Best practices for automated delivery setup, from raw data to CRT, with methods to standardise and streamline programming and delivery processes.

Daniel Carden, PhUSE EU Connect Belfast, November 14th – 16th
Stream: Leadership & Innovation (LI), **Presentation No:** LI06

© 2021. All rights reserved. IQVIA® is a registered trademark of IQVIA Inc. in the United States, the European Union, and various other countries.

Presentation audience and objectives

Programming leads:

- Tips to allow focus on team management and analysis
- Maximise process reusability, contribute to tool development

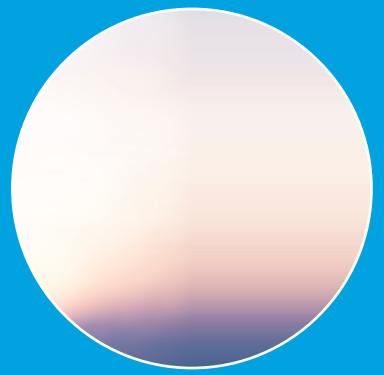
Senior managers:

- Proposals can enhance horizontal and vertical coordination
- Technical knowledge links to management and delegation

All:

- Tips to prioritise all tasks, issues and objectives
- Encourage technical investment and the link to soft skills

The ideal vision and the delivery worst case

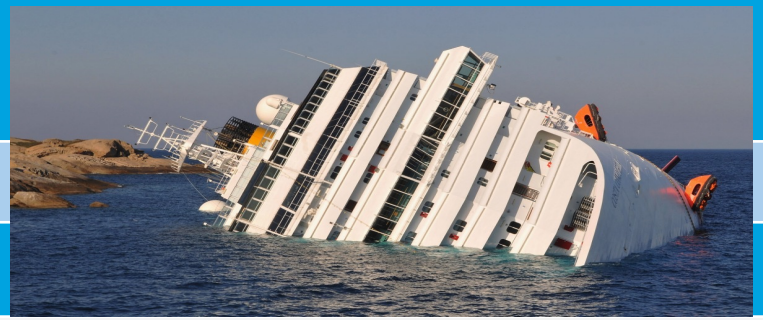


Best case: “It’s taken care of”

Can focus on the other things

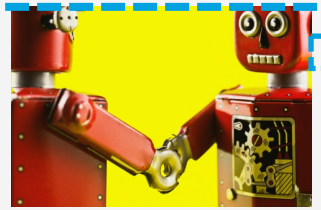
Worst case:

“Oh dear, this or that is missing or not ready, I thought it was fine”

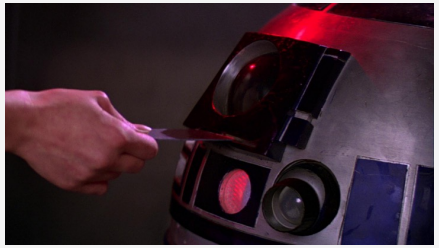


Biostatistical delivery stages and solutions

DATA MANAGEMENT (DM)



OTHER SOURCES



1. RAW DATA

Impact of error: Critical. Challenges: Changes to files, data issues.

Resources: Programming lead. **Solutions:** Delivery check tool, static domain naming of files, documentation, data issue logging proposals.

2. SDTM/ADaM

Impact of error: High. Challenges: Specifications, P21 validation.

Resources: Prod+QC programmers. **Solutions:** Standardised and integrated specifications and codes, less but higher quality P21 commentary.

3. Outputs (TLFs)

Impact of error: Medium. Challenges: Multiple outputs, large teams.

Resources: Prod+QC programmers. **Solutions:** Macros for repeats, lead to direct programming, tracking spreadsheet or equivalent to know status.

4. Define.xml and CRT packages

Impact of error: Medium. Challenges: Skill set to document and explain.

Resources: Programming lead+programmer. **Solutions:** Update the define.xml efficiently, P21 comment tools, see also raw data+SDTM/ADaM.

MEDICAL REVIEW

REGULATORS

Budget versus time for tasks

+ Presentation covers low budget, high difficulty tasks.

+ Management should pay attention to these tasks, to mitigate risk of overspend.

Budget

High budget, low difficulty

High budget, high difficulty ☑

Low budget, low difficulty ☑

Low budget, high difficulty ☠!

