

Automated review of metadata standards, data checks, and trend visualization.

AD11

Giuseppe Di Monaco,
UCB BioSciences GmbH, Monheim, Germany



Inspired by **patients.**
Driven by **science.**

INTRODUCTION

PROBLEMS IN CLINICAL TRIALS PROCESSES

- Professionals engaged in **repetitive manual tasks**
- **Silos** due to **disconnected metadata** and different technologies
- High Drug **cost** and continue to rise

AUTOMATION INITIATIVE

- To **improve standards**, data **access**, data **analytics & visualization**
- To allow staff to have a **leaner, higher-level** footprint

SOLUTION - A LARGE AUTOMATED PROCESS

- **Data review**
- **Metadata** collection and **standardization**
- Data **checks**, trend **visualization** and **notifications**

OBJECTIVES, METHOD & OVERVIEW

OBJECTIVES – PRIMARY AND SECONDARY

- Provide a **guide**
- **Help prevent** system becoming **unmaintainable**

METHOD – INCREMENTAL MODEL

- **Incrementally** built
- **Components** and **modules** designed and built separately
- Code **refactoring** for improving software **architecture**

OVERVIEW – 3 MAIN PRODUCTS

- **Automated** study **report** (**metadata** standards & data **checks**)
- **Trend visualization** dashboard
- **Knowledge** distribution website

OVERVIEW – 3 MAIN PRODUCTS

4

1 STUDY REPORT

- SUPPORT & CREATE METADATA STANDARDS, DATA CHECKS

2 Qlik Sense

- TRENDS VISUALIZATION

3 SHAREPOINT

- KNOWLEDGE DISTRIBUTION

1 Output Report

Data Reviewer Test Results XXXXX							
ID	Module	Data Type	Test Results	Test Status	Grade	Custom Configuration	Number of Findings
28	DRPOSTLASTDISP	SOTM	Later dates than the defined reference date have been found. An example is for patient: XXXXX-YY-ZZZZZ in dataset: SD.SI for variable: S1STOTC which has the issue: 2019-04-23 is greater than 2019-02-27 (Listing created with 2 examples)	FAILINGS	Recommended		2
31	DRDATESEQ	SOTM	Module was not run because it did not pass prerequisite checks: EXHNOTC not on SD.EX. A single secondary	NOT RUN	Recommended (TARGOT gt	Condition used:	
34	DRCHECKVAROBS	ADAM	Module had no findings to report	NO FINDINGS	Recommended		

[See Test Specifications](#)

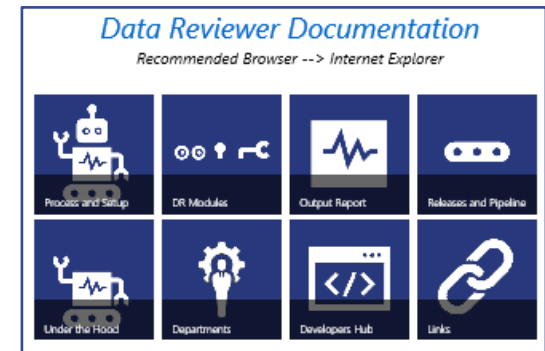
For any queries please contact [UCB Tools and Automation team](#)

Results | TrialSummary | Specs | drlist1 | drlist2 | drlist4 | d ...

