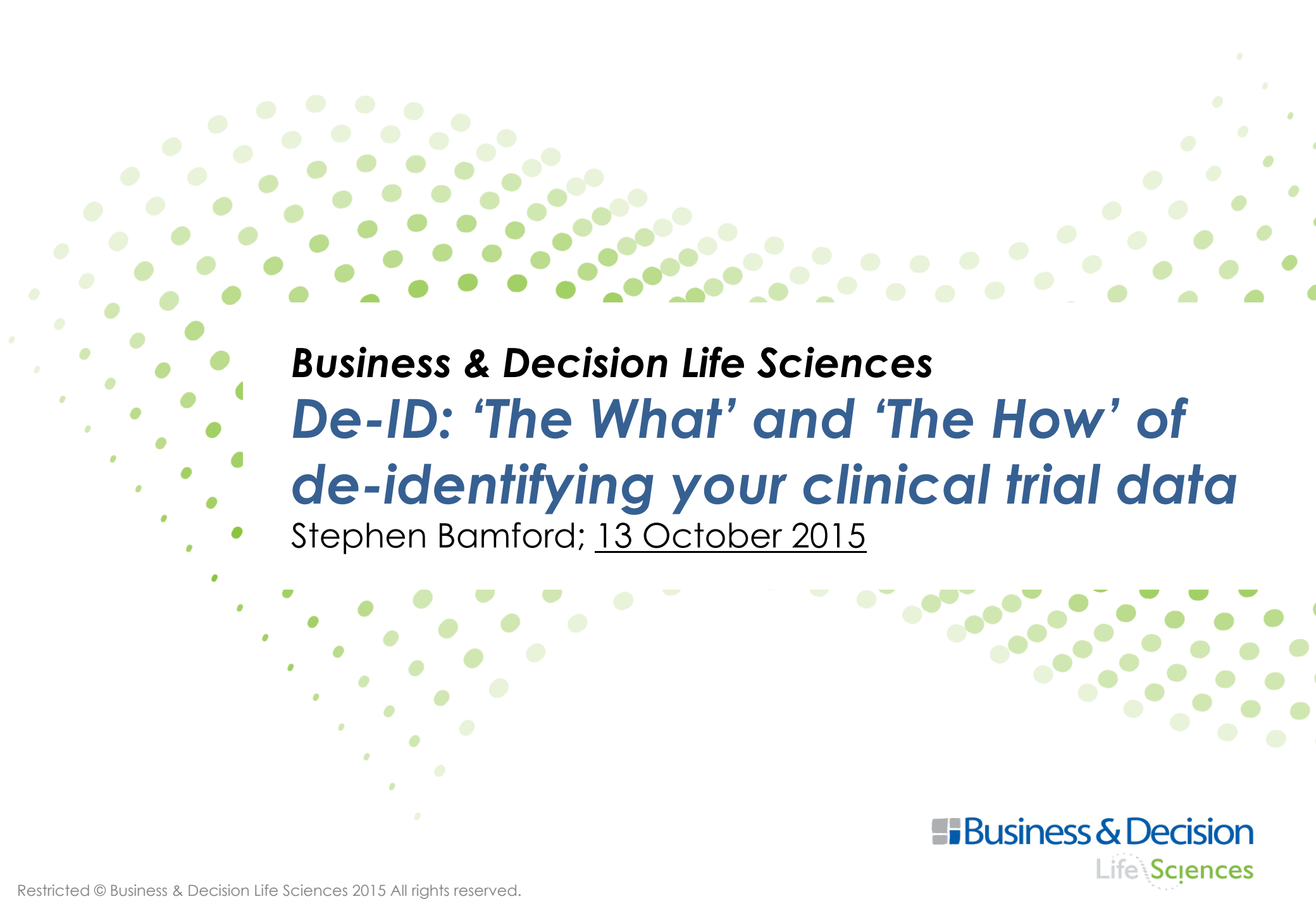


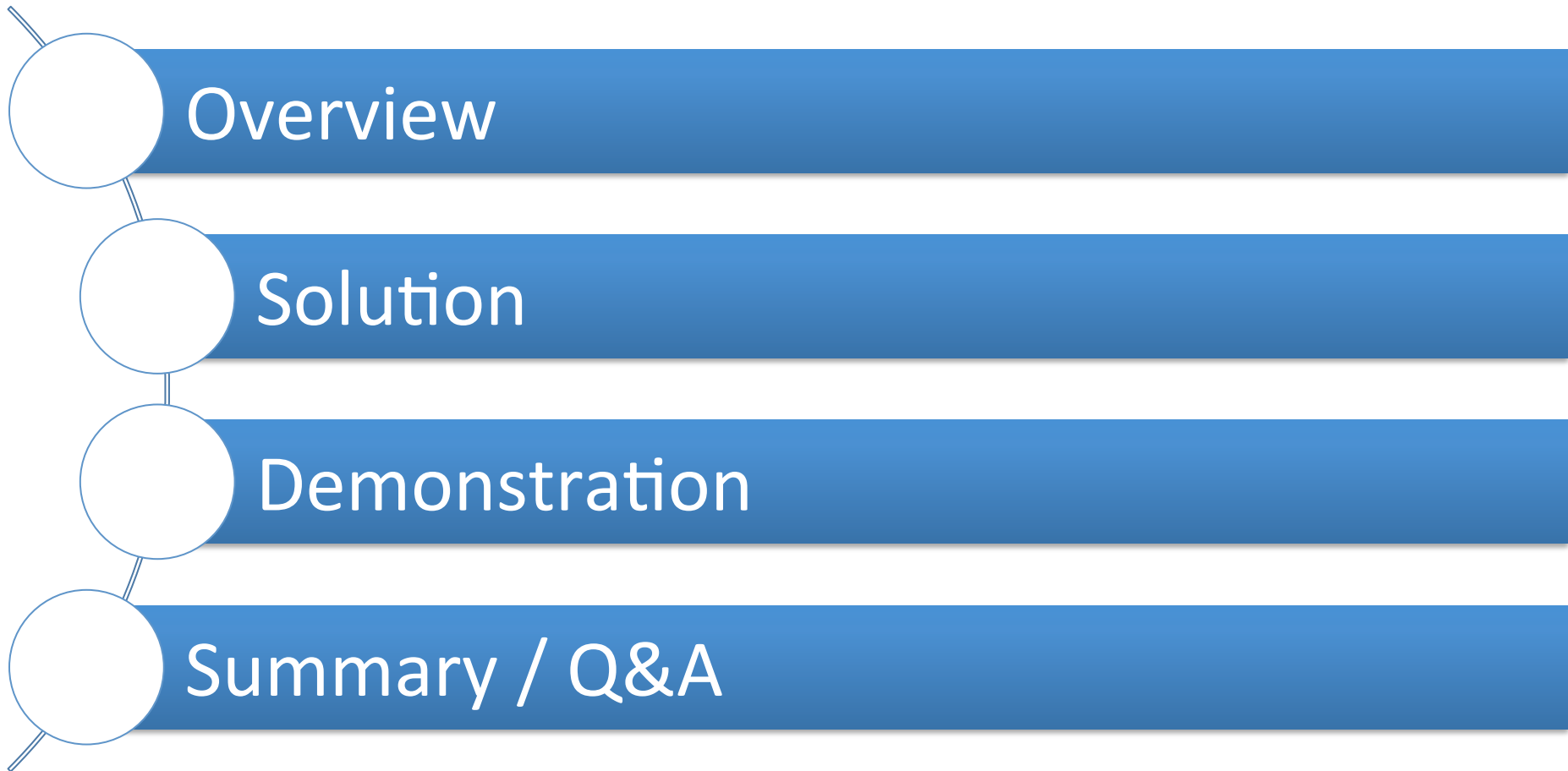


