

DH07 Study Anonymisation: From Request to Delivery

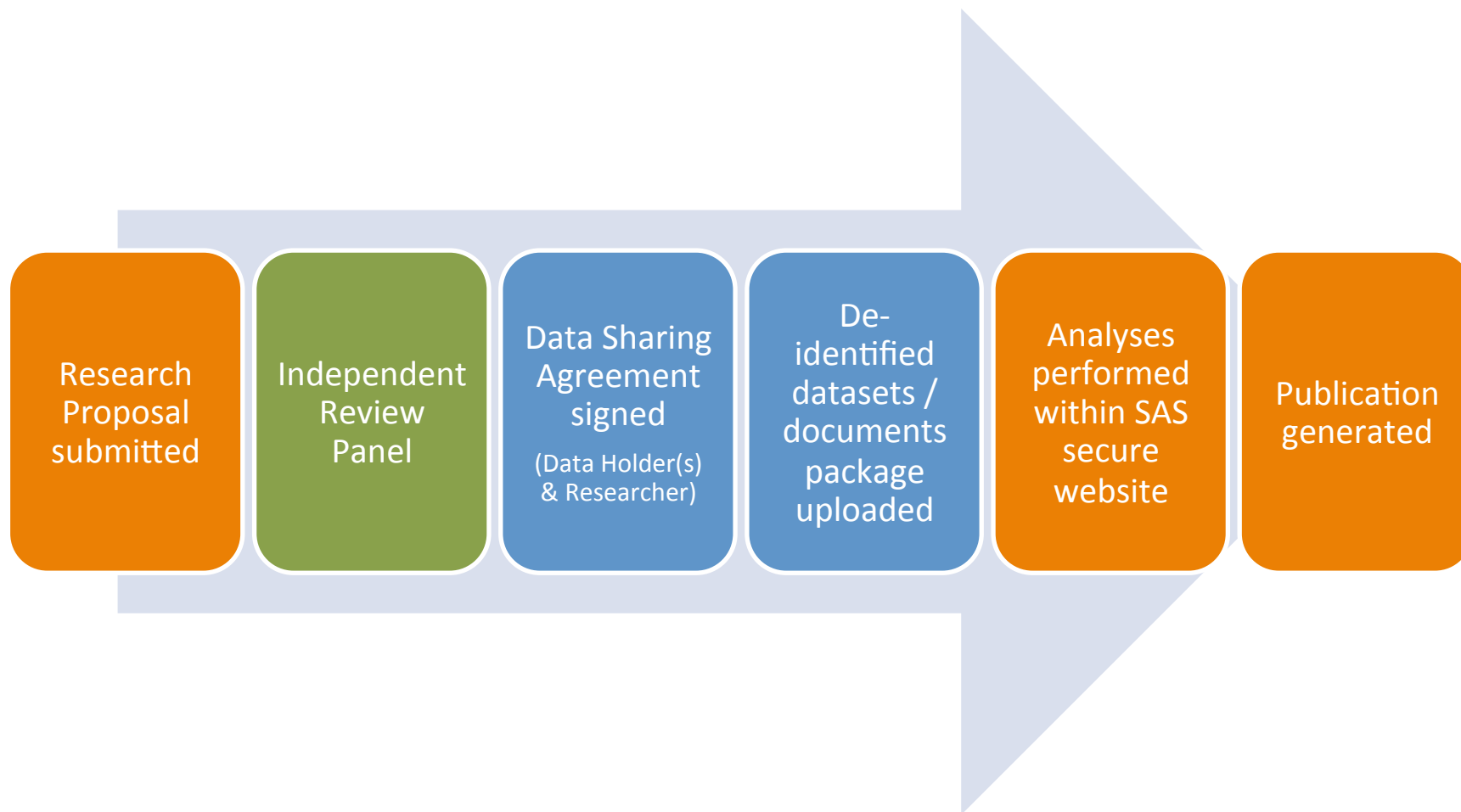
Kelly Mewes, Roche Products Ltd, Welwyn Garden City, UK
Sai Jandhyala, Roche Products Ltd, Welwyn Garden City, UK





- “The opinions expressed in this presentation and on the following slides are solely those of the presenter and not necessarily those of Roche. Roche does not guarantee the accuracy or reliability of the information provided herein.”

