PharmaSUG 2023 - Paper SS-274

F2: FAIR and Filing - Assessing Data Fitness for FAIR and Filing

Bidhya Basnet, Dyuthi Yellamraju, Roche-Genentech

ABSTRACT

Have you ever asked the question: "How are study programming tasks progressing?" When working on programming or filing activities for a study, a lot of time is spent planning what datasets & variables will be required to create the TLGs for a reporting activity. Upfront planning is good; however, keeping track of the progress of the different deliverables such as SDTM/ADaM including conformance, aCRF and Define.xml generation, reviewers guide completion, and TLG creation, proves challenging, especially when there are a lot of data science members involved. The reliance for gauging progress often is on 'word of mouth,' 'guesstimates,' and a 'gut' feeling. This drove us to explore a better approach to assess Filing and FAIR (**F**indable, **A**ccessible, **I**nteroperable and **R**e-usable) progress which resulted in the creation of an R Shiny application. The F2 dashboard, as we call it, creates a color coded visual to demonstrate how the study is progressing - the colors indicate where progress is going well and identify areas where teams may need to place greater emphasis on. The assessment is based on tangible evidence provided by the study team.

The F2 dashboard helps gauge the progress of study deliverables at any point of time. It could help understand how the team is FAIRing, help have conversations to know the road blocks and triage the issues that the team is facing. Our overall objective is to be Filing and FAIR ready by database lock.

INTRODUCTION

Submission activities in the pharmaceutical industry require a lot of effort and planning. There are several components to be considered – number of studies to be submitted, deliverables within the submission, the standards to be used, regulatory requirements and quality to top it off.

The greater use of the study data beyond CSR reporting and Filings is required to generate scientific insights and make more informed clinical decisions. Abiding to global data standards and proactively preparing data for reuse enables efficient data sharing both internally and externally, which means additional insights can easily be generated. The lifespan of our data far exceeds our time with it, so it's our responsibility to make sure that others can easily see how we handled it and the decisions that were made. Remembering this whilst carrying out our project work is crucial for our studies to be FAIR. Having the data in a FAIR (**F**indable, **A**ccessible, **I**nteroperable and **R**eusable) state allows us to maximize the value of our data, sharing it across could help explore new avenues in drug discovery and development activities.

A proactive approach and a mindset shift is required when developing study deliverables, be it for a reporting event or a submission. At Roche-Genentech our aim is to ensure study teams are submission and FAIR ready by the Database Lock (DBL). In an endeavor to achieve this the eSubmission (eSub) team has developed several guidance documents and tools.

E-SUBMISSION TOOLS

READINESS CHECKLIST

The Readiness Checklist has been developed to aid the incorporation of FAIR principles and submission requirements as part of the study development process. The Readiness checklist provides details on all the front loading activities that need to happen during study conduct to ensure generation of good FAIR and filing deliverables leading to DBL. It provides a logical flow of the activities that the team would need to do in order to prepare a good quality submission package; a large portion of which are already being done as a part of CSR reporting activities. The checklist calls out the different deliverables leading to a filing, the dependencies between them, the timings for generating and refreshing the deliverables – the different time points at which these should be generated / refreshed during the study life cycle.

In the screenshots below you can see the readiness checklist and how it emphasizes on frontloading