Business & Decision Life Sciences

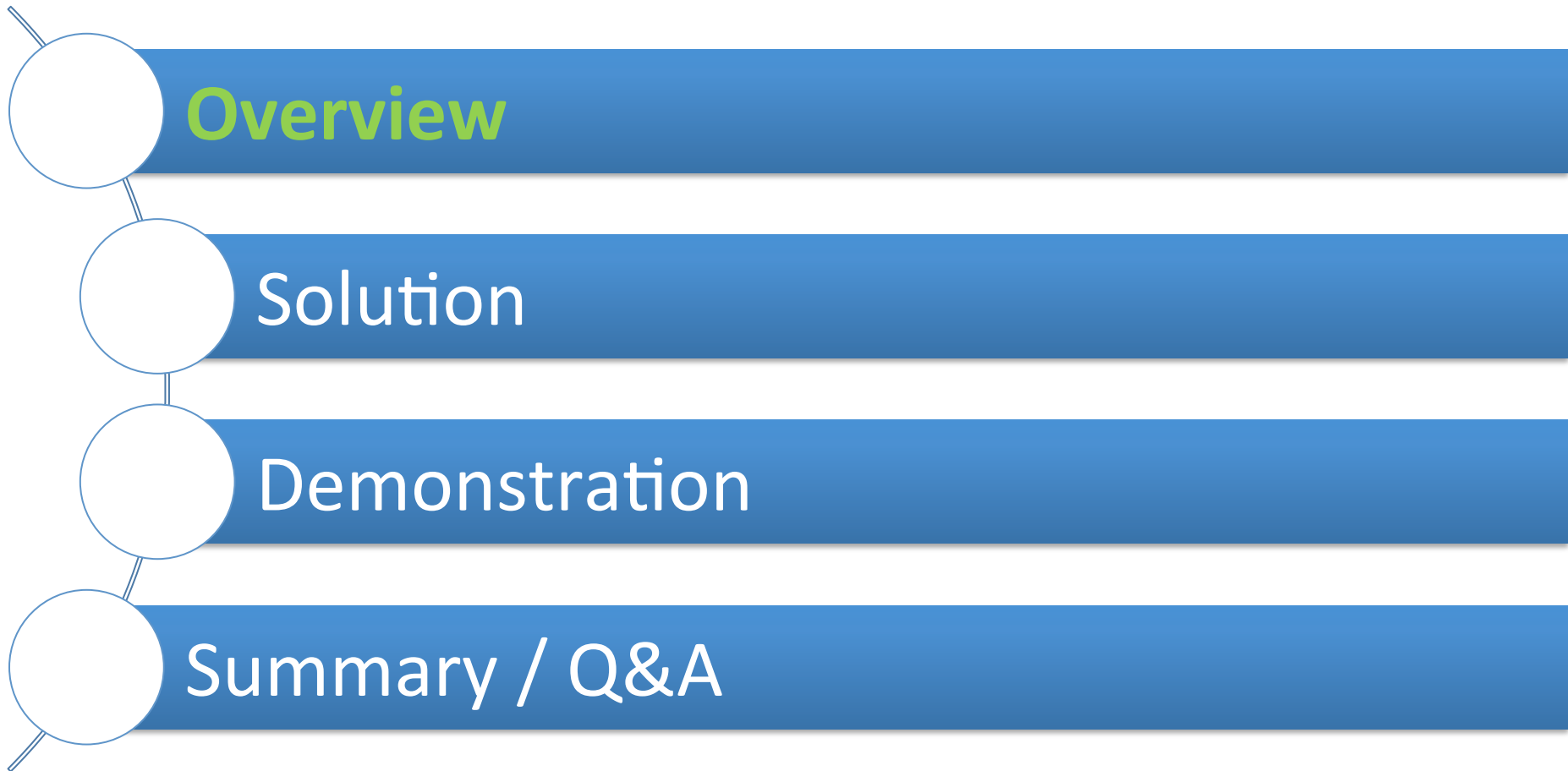
***De-ID: 'The What' and 'The How' of
de-identifying your clinical trial data***