Our Data Access Model





Who are the Requestors?

- All academics from universities



- Evaluating new novel research

“Assessing models for changes in (melanoma) tumour size over time and how they relate to survival times”

“Optimized (Pegasys) Treatment of Hepatitis C and Hepatitis B”

“Identifying immune modulating patterns across diseases from open clinical trial data”

- Multi-sponsor Requests.



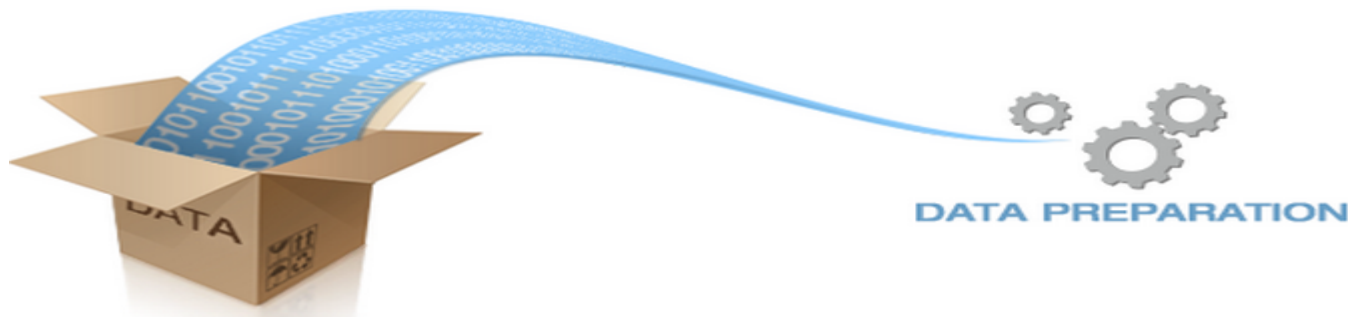
Patient Level Data Sharing Team Operation

- Team
- Remit of group
- Documents and publications
- Governance



Study Data Location and Preparation

- Locating Data
 - Study snapshot to select
 - Electronic submissions
 - Legacy data challenges
 - Other systems for linked documentation
- Provide data in its original state
 - Multiple older data models require preprocessing
 - Able to prepare data packages from multiple areas





Supporting Documentation



- Secure pdf files provided

Dataset
specifications

Annotated Case
Report Form

Protocol

Reporting and
analysis plan

Redacted
Clinical Study
Report

- Anonymisation Orientation Document
- Details all programming applied



Anonymisation

Why do we need it?



- Maintain **patient confidentiality**
- Direct identifiers need to be removed from the data e.g.
 - Subject name, initials
 - D.O.B
 - Address
- Indirect identifiers in combination could lead to re-identification e.g.
 - Age
 - Geographical location
 - Unusual medical conditions
 - Visible characteristics such as race/ethnicity