	A	B	C	D	E	F	G	H	I
1	#	Deliverable	Timing wrt DBL	SPA/DTIA programmer Role	eSub Role	Example timeframe for suggested deliverables	Order of Execution / Execution Time	Dependencies	Recommendations
10	PS8	SDTM Define.xml	Pre-DBL	* Draft: Generate Define.xml from SDTMv specifications (to identify any issues with regards to data, spec,mapping or CT besides	* Consulted	* 3 months prior to DBL REPEAT -	4 * First draft of Define generation could take	* Good quality SDTMv specifications * Clean Controlled Terminology	* Generate a draft define using the specifications. * Clean up specifications based on
11	PS9	SDTM Define Conformance checking	Pre-DBL	* Run P21 conformance checks on define.xml and fix issues if any * Provide comments on issues that cannot be fixed	Consulted	* 3 months prior to DBL REPEAT - * 1 month prior to DBL	5 * Running the Define through P21 could take from a few minutes to half-	* Availability of Define.xml * Availability of SDTM aCRFs * Availability of cSDRG	* Run the define.xml generated through P21 to ensure eSub compliance * Add necessary information / make
12	PS10	DRY RUN - Overall package validation	Pre-DBL	Run P21 checks on complete eSub package; Zip file should include * SDTM XPT * cSDRG.pdf * acrf.pdf	* Consulted in case of any issues	2-4 weeks prior to DBL	6 * Running the entire package through P21 could take up to 30	* Availability of all the eSub package components	* Do a dry-run on the entire package. Run the complete package through P21 to ensure overall package is eSub compliant.
13		DBL						DBL	
14	FS0	**FINAL SDTMv (generation with DBL data)**	Post-DBL	* Refresh SDTMv with final DBL extract	N/A	On the day of DBL	1 * SDTMv refresh would take from a few hours to a	* Good conformant SDTMv from previous runs and/or * Availability of	**FINALIZATION -** ***** Post DBL, refresh SDTMv on the right/identified source data extract to generate the **FINAL SDTMv**
	FS1	**FINALIZE P21 conformance checking on DBL SDTMv**	Post-DBL	* Responsible for running P21 conformance checks on DBL SDTMv data	* Consulted * eSub to provide	On the day of DBL	2 * Execution of P21 checks	* Good conformant SDTMv data	**FINALIZATION -** *** Post DBL run P21 checks on the final DBL SDTMv refresh**