2 Qlik Sense Dashboard

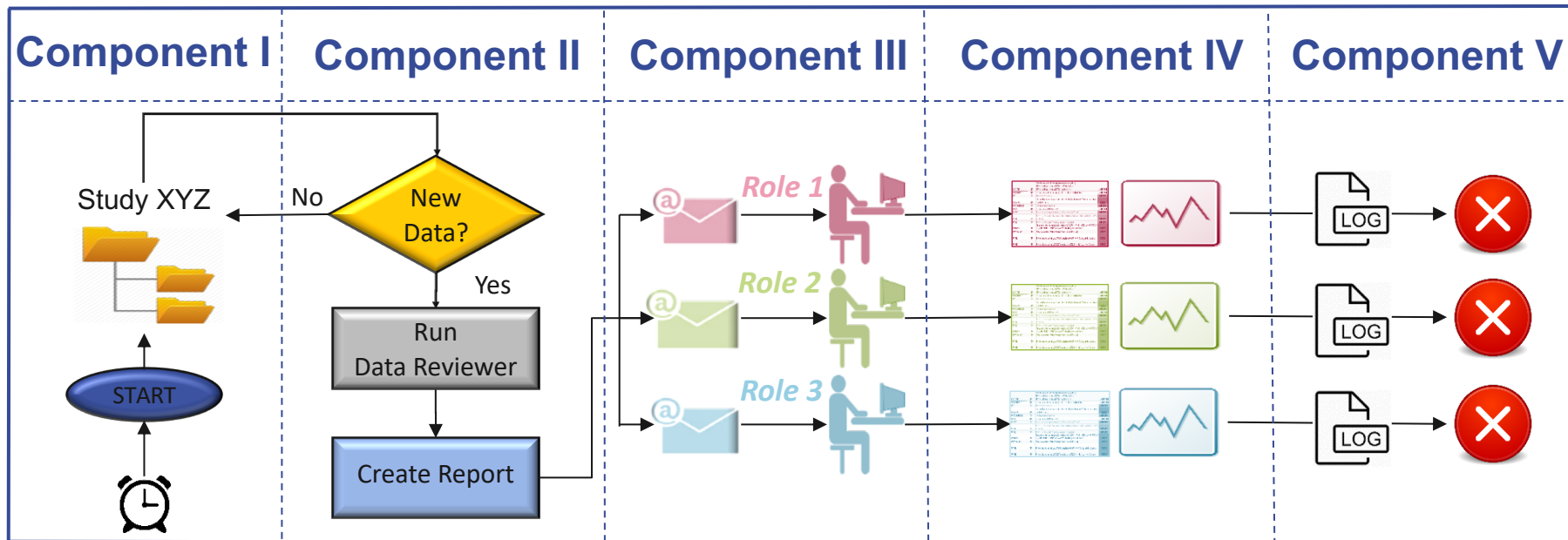


3 SharePoint Site

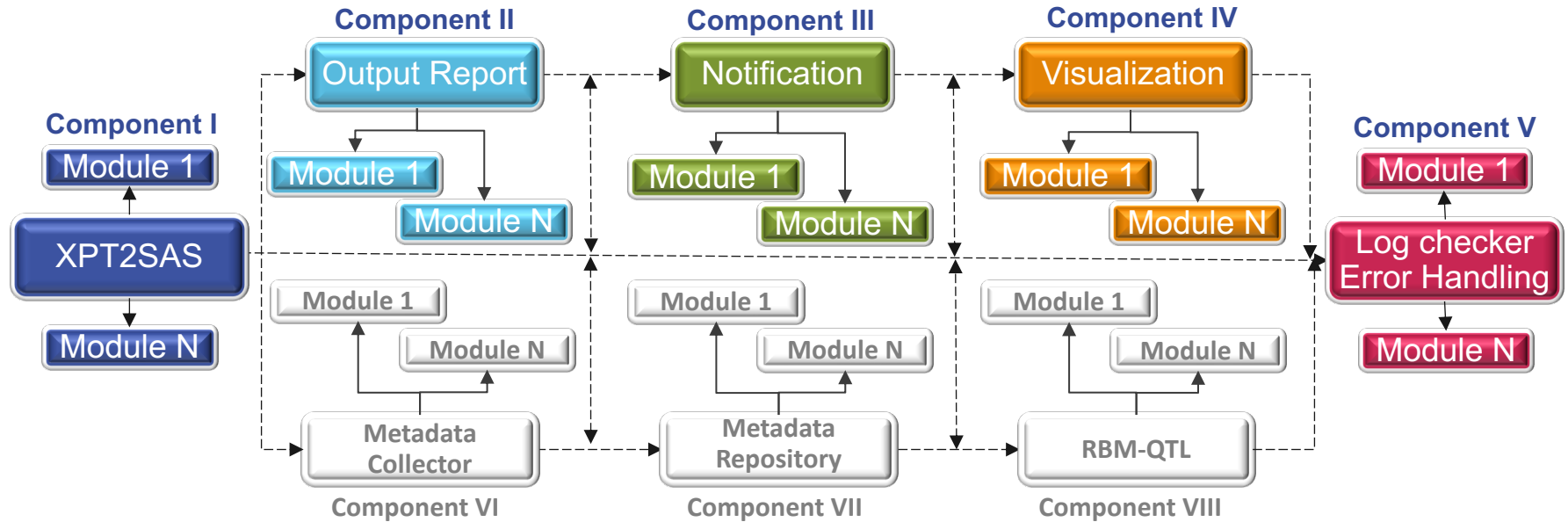


PROCESS FLOW

- Component I** - Data Loading
- Component II** - Data Review
- Component III** - Notification
- Component IV** - Visualization
- Component V** - Logs & Error Handling



PROCESS FLOW - HIERARCHICAL AND SEQUENTIAL DESIGN



Component I: Automatic unzip and extract

Component II: Problems are reported directly to the consumer

Component III: Automated and targeted selection of email notification

Component IV: Trends visualization of data and metadata issues

Component V: Log checker and error handling

Component VI: Creation and maintenance of metadata standards

Component VII: Bulk load of metadata in to MDR

Component VIII: Risk Base Monitoring (RBM), Quality tolerance limits (QTL)

ARCHITECTURE - CONTINUED

COMPONENT I - XPTTOSAS

- Parse **each study** folder and extract SAS datasets from **xpt** files
- A **service account** runs **daily**
- Creates a **report**

STUDY	PATH	STATUS	COMMENT	RUN_DATE
-------	------	--------	---------	----------

COMPONENT II – DATA REVIEW

ID	Module	Data Type	Test Results	Test Status	Grade	Number of Findings
