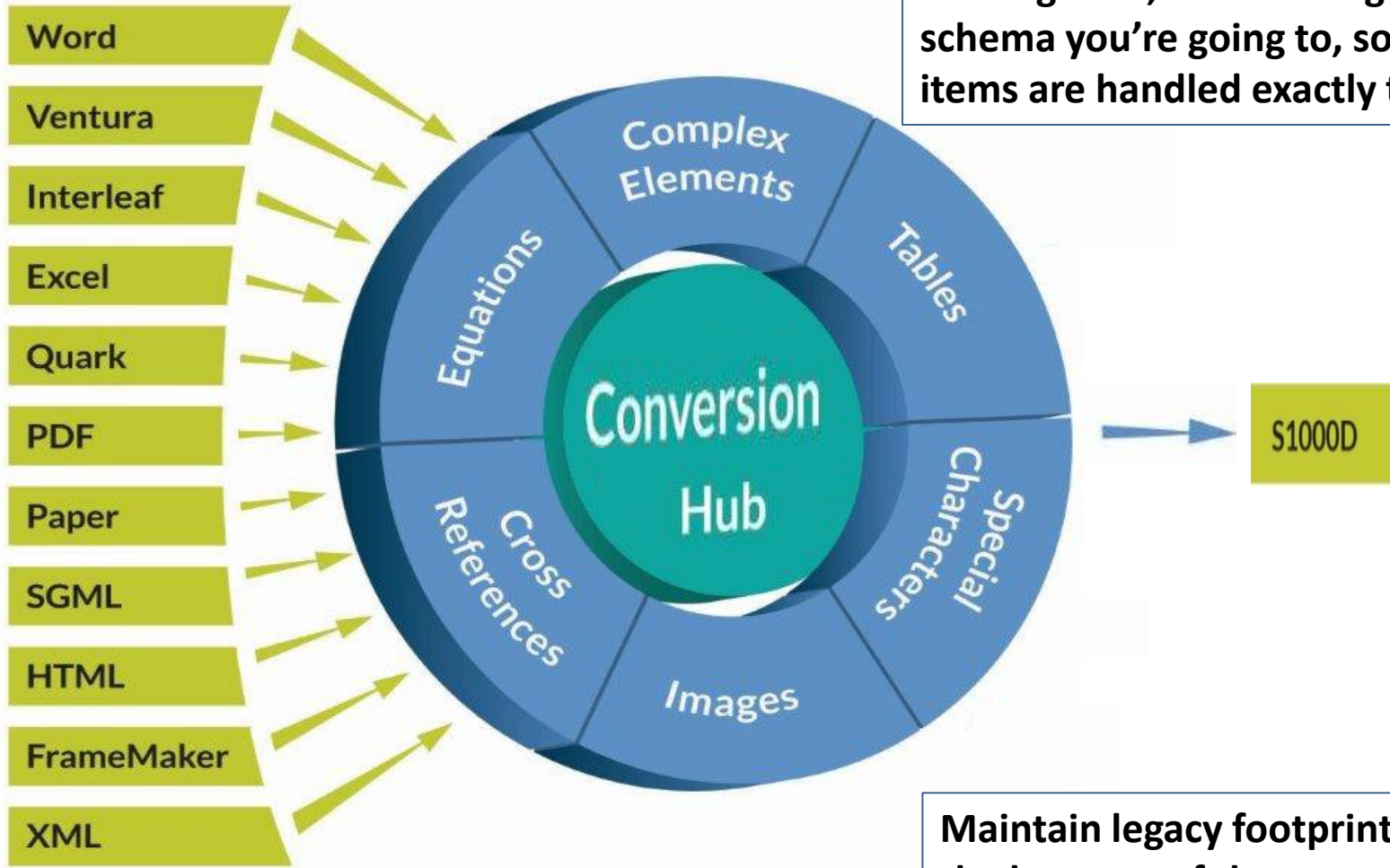
Scripting
 Adobe ExtendScript provides a