- Data issues
- SDTM P21 commentary
- Define.xml (+CRT) updates
- Modify complicated programs

Difficulty

Raw data challenges

Change management

1

Data completeness

Data issues in raw

2

Minimal DM time to talk

Raw data solution 1a, DLEC tool plus static names

Change management | 1 | Data completeness

+ Data Lineage and Export Comparison tool (DLEC) run each time there is a new data transfer

+ Tool works based on static file names and lists out:
  - Whether dataset is still present or if a new dataset may need mapping to SDTM.
  - Whether there are changes to variables and/or number of observations.

+ Ways to enhance this analysis:
  - Create process for converting versioned file names to static names, for use in DLEC and programs.
  - For example ctcv5rf_16MAY2022.xpt becomes **lbctc**.sas7bdat, such that **lbctc** can be analysed.
  - Also consider using compares and flexible summary code (see appendix A), for comparing counts

Raw data solution 1b, external file documentation

+ Document the conversion of files received to static names.
+ NB: That this activity should ideally be performed by DM or equivalent, however in their absence then this process can accommodate adhoc files flexibly.
+ Static names are based on SDTM domains, names are consistent with SAS transport files.

	A	B	C	D	E	F	G	H	I	J	
1	...Data Transfe	Categor	...Data\Extern	Non-DM file(s)	Extensi	Type	SDTM or ADa	DR/IDMC/Other	Update?	Source (such as email subject .msg*)	C
38	2022-07-04-DR2	AE	aecovid	COVID-19 Related_New Terms_23.0-25.0_1.xlsx	xlsx	Import	ADaM	DR only	N/A	RE Pacific 5 files required for DR2 20220715 AECOVID	
42	2022-07-04-DR2	CM	cmpedco	pedconv_29june2021.xpt	xpt	Copy to sas7b	ADaM	Both DR and IDMC	No	RE Pacific 5 files required for IDMC5 20220324	
44	2022-07-04-DR2	DM	dmccodes	CDISC_COUNTRY_CODES.xlsx	xlsx	Import	ADaM	Both DR and IDMC	No	DR1	
46	2022-07-04-DR2	DM	dmrand	d933yc00001_rand.sas7bdat	sas7bdat	Copy/rename	SDTM	Both DR and IDMC	No	A&R meeting, topic no.21	
51	2022-07-04-DR2	LB	lbref	lbref6_6.xpt	xpt	Copy to sas7b	Both	Both DR and IDMC	Yes	RE Pacific 5 files required for DR2 20220629	C

Raw data solution 2, the data issues log

“Three invalid exposure values as per the Protocol,

Seven height measurements recorded as below a metre,

Nine subjects with inconsistent dates and missing causes of death,

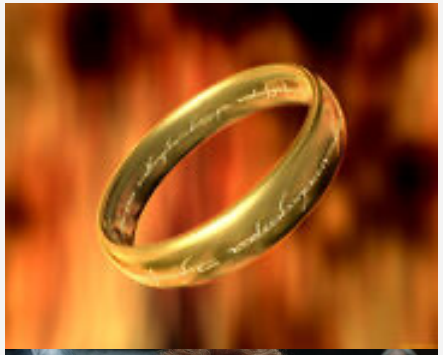
One Data Issues Log for the Programming Lead on SharePoint

In the Collaboration Zone where the queries lie.

One Data Issues Log to rule them all, One Data Issues Log to find them,

One Data Issues Log to bring them all and in Excel bind them

In the Collaboration Zone where the queries lie.”



Raw data solution 2a, the data issues log

+ Create a priority column for each data issue that can be summarized as needed.

+ This will direct DM to the data issues of greatest concern.