Stephen Bamford; 13 October 2015

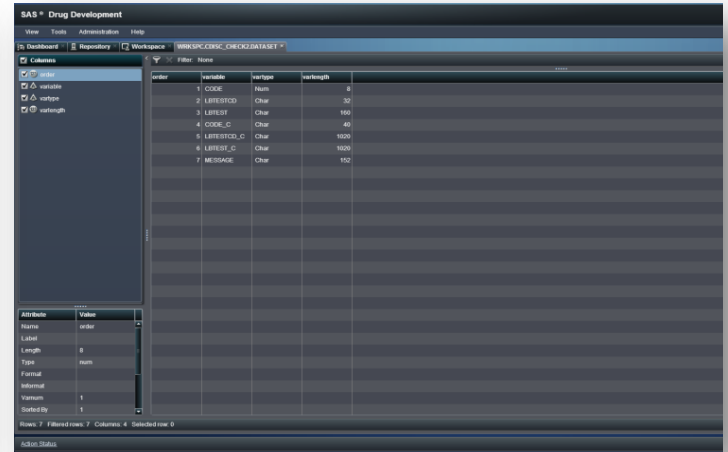
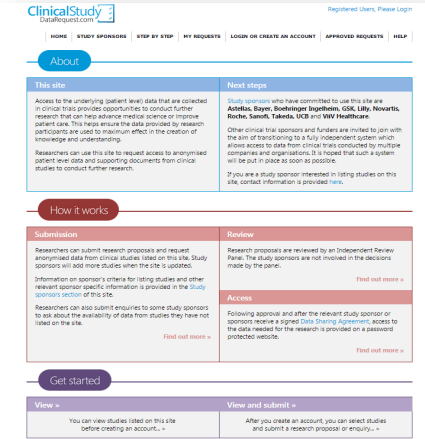
Agenda



Agenda



The industry is tending to follow a standard approach to the de-identification of clinical study data



Public Access Site ➡ Independent Review Panel ➡ Researchers Application

The pharma industry faces challenges when dealing with clinical trial data de-identification

- Providing data access and allowing further research while maintaining the privacy and confidentiality
- Historically de-identification has been a low volume, low-visibility process, lacking urgency
- Growing need for data transparency and sharing initiatives, and regulatory compliance

The pharma industry faces challenges when dealing with clinical trial data de-identification

- Need to address variability in source data
- Reduce effort, time and cost
- Minimise risks to the privacy and confidentiality of research participants
- Ensure compliance with data privacy legal requirements

Variability in Industry De-identification rules

■ Common Rules across sponsors

- Recode Patient ID and Site ID consistently
- Delete or empty all free text (verbatim) fields
- Don't retain ages > 89
- Remove email addresses, URLs, vehicle ID numbers, etc