Anonymisation Process Flow

STEP 1 /2

Read in and unpack source data to PLDS area.
e.g. .xpt or .sas7bdat

STEP 3

Perform any pre-processing required to prepare the source data

STEP 4

Run SAS macro to anonymise SAS datasets

Perform internal independent validation / Code review

STEP 5

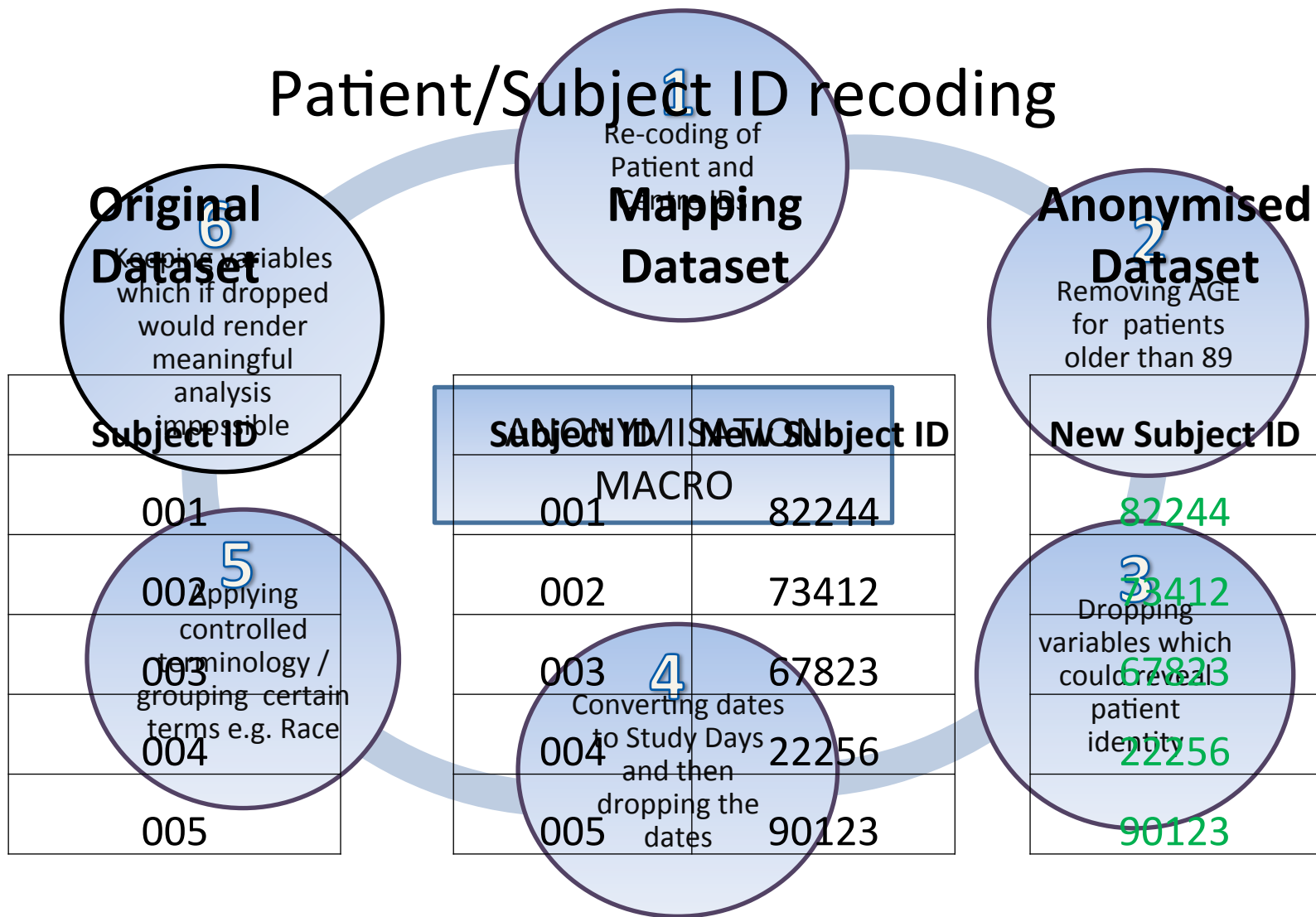
Final anonymised data created ,re moving any link back to the original PT

