

AD10

Data Catalogues – The new Showman to organize your data, A CRO perspective

Limna Salim

India Head – Biometrics & Operations, Genpro Research

limna.salim@genproresearch.com

Mohin K Mohanan

Data Analyst II, Genpro Research

mohin.mohanan@genproresearch.com

DATA CATALOGUES – THE 3 W's

WHAT??

- Library of your Organization's data.
- Neatly Organized & Indexed.
- Data spread across sources.

WHY??

- Clarity
- Accuracy
- Speed
- Data Security

WHERE??

- Data Scientists Across Various Industries
- Healthcare
- CRO

DATA CATALOGUES IN CRO

- Requirements vary based on individual CRO's and across functionalities.

- ❖ Statistics & Statistical Programming
- ❖ Medical Writing & Research scientists
- ❖ Data Analytics & Data Management
- ❖ HR & QA Team

- This paper covers 4 different Use cases across a CRO

- ❖ Macro Library
- ❖ Standard Study Repository
- ❖ Standards & Guidelines Repository
- ❖ Statistical Programming library

BUILDING A DATA CATALOGUE

- Numerous organizations provide tools that develop, and curate catalogues based on your needs.
- JavaScript, Python, PHP, C# etc offers plenty of frameworks and supports several platforms
- Constraints –
 - Continuous dependency on developers
 - Lack of other language awareness
 - Associated Cost

BUILDING A DATA CATALOGUE USING R

Prerequisites:

- ❖ Install latest R version.
- ❖ Setting a Predefined theme
- ❖ Collating Metadata and data to be presented
- ❖ Setting up study folder or Github repository

Packages to be installed:

- ❖ Bookdown
- ❖ RMarkdown

ELEMENTS IN DATA CATALOGUE

1. CSS

CSS

- Cascading Style Sheet
- Theme of your catalogue
- Controls layout of multiple pages at once.

2. RMD

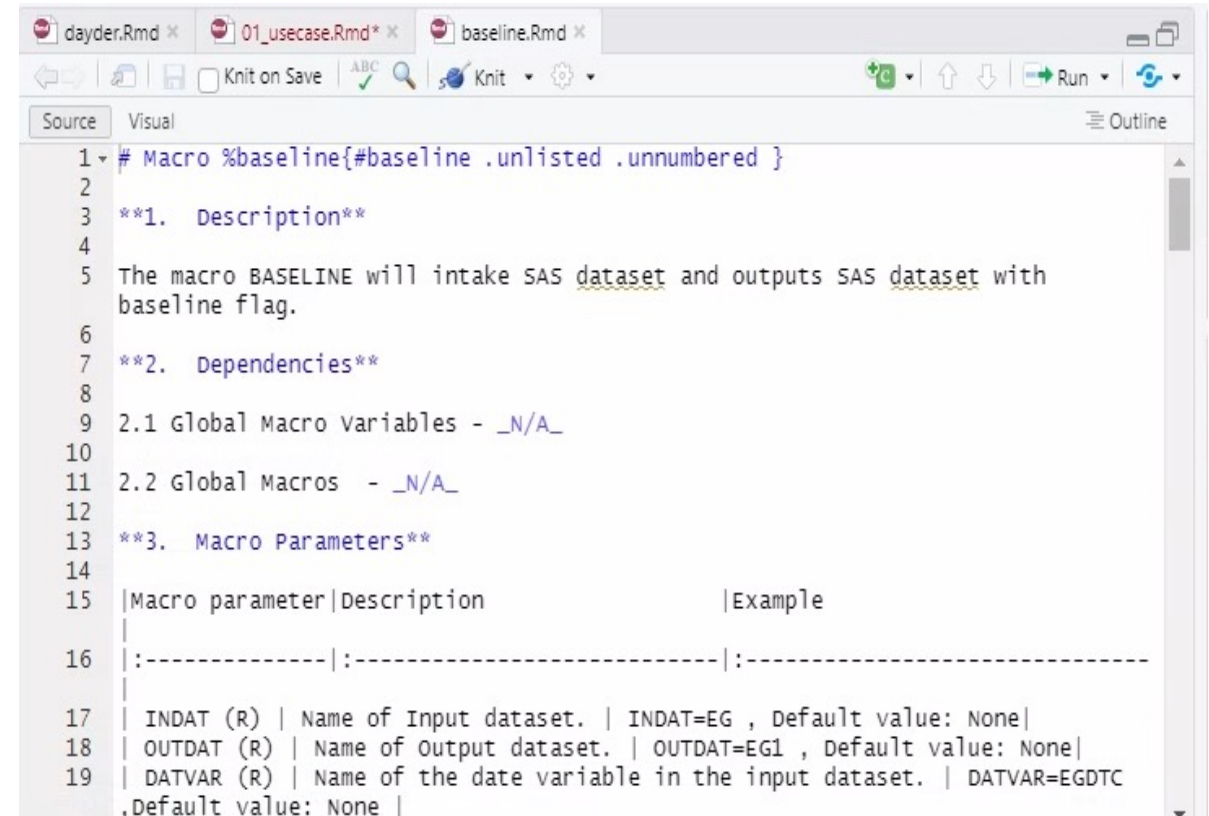
3. BOOKDOWN

```
.wrapper .sidebar{  
  background: rgb(5, 68, 104);  
  position: fixed;  
  top: 0;  
  left: 0;  
  width: 225px;  
  height: 100%;  
  padding: 20px 0;  
  transition: all 0.5s ease;  
}  
  
.wrapper .sidebar ul li a{  
  display: block;  
  padding: 13px 30px;  
  border-bottom: 1px solid #10558d;  
  color: rgb(241, 237, 237);  
  font-size: 16px;  
  position: relative;  
}
```