	A	C	D	E	F	G	H	L	M	N
1	ISS NO		DAT TRA ER	DATAS	VARIABLE	SUBJECT NUM	ISSUE	DM Comments	BIOS Comments	Priority
182	181		21-Jun-21	AE	AETERM		AEANY = '1' and AETERM is missing	SR_11Aug2021 - Query issued by Ops Team.		2. Medium - need DM resp
183	182		21-Jun-21	AE	PT_NAME		AEANY = '1' and AETERM is populated and PT_NAME is missing	NZ 09Aug2021:coding is ongoing action, should be included next round.	DC 16Aug2021: Ok thank you, have downgrade	3. Low - expect to be resolv
184	183		21-Jun-21	LB	Multiple		Records with the same LBDAT LBTIM values have multiple VIS_DAT entries.	TG 24May2022: Comments added on Additional info tab SR_11Aug2021 - Ops comments added to Additional info tab	DC19APR2022: Thanks for investigating, I have added in new entries based on the latest transfer on 7th April 2022.	2. Medium - need DM resp
185	184		21-Jun-21	BMRESOM	BMDAT		May be the typo error as the date is 2002-03-20 which is not expected.	TG 24May2022: date updated to 2002 Apr 05 with query answered as "Updated the correct reporting year.", Please assess if acceptable SR_23Sep2021 - Query issued		2. Medium - need DM resp
186	185		21-Jun-21	BMRESON	BMRLTS		Variable is missing for 4 records.	SR_23Sep2021 - BMRLTS is recorded as "Negative" for subject E7603001. Please advise if there are any other subjects where this is missing in dataset.		2. Medium - need DM resp
187	186		21-Jun-21	AE/CM	AETRT		AETRT = '1' however there is no corresponding record in CM.	TG 23May2022: E6209004-re-queried-site response as "No available info about treatment we have. Sorry.", please assess if acceptable other information are on Additional_Info tab column BQ SR_23Sep2021 This is reviewed and queried on SAS listings review (AE_004).		2. Medium - need DM resp
188	187		21-Jun-21	AE/CM	CMAENO variables		AE record exists but AETRT = '0'.	TG 23May2022: data updated, **issue can be closed** SR_23Sep2021 Similar TRISARC query exists for AETRT = No, but AE #7 is recorded on CM page.		2. Medium - need DM resp
189	188		7-Apr-22	DEATH/DS	Subject		Death recorded in DS but not in DEATH.	TG 18May2022: query raised E1336001-data added	First raised as DI #163.	1. Critical
190	189		7-Apr-22	DEATH/DS	Subject		Death recorded in DEATH but not in DS.	TG 18May2022: query raised	First raised as DI #164.	1. Critical

	A	B
1	Row Labels	Count of Priority
2	1. Critical	11
3	1. High - need DM response	18
4	2. Medium - need DM response	62
5	3. Low - expect to be resolved later	19
6	3. Low - open issue but very small impact	4
7	4. Closed	160
8	**Grand Total**	**274**

Raw data solution 2b, the data issues log code

+ Raw data logic is distinct from SDTM/ADaM logic.

+ Takes time to setup DI check logic but once written it is reusable, it is effectively P21 but on the **raw** data.

Issue Summary						
Source	Pinnacle 21 ID	Publisher ID	Message	Severity	Found	IQVIA Comments
CE	SD1202	FDAB034	CESTDTC date is after RFPENDTC		1	Data issue #193 has been logged.

```
*Data quality issues 163/164: Mismatch between Death and DS as to whether patient died.;

data ds_deaths;
     set c1.ds;
     if upcase(vvalue(DSDECOD)) = 'DEATH';
run;

proc sql;
     create table death_subject as
     select distinct subject, DTH_DAT from c1.death;
quit;

data DQ163_death_ds_only
      DQ164_death_death_only (keep = subject);
     merge death_subject (in = a)
           ds_deaths     (in = b);
     by    subject;
     if      a and not b then output DQ164_death_death_only;
     else if b and not a then output DQ163_death_ds_only;
run;

*DM suggestion is that it is acceptable for screen failure only patients to be recorded as death in DS only.
 So add logic to check for non.missing randomization code (patient).;

data DQ188_death_ds_only
      DQ189_death_death_only (keep = subject)
      death_and_ds;
     merge death_subject (in = a)
           ds_deaths     (in = b);
     by    subject;
     if      a and not b                  then output DQ189_death_death_only;
     else if b and not a and patient ne . then output DQ188_death_ds_only;
     else if b and a                      then output death_and_ds;
run;

data DQ193;
     set death_and_ds;
     where DTH_DAT ne DSSTDAT;
run;
```