STEP 6

Datasets transferred to Zip Data package for upload to the SAS system

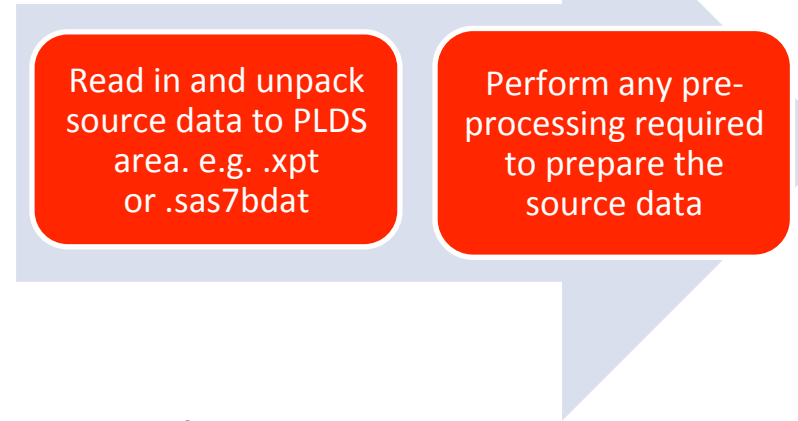
The anonymisation macro

What does it do?





Steps 1 & 2: Prepare Data for Macro



- Copy the study datasets from the original source area to a dedicated anonymisation area
- Unpack datasets and perform pre-processing e.g. date handling
- Data must comply with macro input requirements, such as:
 - Patient and Centre variables must be populated and consistent across selected datasets
 - Dates must be in a compatible SAS format

Run the Anonymisation Macro

Run SAS macro
to anonymise
SAS datasets

```
% ANON_LIBS2(<INLIBS      = libname(s)>
              ,<OUTLIBS    = libname(s)>
              ,<DS         = dataset-list | ALL>
              ,<CENTRE     = variable-list>
              ,<PATIENT    = variable-list>
              ,<DROP       = filename>
              ,<MAPLIB     = libname>
              ,<EXCLUDE    = dataset-list>
              ,<REFDTVAR   = variable-list>
              <DEBUG       = Y | N>);
```



Post Macro Execution

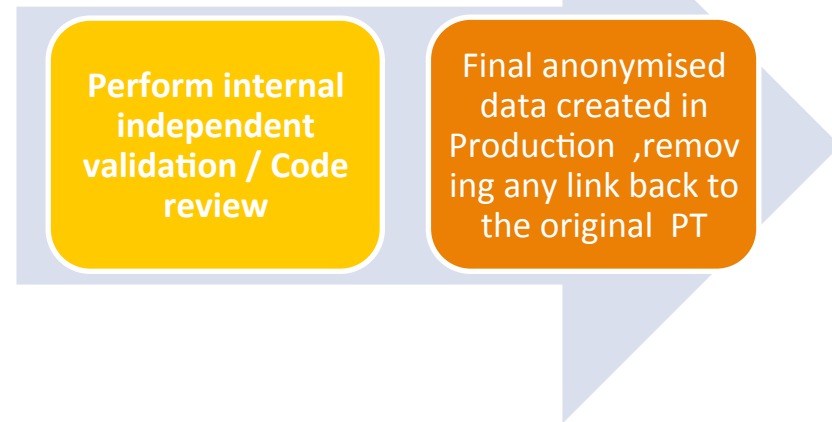


- After the macro has been run:
- The anonymised datasets are produced in a dedicated secure location separate to the source
- A PROC contents output of the anonymised datasets is generated for review
- A summary document is produced to aid review which lists the following:
 - Variables that were removed to ensure anonymisation
 - Blank variables that were removed



Steps 4 & 5:

Independent Validation



- Visual review of anonymised datasets, with a focus on free text fields, timing variables, and direct identifiers
- Review of PROC contents output
- Review of summary document (orientation document)
- Perform code review of the preliminary and anonymisation programs, final run in Production environment



Step 6:

Data Package Creation

Datasets
transferred to Zip
Data package for
upload to the SAS
system

- The data package is created in-line with a template from SAS that contains a standard folder structure as below