■ Varying Rules across sponsors

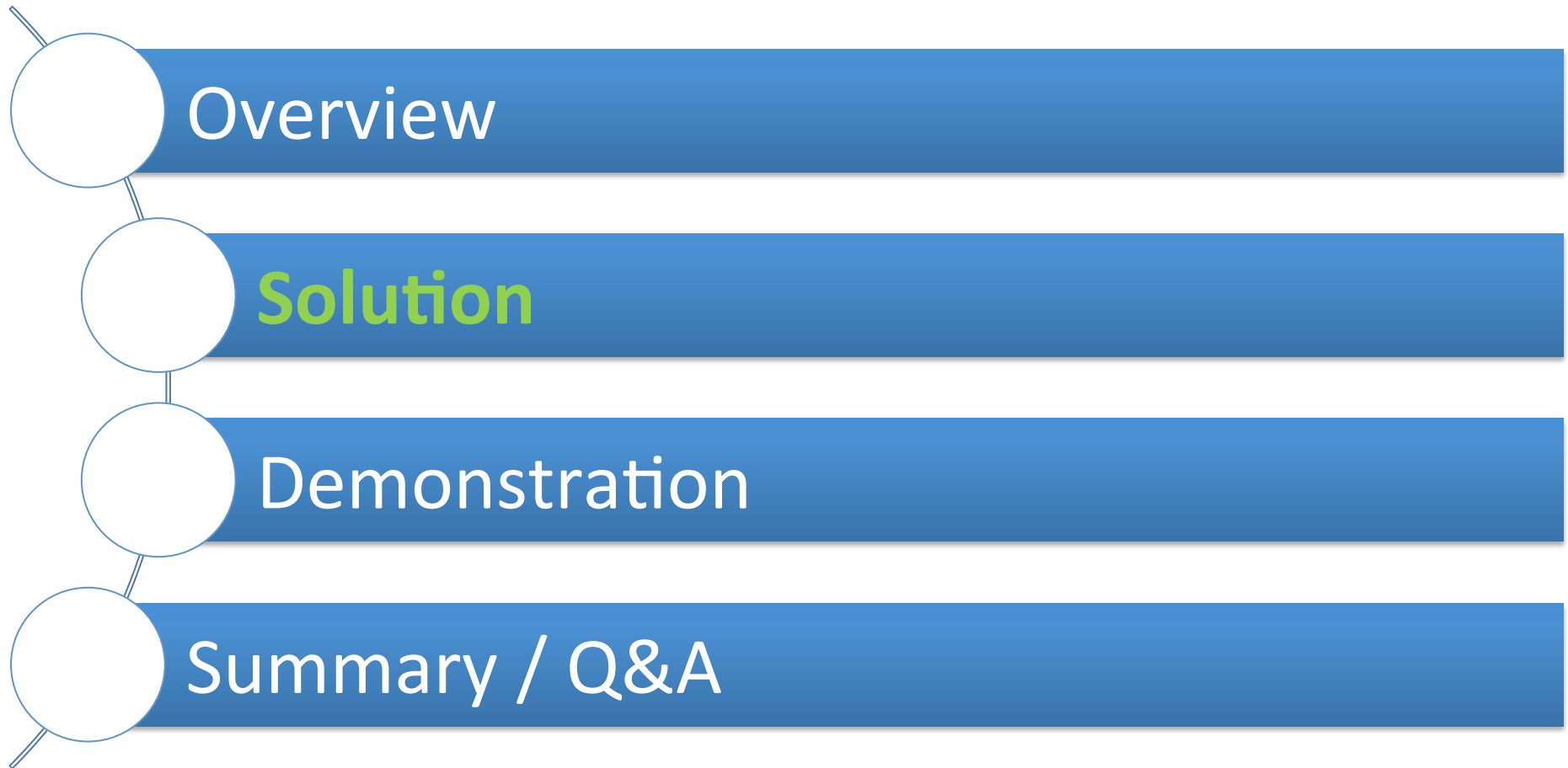
- **Dates**
 - Apply random offsets
 - Replace with calculated Study Day
- **Dates of Birth**
 - Replace with age
 - Delete month and day
- **Re – identification index**
 - (key) Management

The BDLS De-identification solution accommodates variability between trials in terms of data structures and de-identification strategies.

Two stages to the de-identification process

- Removing personally identifiable information (PII) from the dataset.
- Destroying the link (code key) between the dataset that is provided and the original dataset.

Agenda



BDLS's industry-leading solution for De-identification of clinical trial data

- Patient privacy
- Compliance
- Adaptability
- Cost, time and effort reduction

We ensure that patient privacy is maintained without limiting the usability of the data. We understand that the appropriate level of de-identification can vary depending on many factors, such as the planned use of the data, the rarity of the disease under study, and the study design



BDLS Project Logic – Step 1



Understand the data

- Provide expert consultancy services to help Life Sciences organizations determine an appropriate level of de-identification.
- Review the protocol, annotated CRF, clinical study data and intended use for the de-identified data to determine, in liaison with the sponsor, which variables or records contain information that could potentially identify a patient.