ELEMENTS IN DATA CATALOGUE - Contd

RMD

- R Markdown files for the main Data Feed
- Notebook interface for R
- 'rmarkdown' package
- Just "Knit" & produce a report in multiple formats like HTML, PDF, MS Word
- Implement text, code, graphs, all in one interactive report
- Enables reproducibility

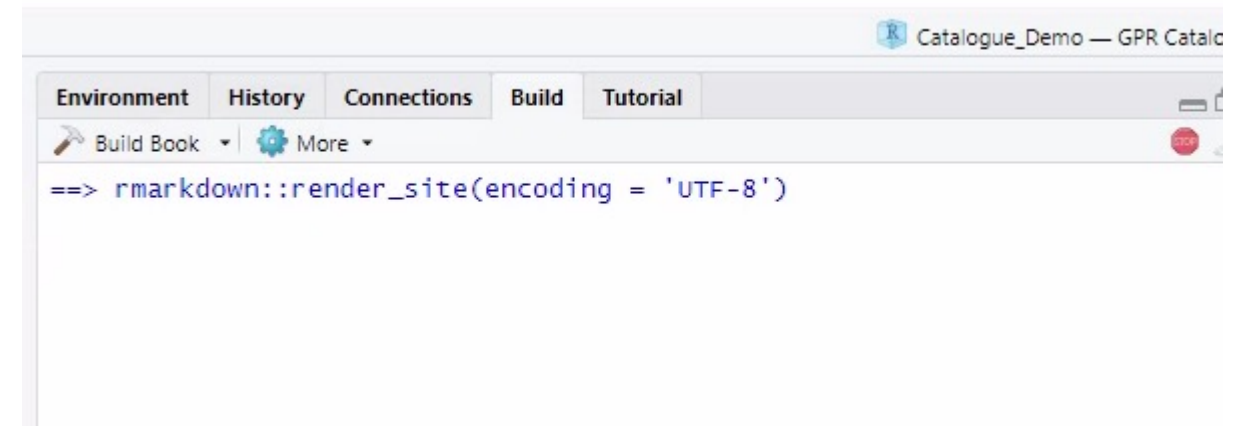


```
1 # Macro %baseline{#baseline .unlisted .unnumbered }
2
3 **1. Description**
4
5 The macro BASELINE will intake SAS dataset and outputs SAS dataset with
6 baseline flag.
7
8 **2. Dependencies**
9
10 2.1 Global Macro Variables - _N/A_
11
12 2.2 Global Macros - _N/A_
13
14 **3. Macro Parameters**
15
16 |Macro parameter|Description|Example|
17 |-----|:-----|:-----|
18 | INDAT (R) | Name of Input dataset. | INDAT=EG , Default value: None|
19 | OUTDAT (R) | Name of Output dataset. | OUTDAT=EG1 , Default value: None|
20 | DATVAR (R) | Name of the date variable in the input dataset. | DATVAR=EGDTC
21 ,Default value: None |
```

ELEMENTS IN DATA CATALOGUE - Contd

BOOKDOWN

- R package that facilitates long-form reports with R Markdown
- Just “Build” & produce a book
- `render_book()` function
- Supports multiple output formats
- Integrated with the RStudio IDE
- Enables dynamic graphics and interactive applications
- Easy publishing to GitHub



USE CASE 1 – MACRO LIBRARY

Glossary of in-house macros with information like what each macro does and its outputs & dependencies.

Macro categories and macros with further drill downs leading to details

Features implemented–

- ❖ Macro codes displayed in a box and scrolling down/sliders incorporated.

Future scope of expansion in this use case include –

- ❖ Enabling Download option for the macro codes.
- ❖ Long term scope to have the macros executed from the Catalogue page.

USE CASE 2 – STANDARD STUDY REPOSITORY

Basic study repository with therapeutic areas, phase of study, number of subjects/sites as part of the study, niche areas of contribution.
Useful for reference, QA/Audit purposes etc.

Features implemented–

- ❖ Possibility of importing the excel file into our rmd file

Future scope of expansion in this use case include –

- ❖ Presenting small summary counts like number of studies across phases/TA's

USE CASE 3 – STANDARDS & GUIDELINES

Glossary of standards and guidelines set forward by the FDA/CDISC.
Multiple versions can be represented along with information like version number, release date, links to these documentations etc.

Features implemented–

- ❖ Accounted for tabular representation of data.
- ❖ Provided hyperlinks within the data that takes us to external web links.

Future scope of expansion in this use case include –

- ❖ Hyperlinks can link to data stored within the organization's servers like SharePoint.
- ❖ Automation of Version changes to a particular document.

USE CASE 4 – STATS & PROGRAMMING LIBRARY

Library of code snippets from complex analysis performed by the team saved for future reference and reusability.

Goal to reduce the research time for our team.

Features implemented–

- ❖ Having images or graphs incorporated into the presented details.

Future scope of expansion in this use case include –

- ❖ Hyperlinks to access represented data directly.
- ❖ Enabling download options for the code patch.

FEATURES

1. Search Capability
2. Data Preview
3. Data Privacy
4. Expansion/Upgradation

FUTURE SCOPE

1. Tags/Hyperlinks within server
2. Integrating with other apps
3. Data access from cloud
4. Users provided with data recommendations

HOSTING & CONCLUSION

Host the Catalogue application as a Web Page.

RStudio itself offers different ways to host your application.

- RStudio Connect
- Bookdown.org
- Rpubs
- GitHub

Easy cost-effective solution for a CRO to temporarily kick start the process which can later be improvised and developed.



THANK YOU

Any Questions ?