Analysis Ready Datasets

Documents

Raw datasets

Initialisation file

- Parameters of the initialisation file

Name=RCH-1085-XXXXXX /*Roche identifier containing Proposal & Protocol ID*/

Description=Study title /*As specified in the protocol*/

SponsorID=RCH-11-2014 /*Default Roche ID value*/

AppendFiles=Yes /*Option if you are adding new files*/

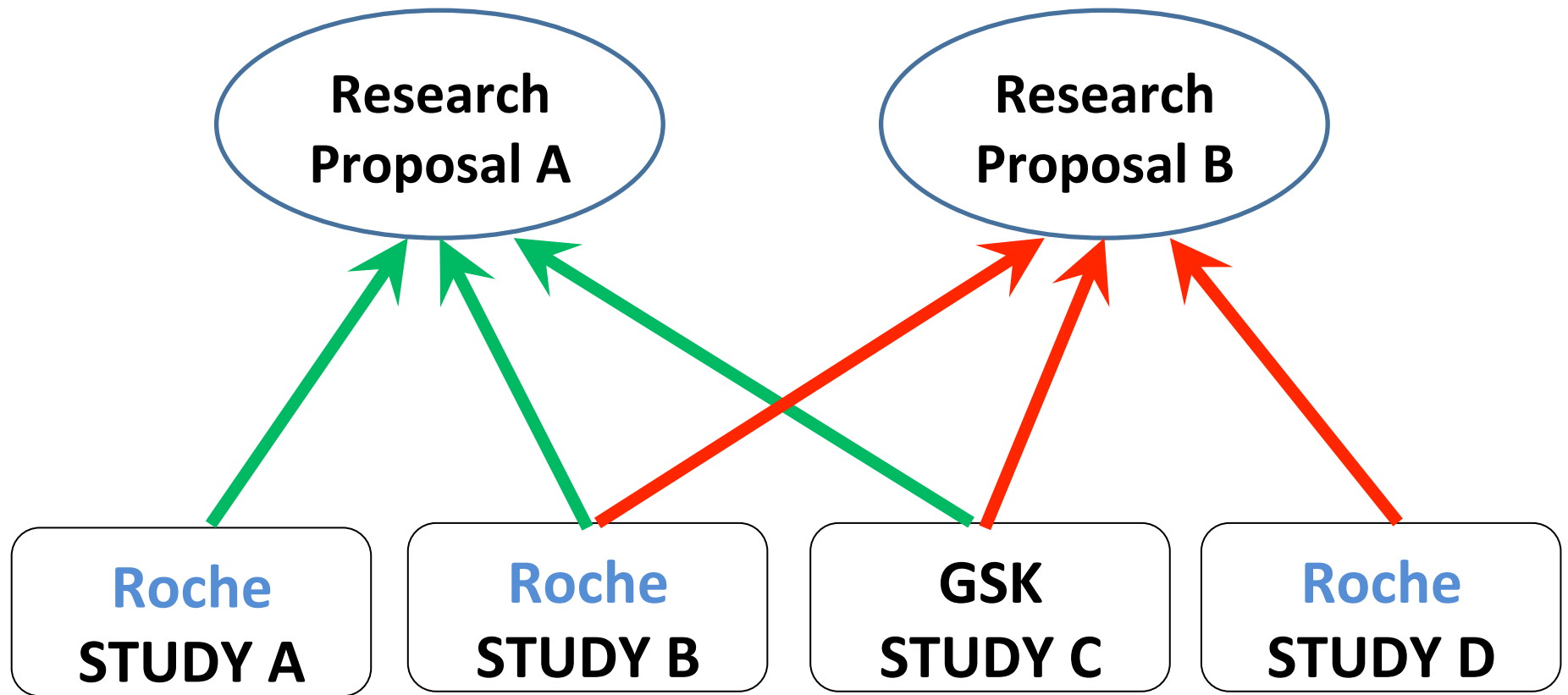


Delivery to the SAS CTDT System

- Data delivered via the SAS CTDT MSE
- Secure area for researchers to access the data
- Enables analyses to be performed in SAS/R
- Pre-requisites:
 - Setup of Study Area(s)
 - Creation of Research Area(s)
 - User account creation and access



A multi-sponsor scenario



Future Plans

- **Proactive anonymisation**
- **PhUSE Data De-Identification Standards pilot.**
- **Automated redaction of free-text variables**
- Increased enabling of **in house** data sharing



Any Questions ?



Kelly.mewes@roche.com

Sai.jandhyala@roche.com