	A	C	D	E	F	G	H	L	M	N
1	ISSUE NO.	M	DATE C TRANSFER	DATASET	VARIABLE	SUBJECT NUMBE	ISSUE	DM Comments	BIOS Comments	Priority
193	192		7-Apr-22	CE	CETHW	E7000004	Covered by data issue 55 have been excluded.	Please consider to add logline number into programming.	Coming through.	1. High - need DM response
						SUBJECT DTH_DAT DSSTDAT		TG 19May2022: details are listed as below, please check		
						[redacted]021-02-07 2021-01-18		[redacted] subject withdrew in Jan, but allow for survival FU via phone-However, subject died on 07feb2021 NZ: Query in Death.		
						[redacted]2021-11-21 2021-11-22		[redacted] the team was informed about death on 22.11.2021-NZ Query in DS, ds date should be the death date.		
194	193		7-Apr-22	DEATH/DS	SUBJECT DTH_DAT DSSTDAT	[redacted]021-06 2021-06-25	Death dates recorded in DS and DEATH do not match.	[redacted] We do not have data because the patient died in an external center.		1. High - need DM response

SDTM/ADaM P21 challenges

+ Documentation of CDISC and controlled terminology versions.

+ ADaM P21 commentary tends to be simpler than SDTM, which has more data consistency checks. Commentary is hard to delegate because of complexity.

+ See below some complicated issues:
  - Multiple raw data issues contributing to a single P21 item, all of which need separate ID logic.
  - Sponsor standards that are inconsistent with CDISC and required more detailed knowledge.

32	CM	SD0021	Missing End Time-Point value		14	Data issues #217 (2), #233 (6), #234 (1), #237 (4), have been raised. 1 record for conmeds not taken because of pandemic is as per sponsor raw mapping to SDTM guidance i.e. neither CMENDTC or CMENRTPT are assigned. (2 + 6 + 1 + 4 + 1 = 14).
35	CM	SD1040	CMSCAT value is redundant with other variable values		450	Mapping of radiotherapy records to the CM domain is consistent with sponsor standards, with CMSCAT and CMTRT both being assigned as RADIOTHERAPY. However it is not consistent with the SDTM IG v3.3, which says that radiotherapy data should be mapped to the PR domain.
192	SUPPRS	SD1120	Comments in SUPPRS domain		1555	As per BICR mapping sponsor guidance "ICR Mapping Package 1.1.1.xlsx".
194	SUPPTR	SD1120	Comments in SUPPTR domain		137	As per BICR mapping sponsor guidance "ICR Mapping Package 1.1.1.xlsx".

Output (TLF) solutions

Tracking document in MS teams or similar

- Opt for applications which have best multi-user functionality
- Lead should refer to tracker in all team meetings
- Status column(s) to check progress and actions

Reporting macros, such as Gen tools

- Many outputs can be created with a single macro call, without the need for complex manipulations
- This approach gets programmers to think more about the data selection and variables used

Annotated shells and Q&A spreadsheets

- Annotations help with programming and statistician review
- The shells spreadsheet should ideally use macro variables for footnotes and titles, such that updates can be applied once
- Q&A spreadsheet for programmers to raise shell queries

Reporting environment (Gen tools Interactive Environment: GENIE)

- Interface for creation and update of delivery metadata such as titles, footnotes, RTF templates etc.
- Easy to reference template programs or to clone programs from other studies or same study

Tracking programming team status

The programming tracker is to leads what the chess board is for Chess - so do know what's there!

- **Ensure consistency** with output shells.
- Keep the tracker **up to date**, remove old info.
- Use as the basis for **daily status update** calls:

9	⊟3-EXTENT OF EXPOSURE	8
10	8. SBR gate passed	8
11	⊟3-PK/ADA	11
12	1. QC not passed	1
13	4. STL gate not passed => Prog	5
14	6. STL gate passed => Ready for SBR	5
15	⊟4-ADVERSE EVENTS	22
16	3. Output ready for STL	1
17	4. STL gate not passed => Prog	2
18	6. STL gate passed => Ready for SBR	19

Programming tracker example

+ The spreadsheet has a tab for status of all steps from receipt of the raw data to outputs:

Signature Page | External_1 | DCO_1 | SDTM_2 | ADaM_3 | **Outputs_4** | Pivot PTL | Pivot STL

+ These are the main items the lead should **own**:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1										Production				QCer			
2	Category	Table/ Figure/ Listing	Output #	Title	Shell	Dataset Used	Macro Name (pplicable)	Program	Overall Status	Init	Init2	Status (Gate 3)	Date	Init1	Init2	Status	Date (Gate 3)
61	1-BASELINE AND DEM	Table	14.1.19.a	Summary of COVID-19 study disruptions (Modified intent-to-treat set)	COVID_summary3		Prod_ COVID_su	ds208a	DD/output ready for Delivery	SG		Completed (no issues	10-Oct-2022	DC	ES	Completed (issues)	10-Oct-2022
177	9-OTHER SAFETY	Table	14.3.8.1	Vital signs variables over time	S25_VITAL			vs200	DD/output ready for QC	RC		Completed (no issues	14-Oct-2022	JP		Completed (issues)	14-Oct-2022
	8-OTHER ADVERSE	Table	14.3.6.2.3		TAE300_R1		N/A	aeonm215a	DD/output ready for QC	SG		Completed (no issues	10-Oct-2022	ES		Completed (issues)	10-Oct-2022

Matches delivery PDF and groups help assignments

Macro and program names are <u>CRUCIAL</u> !!:

- **The lead MUST ensure that if repeats are in plan, then macros are setup.**
- **Naming of macros and programs should be logical and consistent, such as macros based on shells.**

Overall Status column used by Pivot Table to track overall status.

Efficient output programming #1, disposition

+ For repeated processes, focus on **how** it is done rather than **what** is produced.

+ Disposition table program before:

Program Editor

Program Editor DS200 - Subject disposition

Output Number: 14.1.1

Default Output Filename:

Population: AllSubjects - All subjects

Manage Populations

Output Template: Standard

Manage Output Settings

Program Elements

Type	Description
catgen	Subjects who received treatment
catgen	Number who received [redacted]
catgen	Number who received
catgen	Subjects ongoing treatment at data cut-off
catgen	Subjects ongoing study at data cut-off
catgen	Subjects who discontinued [redacted]
catgen	Subjects who discontinued
catgen	Subjects who discontinued
catgen	Subjects who discontinued
catgen	Subjects who completed or are ongoing on the study
catgen	Subjects who terminated study
catgen	Subjects who terminated study

After:

Program Editor DS200_DE - Subject disposition

Output Number: 14.1.1.B

Default Output Filename:

Population: EnrDE2Pop - All Subjects

Manage Populations

Output Template: Standard

Manage Output Settings

Program Macros | Titles and Footnotes | Code Peek | Batch Submit

Modification History

Turn Modification History On

Program Elements

Type	Description	Last Edited By	Active
customcode	Create work.adsl, DCTREASn and DCREASn informats	q1055929 03Nov2021	☑
popgen	Create POP macro variables for N = counts for Cohorts	q1055929 18Nov2021	☑
popgen	Create POP macro variables for N = counts for Total	q1055929 03Nov2021	☑
catgen	Create work.catgen_cohort with catgen call for cohort counts based on work.adsl	q1055929 18Nov2021	☑
catgen	Create work.catgen_total with catgen call for total counts based on work.adsl	q1055929 03Nov2021	☑
customcode	Select rows and format text	q1055929 18Nov2021	☑
aligngen	Align Values	q1050638 11Aug2020	☑
pageformat	Set page breaks	q1055929 18Nov2021	☑
compare	QC Dataset output	q1055929 18Nov2021	☑
masterreport	Output report	q1055929 02Dec2021	☑

Efficient output programming #2, adverse events

+ Adverse event tables typically have lots of repeats, there can also be population repeats.

+ As lead I setup a macro then delegated updating titles and footnotes for 40+ outputs.