Collect Titles to a Pre analysis Spreadsheet or DMRL

Para No	Original Title	DMC	Type	Comment	Tools	Ref
	FOREWORD	SAMPLESYSTEM-A-00-00-00-00A-018A-D	Desc			
1	INTRODUCTION.	-----				
2	MANUAL ORGANIZATION.	-----				
CHAP 3	PREPARATION FOR USE AND RESHIPMENT	NA				
SECT I	PREPARATION FOR USE	SAMPLESYSTEM-A-00-00-00-00A-006A-A	Desc			
3.1	INTRODUCTION.	-----		Repeat: 3.2, 3.3		
3.1.1	Nomenclature Setup and Initialization Instructions.	SAMPLESYSTEM-A-00-00-00-00A-007A-A	proc		Wrench[1];Hammer[1]	Para 3.3;Chapter 2;
SECT II	PREPARATION FOR RESHIPMENT	SAMPLESYSTEM-A-00-00-00-00A-530A-A	proc		Screw[4];Washer[4]; O-ring[2]	Para 3.1.1, TO 14-77-001
3.2	COMMON TOOLS.	-----	Reconst			
3.3	CONDITIONS AND METHODS OF RESHIPMENT.	-----				
CHAP 8	ILLUSTRATED PARTS BREAKDOWN	SAMPLESYSTEM-A-00-00-00-00A-018B-D	Desc			
SECT I	INTRODUCTION	NA				
8.1	INTRODUCTION.	-----				
8.2	MAINTENANCE PARTS LIST.	-----				
SECT II	MAINTENANCE PARTS LIST	SAMPLESYSTEM-A-00-00-00-00A-941A-D	IPD			
SECT III	NUMERICAL INDEX	N/A				
SECT IV	REFERENCE DESIGNATOR INDEX	N/A				
	GLOSSARY	SAMPLESYSTEM-A-00-00-00-00A-006A-D	Desc			

Automated Conversion Software

Regardless of the legacy you're starting with, and the target XML schema you're going to, some items are handled exactly the same



Maintain legacy footprint until the last step of the process

Reporting, Audit & Reconciliation

```
.....
Invocation: airqc($Revision$1.2.0) -s -pgraphics -mk:\P27078\sw\qc\qc#.ctl WP002.sgm
* ERR * EOF, tags not closed: wptablelist wplep wpilluslist , msg=906

* ERR * Table orient=port width is not standard: 7 (not 6.5) Inches; line #=106, msg=337
* ERR * Table orient=port width is not standard: 7 (not 6.5) Inches; line #=196, msg=337
* WARN * <emphasis> ends on space: 'Help '; line #=164, msg=569
* WARN * Check space: '/ ' in <para>When the symbol &lt;graphic1&gt; is placed between a procedure/step
* WARN * Check title text 'FOREWORD' in <title>FOREWORD</title>; line #=13, msg=121
* WARN * Check if not WP ref 'See WP 007' in <para>See WP 007 (Maintenance Procedures for Operators) ;
* WARN * Graphic with boardno='703_0356' occur 2 times; line #=526, msg=252
* WARN * Graphic with boardno='703_0418' occur 2 times; line #=3443, msg=252
* WARN * Table in body/rear without title text: <table frame="all">; line #=91, msg=320
* WARN * Table with width less then 3.5 inches spans page Table width 2.32 < 3.500001388888889 Inches; :
* WARN * Check if not figure or figure ref 'FIG 3/' in <entry>FIG 3/</entry>; line #=77, msg=509
* WARN * Check if not reference 'to paragraph 3.203).' in <para>button. The system saves the cable swe:
* WARN * Content is absent '' in <title></title>; line #=92, msg=604
* WARN * Element has no content '' in <title></title>; line #=92, msg=606
* WARN * Step without label: Wait until Windows safely shuts down and the screen goes black. 'label'; :
* WARN * Content is absent '' in <para></para>; line #=75, msg=704
* INFO * ***** *QC CTL file: k:\P27078\sw\qc\qcLEG.ctl ***** , msg=800
```