33	DRIMPORTDDT	ADAM	DDT was successfully imported from \ddt\adam\YYYYY SSD ADaM Specification.xlsx. Metadata issues found e.g. Variable SRAGEONS has a bad format assigned Please note that transcoding issues were found - this is typically caused by special characters in input files.	FINDINGS	Recommended	16
66	DRTEPARMS	ADAM	No time-to-event parameters were detected in non-ADTTE datasets	NO FINDINGS	Recommended	0
2	DRIMPORTDEFINE	SDTM	Module was not run because it did not pass prerequisite checks: No DEFINE candidates were defined or found in the expected location. For transfer data this is [spi\SDTM data\working]	NOT RUN	Recommended	

- **Report** is **simple** and **easy** to understand
- **Easy** navigation within the **file** and to the **SharePoint** site
- **Metadata** check are **connected**

Component I

Module 1

XPT2SAS

Module N

Component II

Output Report

Module 1

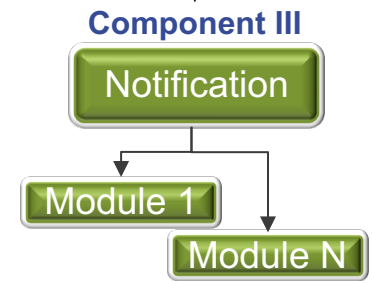
Module N

ARCHITECTURE - CONTINUED

8

COMPONENT III - NOTIFICATION

- Includes 3 main parts.
 1. Component **Settings**
 2. e-mail containing output **report** for each study
 3. e-mail containing process **logs** for each study