Program Editor

Program Editor AE202_DE - Adverse events by system organ class and pre

Output Number: 14.3.2.3.B

Default Output Filename:

Population: SafDE2Pop - Safety analysis set

Manage Populations

Output Template: Standard

Manage Output Settings

Program Elements

Type	Description
popgen	Population counts by cohort
customcode	Create dataset via Prod_TAEVARPT macro call
aligngen	Align
customcode	create dummy column for RTF spacing after col1
pageformat	Pagination
compare	Compare the dataset for QC
masterreport	Report

```
%macro Prod_TAEVARPT (popwhere, /*Population flag where condition, do not i
                      addwhere, /*Additional where condition*/
                      var1,     /*First variable to summarise by*/
                      sort1,    /*Sort varriable for first variable*/
                      artxt     /*AnyRowTxt for occgen*/);

    data adae;
        set adam.adae;
        %if &addwhere ne %then where &addwhere;;
        if index(COHORT,'SARC') then COHORT_GRP = 'SARCOMA';
        else if COHORT = ''     then COHORT_GRP = 'Missing';
        else                         COHORT_GRP = COHORT;
    run;

    data adsl;
        set adam.adsl;
        if index(COHORT,'SARC') then COHORT_GRP = 'SARCOMA';
        else if COHORT = ''     then COHORT_GRP = 'Missing';
        else                         COHORT_GRP = COHORT;
    run;

    %if %index(&progname,_DF) %then %do;
        %let covars = WGHTBLG1 TRT01AN;
        %let ovcols = WGHTBLG1;
        %let cvlvls = (>=35|<35) (1|2);
        %let pwhere = studpern = 1;
    %end;
    %else %if %index(&progname,_DE) %then %do;
        %let covars = COHORT_GRP;
        %let ovcols = COHORT_GRP;
        %let cvlvls = (SARCOMA|STO);
        %let pwhere = studpern = 2;
    %end;
    %else %put ERROR: progname macro variable is unexpected, &progname;;

    %OCCGEN( InDsn              = WORK.ADAE
    , WhereGrp                  = &popwhere and &pwhere
    , Style                     = NPCT
    , VarList                   = &var1 AEDECOD
    , colVars                   = &covars
    , OvrlCols                  = &ovcols
    , colVarLevels              = &cvlvls
    , SortList                  = (&sort1) (AEDECOD)
    , InclAnyRow                = Y
```

Other items for efficient delivery packages

Item should have a centralised tool+standard ☑	Ways of working best avoided ☒
Log checking process, to be applied to existing logs or when batching.	Adhoc study level log checking processes, that are run from standalone SAS programs.
Compare process, that can be run efficiently for all areas or individual programs.	
Data transfer guide template and delivery checklist, for use across studies and clients.	Every study doing their thing, with the additional effort for delivery of extra items not budgeted.
Auto-copy process for files into client provided folders, with a checklist file created.	Leads manually copying and pasting files, this is slow and introduces execution risk.
Archive process for creating snapshots of files delivered, at each delivery, that reports the delivery.	Anything that needs to be done manually for this.
File lock manager, to unlock files needed.	Having datasets lock close to the delivery ☹💣.

Summary and conclusion

Programming leads:

- Follow best practices, focus on team management and analysis
- Maximise process reusability, contribute to tool development

Senior managers:

- Leads must enable and not hold back the programming team
- If processes are better: on time deliveries, staff will be happy

All:

- Be sure to prioritise all tasks, issues and objectives
- Where possible, invest in processes and items that will be reused
- Focus the time of specialists where it is needed most

Appendix A: Flexible summary code

```
%let plista=%str(COL1);
%let plistb=%sysfunc(tranwrd(&plista,%str(,),%str( )));
%let whr=%str();

data MAIN; set ctable.; where &whr.; run;
data QC; set qctable.qc_; where &whr.; run;
proc sql;
create table prim as
select &plista.,count(*) as primcount
from MAIN
group by &plista.;
create table val as
select &plista.,count(*) as valcount
from QC
group by &plista.;
quit;

data diff;
merge prim(in=a) val(in=b);
by &plistb.;
if primcount ne valcount;
run;
```

Appendix B: Glossary

+ ADaM: Analysis Data Model
+ ADRG: Analysis Data Reviewer’s Guide
+ CSDRG: Clinical Study Data Reviewer’s Guide
+ CRT: Case Report Tabulation
+ DM: Data Management
+ Q&A: Question & Answer. QC: Quality Control
+ RTF: Rich Text Format
+ SDTM: Study Data Tabulation Model