BDLS Project Logic – Step 2



■ Identify the appropriate De-identification methods

- Develop a de-identification plan based on the metadata for each individual clinical studies
- Fully document the de-identification functions that will be applied to the applicable variables and records

BDLS Project Logic – Step 3



■ Implement the De-identification methods

- A metadata-driven approach automates the application of the specified de-identified methods
- Automatic generation the de-identification documentation that includes a description of the function used for each variable or record and the de-identification methods and parameters
- Destroy all elements of the seed keys and commendation to ensure complete anonymization of the patent data

BDLS's approach

- All clinical trials are different... They are unique
- De-I.D. works with any clinical data
- Principle of 7 de-identification functions
- More principles can be added in consultation with a client



BDLS has been working with two of the world's top 20 pharma companies and has launched the De-I.D. solution for de-identifying clinical trials.

PhUSE Project Goals



- **Provide peer-reviewed de-identification standards for CDISC data models to the industry**
 - Facilitate the **assessment of direct and quasi identifiers** in CDISC datasets
 - **Ensure consistency** in de-identified data shared across sponsors

De-Identification Standards for CDISC SDTM 3.2



Observation Class	Domain Prefix	Variable Name	Variable Label	Type	Direct/Quasi Identifier	DI Primary Rule	DI Alternative Rule	DI Comment
Special-Purpose	DM	RFPENDTC	Date/Time of End of Participation	Char	Quasi Level 2	Offset		
Special-Purpose	DM	DTHDTC	Date/Time of Death	Char	Quasi Level 1	Offset		In case of Fatal event, this may be considered for further de-identification for low-frequency of dead patients. This is the responsibility of the sponsor to conduct such assessment considering among other occurrence of such death for the concerned subjects in the general population.
Special-Purpose	DM	DTHFL	Subject Death Flag	Char	Quasi Level 2	Keep		In case of Fatal event, this may be considered for further de-identification for low-frequency of dead patients. This is the responsibility of the sponsor to conduct such assessment considering among other occurrence of such death for the concerned subjects in the general population.
Special-Purpose	DM	SITEID	Study Site Identifier	Char	Quasi Level 1	Remove	Recode ID variable	If SITEID is required and is recoded as per the alternative rule, it must be considered within the risk assessment.
Special-Purpose	DM	INVID	Investigator Identifier	Char	Quasi Level 1	Remove	Recode ID variable	If INVID is required and is recoded as per the alternative rule, it must be considered within the risk assessment.
Special-Purpose	DM	INVNAM	Investigator Name	Char	Quasi Level 1	Remove		Such information is related to other individuals than the patients and can also reveal geographic location of site. In addition, it holds little data utility.
Special-Purpose	DM	BRTHDTC	Date/Time of Birth	Char	Quasi Level 1	Remove		
Special-Purpose	DM	AGE	Age	Num	Quasi Level 1	Derive Age	Aggregate Age	
Special-Purpose	DM	AGEU	Age Units	Char				
Special-Purpose	DM	SEX	Sex	Char	Quasi Level 1	Keep		
Special-Purpose	DM	RACE	Race	Char	Quasi Level 1	Keep		If necessary remap to CDISC code lists and consider races with low frequency into a category "OTHERDI".
Special-Purpose	DM	ETHNIC	Ethnicity	Char	Quasi Level 1	Keep		
Special-Purpose	DM	ARMCD	Planned Arm Code	Char				
Special-Purpose	DM	ARM	Description of Planned Arm	Char				
Special-Purpose	DM	ACTARMCD	Actual Arm Code	Char				
Special-Purpose	DM	ACTARM	Description of Actual Arm	Char				
Special-Purpose	DM	COUNTRY	Country	Char	Quasi Level 1	Elevate to continent	Keep	If country is critical to the analysis (e.g. required to reproduce a result), it may be kept and it is the responsibility of the sponsor to assess whether the residual risk is acceptable and take further actions on other variables if necessary. Countries with less than 10 patients must be grouped in country OTHERDI.
Special-Purpose	DM	DMDTC	Date/Time of Collection	Char	Quasi Level 2	Offset		
Special-Purpose	DM	DMDY	Study Day of Collection	Num	Quasi Level 2	No further de-identification		