SDTM | ADaM

Figure 1. Screenshot of Readiness Checklist

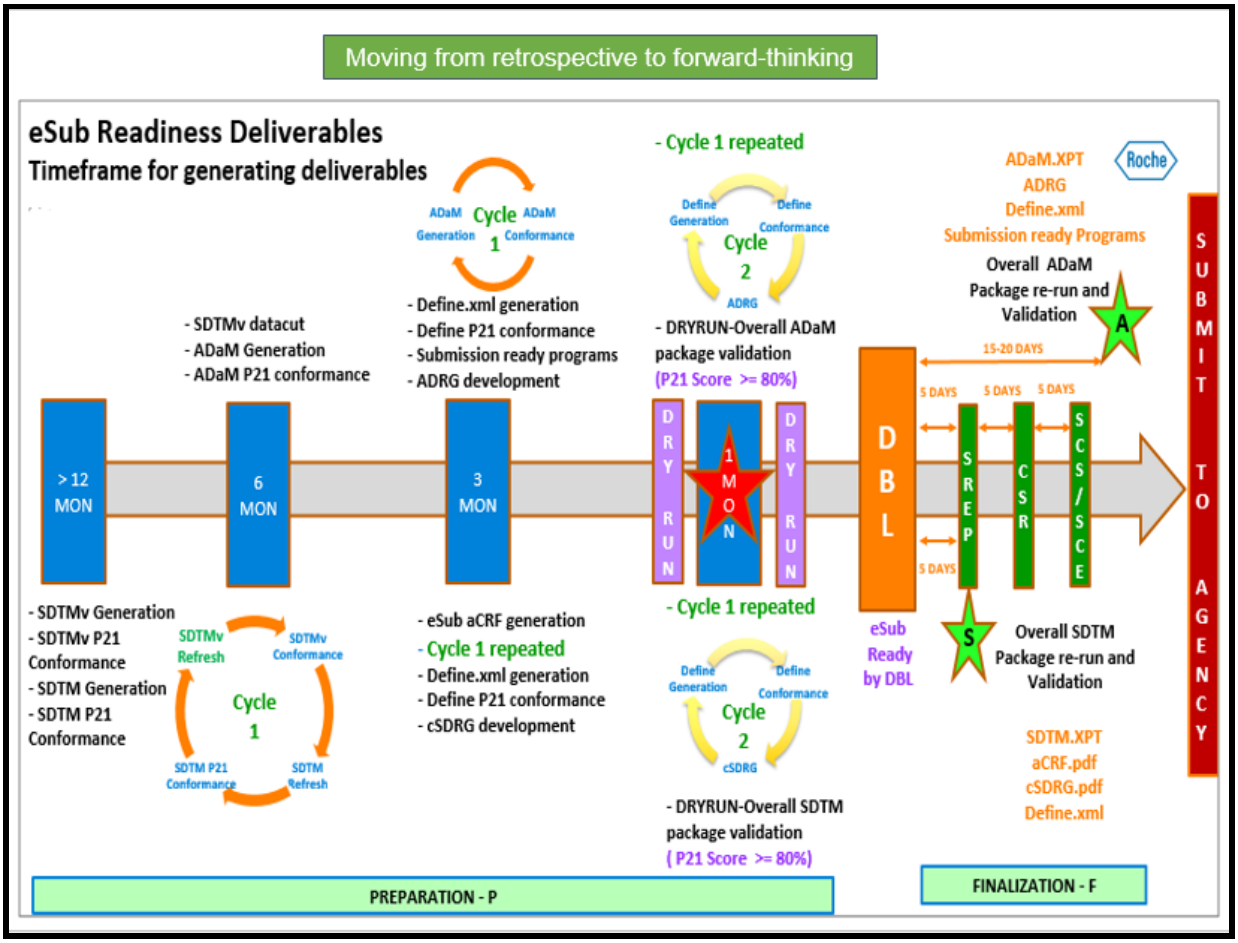
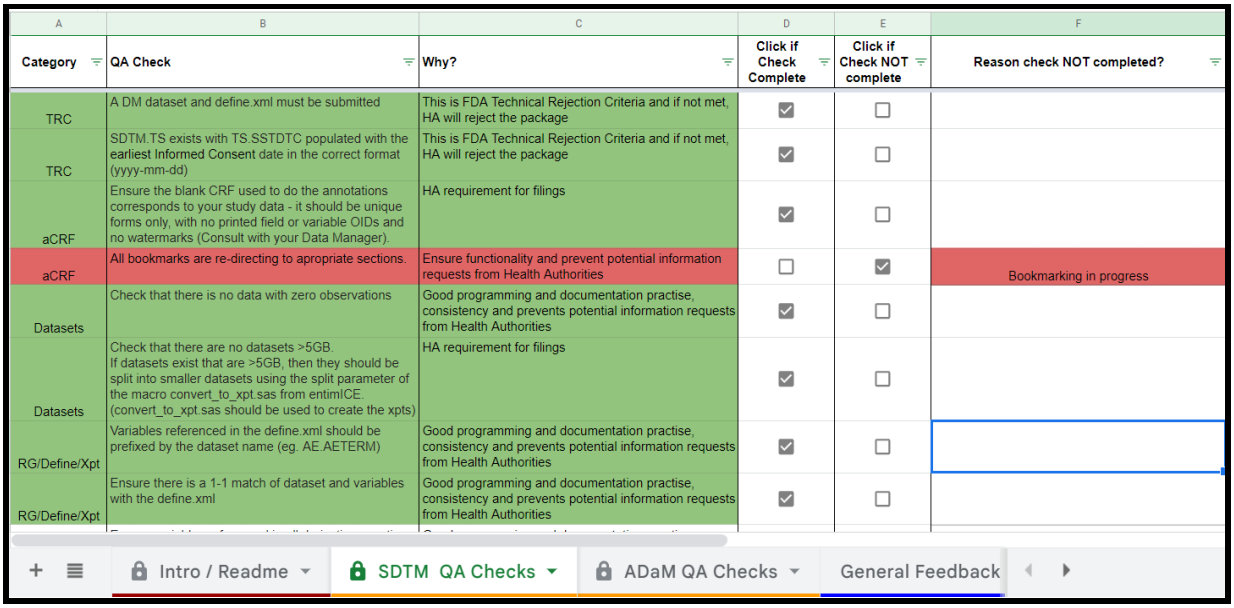