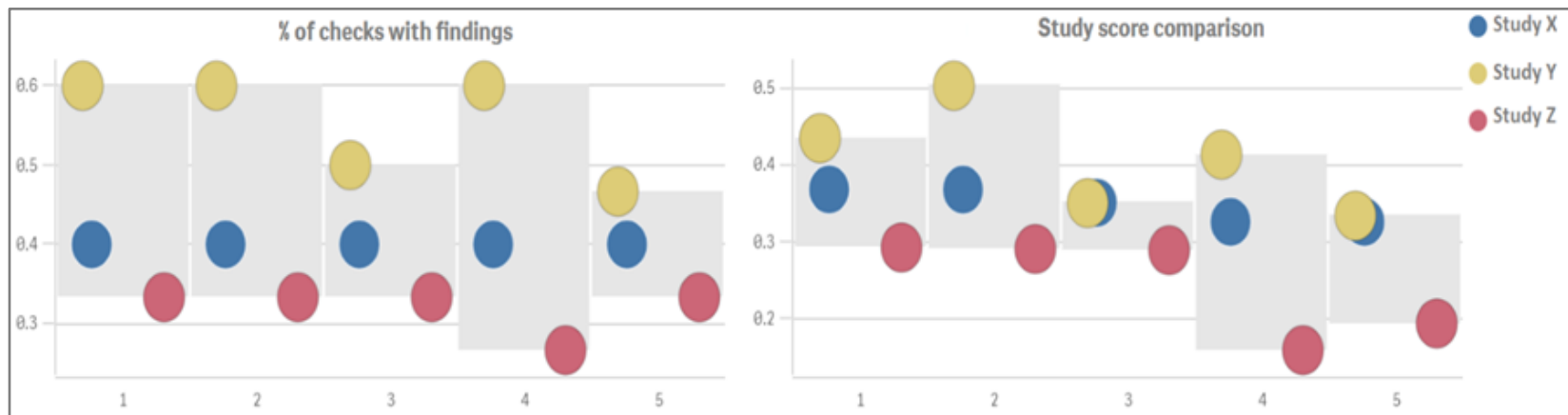
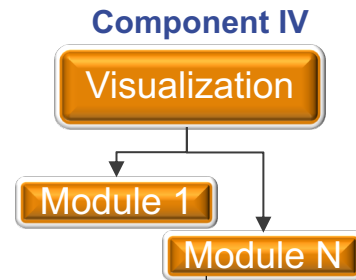


1 - Specification	2 - Report sent via e-mail	3 - Log sent via e-mail
PROJECT_ID	 data_reviewer_SDTM_ADAM_201908... 140 KB	 Logreport_20190826.txt 338 bytes
STUDY_ID		
RECEIVER		
EMAIL SUBJECT		
BODY TEXT	Dear Receiver This is an automated email sent on behalf of Tools and Automation Standard Team. Please review the attached	Dear Receiver This is an automated email sent on behalf of Tools and Automation Standard Team. Please review the attached
EMAIL ATTACHMENT		
OUTPUT		
COMMENTS	Best Regards, Tools and Automation Standard Team	Best Regards, Tools and Automation Standard Team

ARCHITECTURE - CONTINUED

COMPONENT IV – VISUALIZATION

- Dashboard generated from **reports** helps **identify**:
 - Significant **trends** and **deviations**
 - **Hidden** patterns
 - Detect important **exceptions**
 - Study **Y** performed **worse** than study **X** and study **Z**
 - NOTE: scores were **normalized** (between 0-1)