Dates

Low frequency & rare events

Recoding of unique identifiers

Handling of free-text variables

Extensible code lists

Geographical location

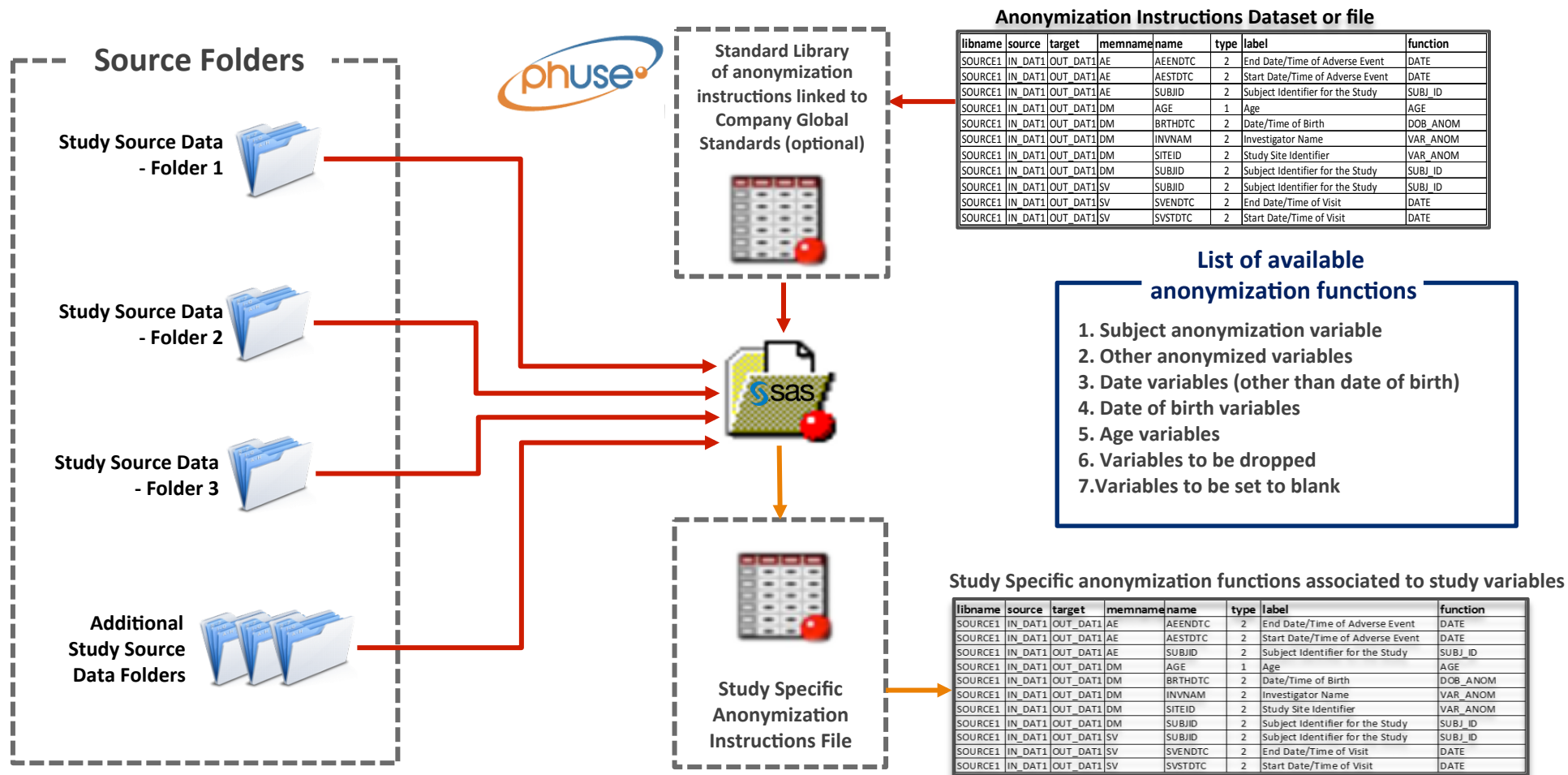
Sensitive data