Figure 2. Timeframe for generating FAIR and Submission Deliverables

QUALITY ASSURANCE (QA) CHECKS

In order to address the common issues repeating across studies, molecules and therapeutic areas (TA's), resulting in variations in quality of submission deliverables, the QA checks were developed. These checks help teams validate their own submission deliverables and produce good quality packages. The QA checks are embedded within an excel file with a tab each for SDTM and ADaM related deliverables; with a column to check mark or add comments during the validation process. These checks serve as reminders to improve quality. The checks will help reduce the back and forth between study teams and eSub team reviewers. Implementing them right from the very beginning could help save resources be it time or personnel.

Below is a screenshot of the QA checks that help understand the issues to be looked into from a quality perspective in SDTM data package. It lists out contents to be checked in order to ensure the deliverables adhere to Health Authority requirements and will avoid failing any Technical Rejection Criteria



A	B	C	D	E	F
Category	QA Check	Why?	Click if Check Complete	Click if Check NOT complete	Reason check NOT completed?
TRC	A DM dataset and define.xml must be submitted	This is FDA Technical Rejection Criteria and if not met, HA will reject the package	☑	☐	
TRC	SDTM.TS exists with TS.SSTDTC populated with the earliest Informed Consent date in the correct format (yyyy-mm-dd)	This is FDA Technical Rejection Criteria and if not met, HA will reject the package	☑	☐	
aCRF	Ensure the blank CRF used to do the annotations corresponds to your study data - it should be unique forms only, with no printed field or variable OIDs and no watermarks (Consult with your Data Manager).	HA requirement for filings	☑	☐	
aCRF	All bookmarks are re-directing to apropriate sections.	Ensure functionality and prevent potential information requests from Health Authorities	☐	☑	Bookmarking in progress
Datasets	Check that there is no data with zero observations	Good programming and documentation practise, consistency and prevents potential information requests from Health Authorities	☑	☐	
Datasets	Check that there are no datasets >5GB. If datasets exist that are >5GB, then they should be split into smaller datasets using the split parameter of the macro convert_to_xpt.sas from emlmICE. (convert_to_xpt.sas should be used to create the xpts)	HA requirement for filings	☑	☐	
RG/Define/Xpt	Variables referenced in the define.xml should be prefixed by the dataset name (eg. AE.AETERM)	Good programming and documentation practise, consistency and prevents potential information requests from Health Authorities	☑	☐	
RG/Define/Xpt	Ensure there is a 1-1 match of dataset and variables with the define.xml	Good programming and documentation practise, consistency and prevents potential information requests from Health Authorities	☑	☐	

Intro / Readme | SDTM QA Checks | ADaM QA Checks | General Feedback

Figure 3. Screenshot of SDTM QA Checks

While the screenshot highlights the SDTM checks, similar ones are available for ADaM data and related deliverables as well (**ADaM QA Checks** in screenshot above).

F2 DASHBOARD (F2DB)

As part of our submission activities to the Health Authorities, we had to evolve our approach to create and deliver CDISC compliant eSub data packages. While the readiness and QA checks were helping teams develop quality packages, we never really could gauge or assess the progress of eSub activities. We needed a tangible way to help track progress and identify areas to prioritize and focus on. To address this issue the F2 Dashboard was developed.

The F2DB is an R Shiny app developed internally by Roche-Genentech eSub team. It helps study teams assess their progress towards being FAIR and submission ready by DBL. We first prototyped this tool in excel before making it an R shiny app. The app creates a color coded visual to demonstrate how the study is progressing.