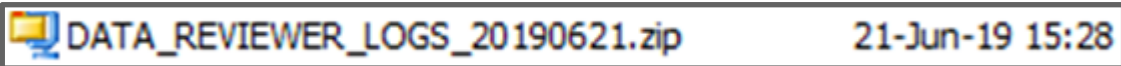


ARCHITECTURE - CONTINUED

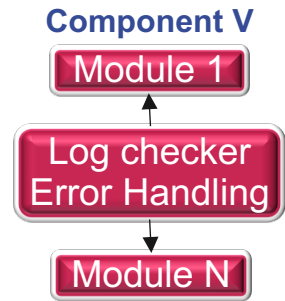
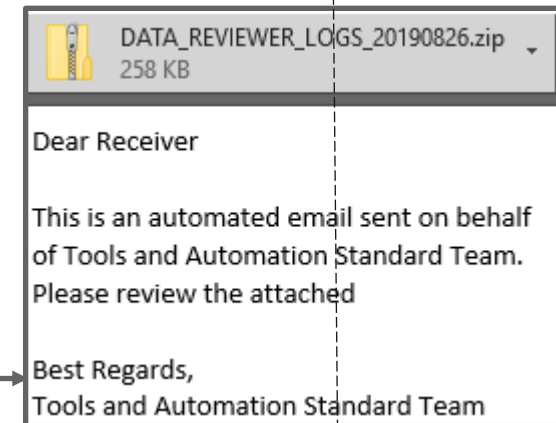
10

COMPONENT V - LOGS AND ERROR HANDLING

- Run as **last** component
- **Parses all** module's log
- **Zips all logs** under the project and study folders



- Creates a **summary** of all logs
- If **error/warning** then call Component III

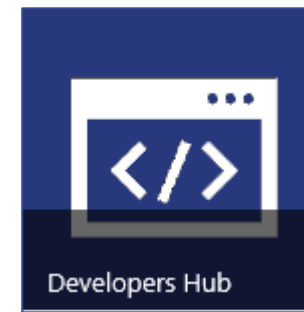
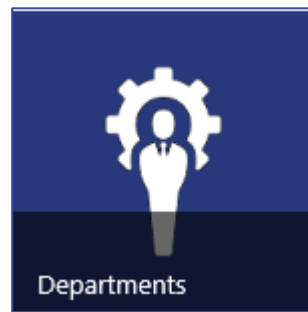
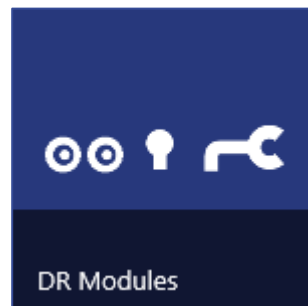


SHAREPOINT SITE

- **Internal** and **external** documents
- **“Three-click-rule”**: information reached within three clicks

Welcome to the Data Reviewer Homepage!

Automated review of metadata standards, data checks, and trend visualization



SHAREPOINT SITE

- Releases and pipeline

Welcome to the Data Reviewer Homepage!

Automated review of metadata standards, data checks, and trend visualization

Releases and Pipeline

Versions

In 2019, 4 versions are scheduled to be released, one for each quarter. Click on the links to navigate to the designated version.

Up to Q4-2018

Q1-2019

Q2-2019

Q3-2019

Q4-2019

[Old Version](#)

[Pre-Alpha](#)

[Alpha](#)

[Beta](#)

[Stable](#)

- **Pre-Alpha:** to **gain confidence** of management and stakeholders
- **Alpha:** **new modules**, white-box and black box **testing**
- **Beta:** **new components** and bug fixing
- **Stable:** **End of first development phase.** Process is **fully automated**

MAINTAINABILITY

PROBLEMS - LARGE PROJECTS

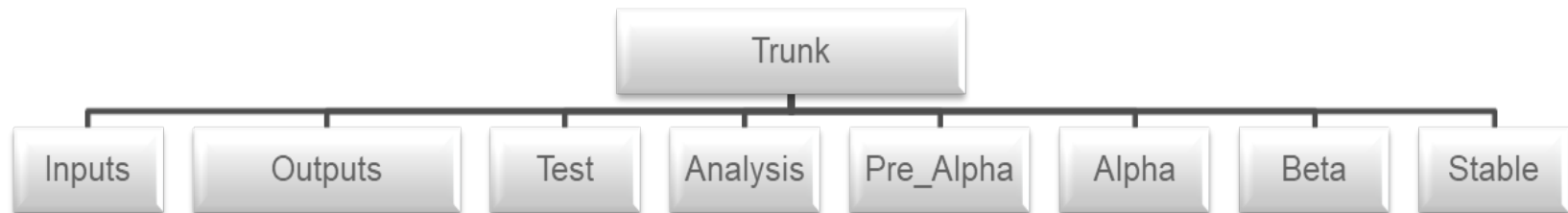
- **Risky, complex** and need significant **budget**
- Require **structure** and **effective** management
- **High maintenance when** new developers join or new components are added
- **High %** of the lifecycle cost for **corrective, adaptive and perfective maintenance**