Reporting, Audit & Reconciliation

```
rr.sum x Document2* DMC-CL94-0415-22-21-00-010-941A-... DMC-CL94-0415-11-30-00-020-941A-... iai01-29fi-proposal_chp2.sgm Document1.txt*
errs#: 2 , Attribute @dmcode is not allowed on element <dmcode> (See http://www.w3.org/TR/xmlschema-1/#cvc-complex-type
errs#: 2 , Attribute @inschlvl is not allowed on element <levelledPara> (See http://www.w3.org/TR/xmlschema-1/#cvc-con
errs#: 3 , Attribute @label is not allowed on element <listItem> (See http://www.w3.org/TR/xmlschema-1/#cvc-complex-type
errs#: 2 , Attribute @leftind is not allowed on element <tgroup> (See http://www.w3.org/TR/xmlschema-1/#cvc-complex-type
errs#: 5 , FORG0001: The value ' ' is not a valid NCName (See http://www.w3.org/TR/xmlschema-1/#cvc-complex-type clause
errs#: 3 , FORG0001: Value "unknown" contravenes the enumeration facet "modify, delete, add" of the type changeTypeAtt
errs#: 1 , In content of element <action>: The content model does not allow element <brk> to appear here. Expected one
errs#: 3 , In content of element <action>: The content model does not allow element <note> to appear here. Expected or
errs#: 7 , In content of element <action>: The content model does not allow element <note> to appear here. Expected or
errs#: 1 , In content of element <action>: The content model does not allow element <query> to appear here. Expected c
errs#: 1 , In content of element <dmIdent>: The content model does not allow element <dmcode> to appear here. Expectec
errs#: 1 , In content of element <dmIdent>: The content model does not allow element <dmcode> to appear here. Expectec
errs#: 11 , In content of element <isolationProcedureEnd>: The content model does not allow element <dmRef> to appear h
errs#: 5 , In content of element <levelledPara>: The content model does not allow element <sequentialList> to appear h
errs#: 11 , In content of element <para>: The content model does not allow element <brk> to appear here. Expected one c
errs#: 8 , In content of element <para>: The content model does not allow element <brk> to appear here. Expected one c
errs#: 4 , In content of element <para>: The content model does not allow element <brk> to appear here. Expected one c
errs#: 1 , In content of element <para>: The content model does not allow element <note> to appear here. Expected one
errs#: 2 , In content of element <para>: The content model does not allow element <note> to appear here. Expected one
errs#: 4 , In content of element <para>: The content model does not allow element <table> to appear here. Expected one
errs#: 3 , In content of element <para>: The content model does not allow element <table> to appear here. Expected one
errs#: 1 , In content of element <para>: The content model does not allow element <table> to appear here. Expected one
errs#: 1 , In content of element <para>: The content model does not allow element <table> to appear here. Expected one
errs#: 2 , In content of element <proceduralStep>: The content model does not allow element <caution> to appear here.
errs#: 9 , In content of element <proceduralStep>: The content model does not allow element <caution> to appear here.
errs#: 4 , In content of element <proceduralStep>: The content model does not allow element <figure> to appear here. E
errs#: 2 , In content of element <proceduralStep>: The content model does not allow element <note> to appear here. Exp
errs#: 1 , In content of element <proceduralStep>: The content model does not allow element <warning> to appear here..
errs#: 15 , In content of element <proceduralStep>: The content model does not allow element <warning> to appear here.
errs#: 1 , In content of element <procedure>: The content model does not allow element <proceduralStep> to appear here
errs#: 8 , In content of element <procedure>: The content model does not allow element <proceduralStep> to appear here
errs#: 2 , Required attribute @assyCode is missing (See http://www.w3.org/TR/xmlschema-1/#cvc-complex-type clause 4)
errs#: 2 , Required attribute @disassyCode is missing (See http://www.w3.org/TR/xmlschema-1/#cvc-complex-type clause 4)
errs#: 2 , Required attribute @disassyCodeVariant is missing (See http://www.w3.org/TR/xmlschema-1/#cvc-complex-type c
errs#: 2 , Required attribute @infoCode is missing (See http://www.w3.org/TR/xmlschema-1/#cvc-complex-type clause 4)
errs#: 2 , Required attribute @infoCodeVariant is missing (See http://www.w3.org/TR/xmlschema-1/#cvc-complex-type cla
errs#: 2 , Required attribute @itemLocationCode is missing (See http://www.w3.org/TR/xmlschema-1/#cvc-complex-type cla
errs#: 2 , Required attribute @modelIdentCode is missing (See http://www.w3.org/TR/xmlschema-1/#cvc-complex-type claus
errs#: 2 , Required attribute @subSubSystemCode is missing (See http://www.w3.org/TR/xmlschema-1/#cvc-complex-type cla
```

Q&A



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