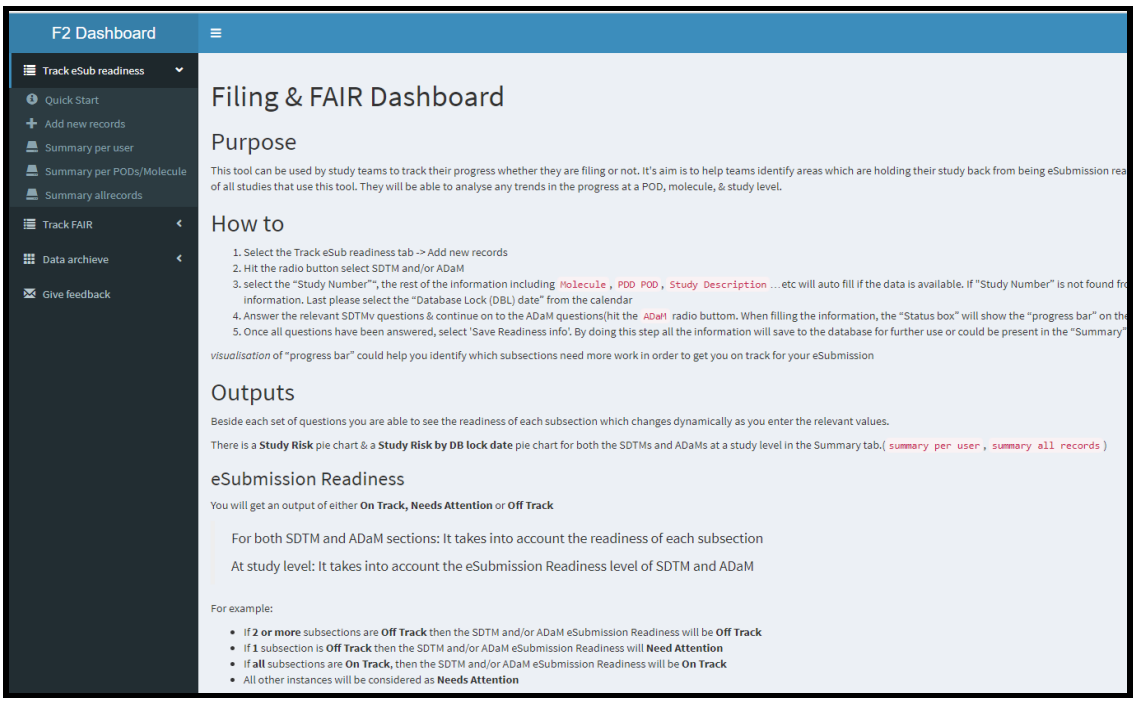


Figure 4. Screenshot of FAIR and Filing Dashboard (F2DB)

The app requires teams to input study information and further provide details on the deliverables. At Roche-Genentech the app has been linked to an internal portfolio report to get basic study information like study name, study number, and milestone dates. This along with study specific deliverable information helps generate a color coded report as seen in the screenshot below (Figure 5). The color coding is determined by a backend formula that evaluates the progress of the deliverables relative to the DBL date and assigns colors accordingly. The colors indicate if you are on track, off track or need attention.