SOLUTION - CODE REFACTORING

- Prevents **redundancy**, increases **maintainability**
- Facilitates **hierarchical and sequential** architecture
- **Benefit** of future **maintenance** done at any level by any **user**

WHAT TO MAINTAIN

1. **Process and Architecture** (Programs, configuration, etc....)
 - **Structured** and Intuitive **Architecture** requires minimal **maintenance**



2. **SharePoint** Site (Any type of document is in SharePoint)
 - Documents are **automatically** created in “.aspx” and loaded on the site

SYSTEM'S USERS

1. **Regular** Users
2. System **Developers**

MAINTAINABILITY – CONTINUED

4 LEVELS OF MAINTANANCE

1. **Process Run** (Study specific settings)
2. **Process configuration** (Architecture and process flow)
3. **Components** (Integration of software components)
4. **Modules** (Integration of software Modules)

3 LEVELS OF TESTING

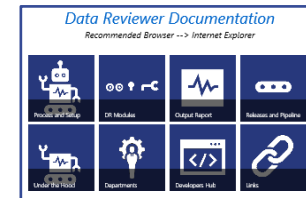
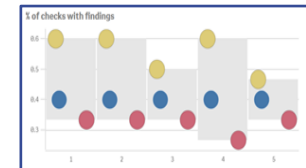
1. **White Box** (Developer's testing)
2. **Black Box** (Full QC by study leads)
3. **Regression Testing** (Full automated Regression Testing)

BENEFITS OF THE 3 PRODUCTS

1. **Automated study report**
 - Speed with Data review & Metadata standards
2. **Trend Visualization**
 - Trends, deviations, exceptions, Hidden patterns
3. **SharePoint Site for knowledge distribution**
 - Information reached in 3 clicks



Test ID	Test Name	Test Type	Test Date	Test Status	Test Results
1	Test 1	Unit Test	2010-10-10	Passed	Test 1 passed
2	Test 2	Integration Test	2010-10-10	Failed	Test 2 failed
3	Test 3	System Test	2010-10-10	Not Run	Test 3 not run



LEVELS OF MAINTAINABILITY & TYPE OF USERS

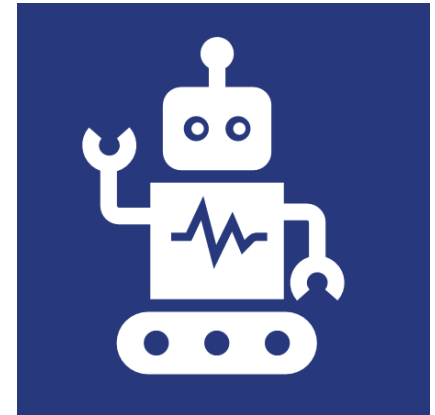
1. Process run, Process configuration, Component, Module
2. Regular users and expert developers

3 LEVELS OF TESTING

1. **White Box** (Developer's testing)
2. **Black Box** (Full QC by Study Leads)
3. **Regression Testing** (Automated)

CONCLUSION

This is an unattended Bot running on a server and sending **notifications** with output **reports**



Professionals are not engaged in **repetitive manual** tasks

Metadata not disconnected

Increased quality and efficiency

No learning curve for the end-user

Reduced cost of Drug

Questions?

Thanks!