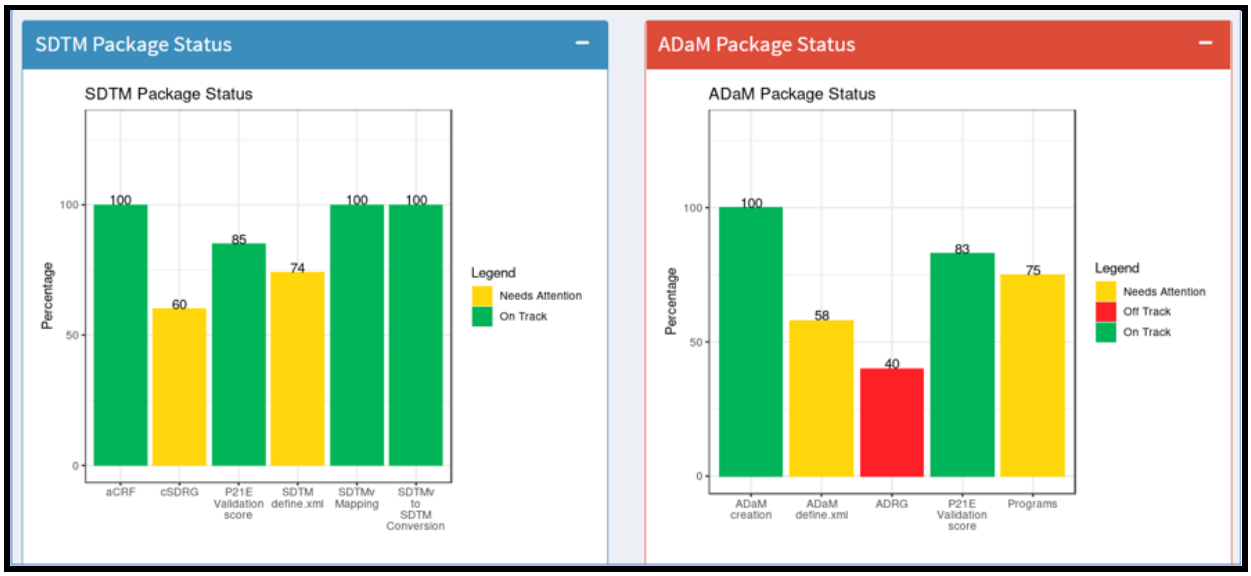


Figure 5. Example - F2DB Color Coded Report

READINESS – TOOL INPUT AND METRICS

The F2 dashboard consists of windows that require user input. Described below are details about each of them

STUDY INPUT or READINESS CHECK

This panel displays basic study information. It is linked with Roche-Genentech's internal portfolio report that helps populate the basic study information like study number, molecule, study description, and DBL date. Users can check the readiness of either SDTM or ADaM by choosing one of them. The orange panel in the screenshot below displays the study input information.

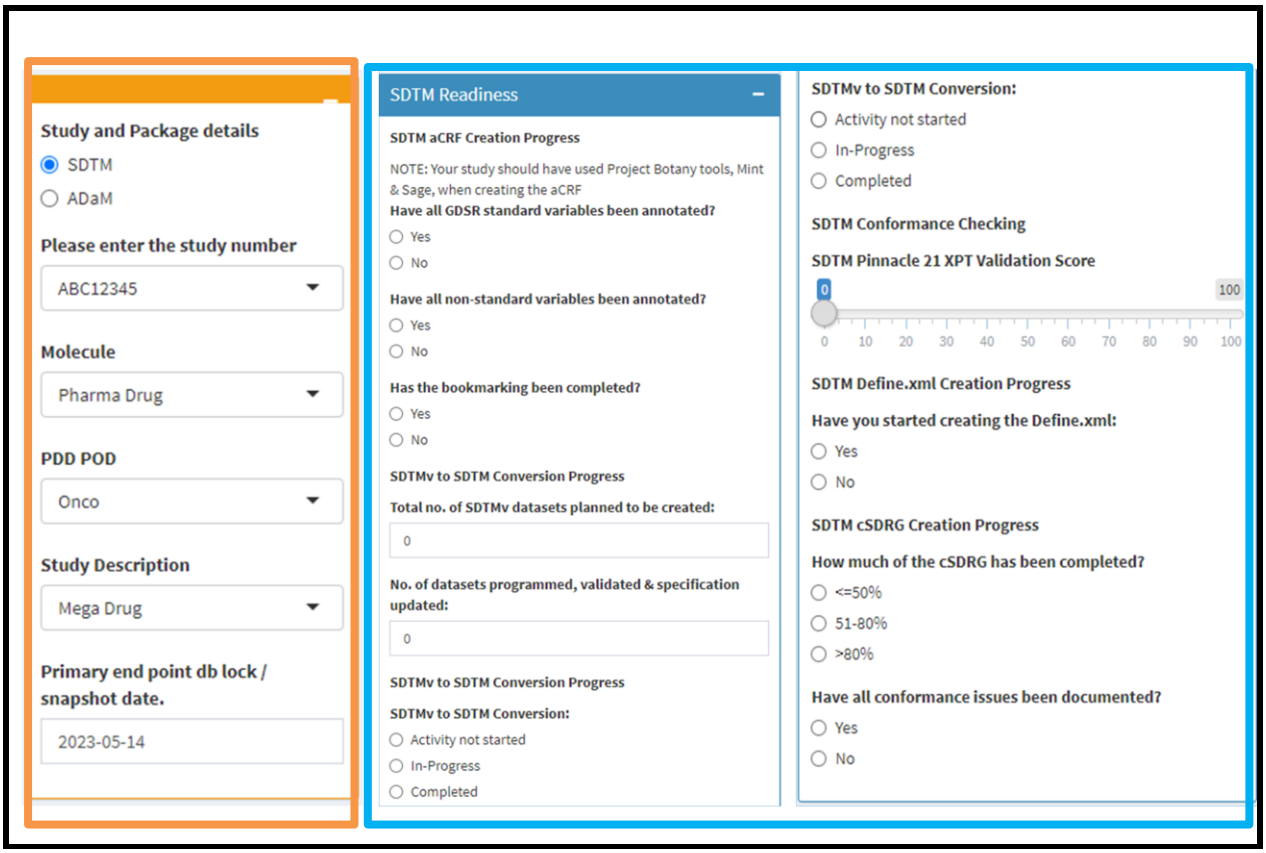


Figure 6. F2DB – Input information panels screenshot

SDTM / ADaM READINESS INPUT

This panel includes a series of questions about the status and extent of completion of deliverables that contribute to the SDTM and ADaM packages respectively. Users would need to provide input on each of the listed fields. Alternatively, they could enter all the relevant information into an EXCEL file and use the upload functionality to input the information for generating the reports. The above screenshot in blue shows this panel's contents (readiness criteria of the FAIR and Filing).

OVERALL PACKAGE STATUS

Once the study and deliverable status input is provided in the previous two panels, an OVERALL package status report is generated and displayed in the SDTM/ADaM Status package panel. A color coded visual showing the status of readiness of each of the fields suggested in the readiness panel is output. The report visually provides tangible evidence on the progress of tasks in relation to the DBL date. It highlights areas where -

- **Progress is good (ON-TRACK)**
- **Average (NEEDS ATTENTION)**
- **Poor (OFF-TRACK)**

The screenshot below displays the Overall status of the SDTM and ADaM deliverables for an example study.

Figure 7. Example - F2DB Overall SDTM and ADaM Status Report

The tool can be used multiple times, as many time as the team would like to gauge their progress during the study life cycle. Managers and study / programming leads could look at the overall status report to understand the progress and address potential road blocks to help the teams be FAIR and filing ready in time. In addition to studies, summary level reports could be generated and reviewed from a molecule perspective to understand patterns and trends.

QR CODE to F2DB

Please find below a QR Code that will give you access to try out the F2DB that has been developed by Roche-Genentech

PLANNED FUTURE ENHANCEMENTS

While the F2DB is completely functional and is helping us gauge the progress of our deliverables, we intend to evolve this tool to make it more user friendly based on the initial feedback received. We plan to add more options to improve the visualizations, bring in some automation to avoid manually entering fields, and also have a similar dashboard for non-filing studies, to name a few.

CONCLUSION

The F2 dashboard is the first of its kind within our function that will help teams gauge their progress, a tangible way to assess the completeness of their deliverables at any point of time. The dashboard is particularly helpful to know how the team is fairing, understand the bottlenecks and triage the issues that the team is facing due to resourcing, time or any other constraints. The color coded user friendly visuals help understand areas to focus on. Timely and regular use of F2DB, along with the rest of the tools listed, could help avoid panic situations and last minute fire-fighting and ensure creation and delivery of good quality packages.

ACKNOWLEDGEMENTS

The authors would like to thank the Analytical Data Sciences eSub Chapter and management teams from Roche-Genentech, for their advice and support on this paper/presentation.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Bidhya Basnet
Roche-Genentech, 1 DNA Way, South San Francisco, CA - 94080
e-mail: basnet.bidhya@gene.com

Dyuthi Yellamraju
Roche-Genentech, 1 DNA Way, South San Francisco, CA - 94080
e-mail: yellamraju.dyuthi@gene.com