Quasi identifiers to keep

PII of third-party

+1300 variables

BDLS's Software Approach

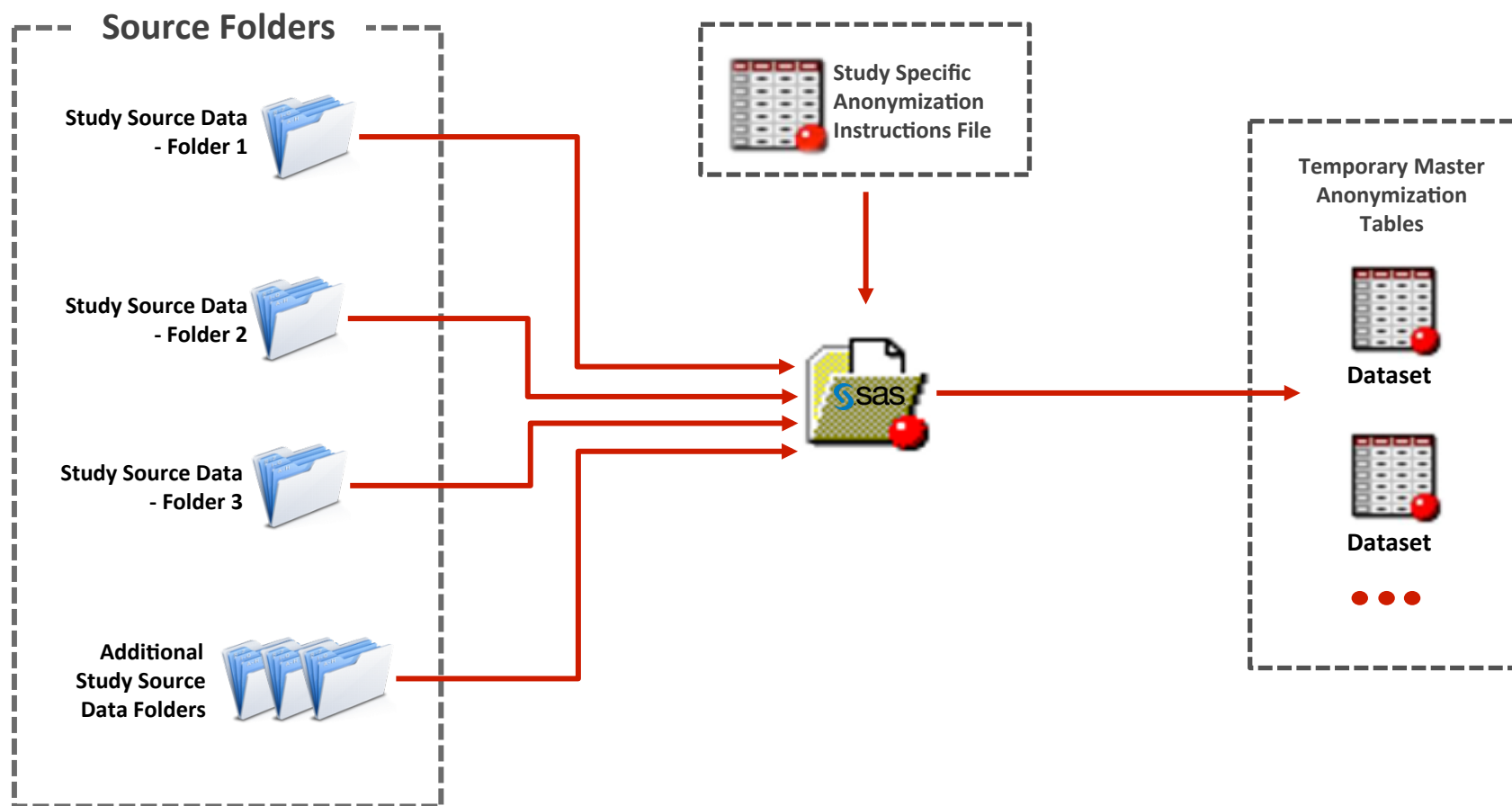


Step 1 – Preparing anonymization instructions

7 Anonymization Functions

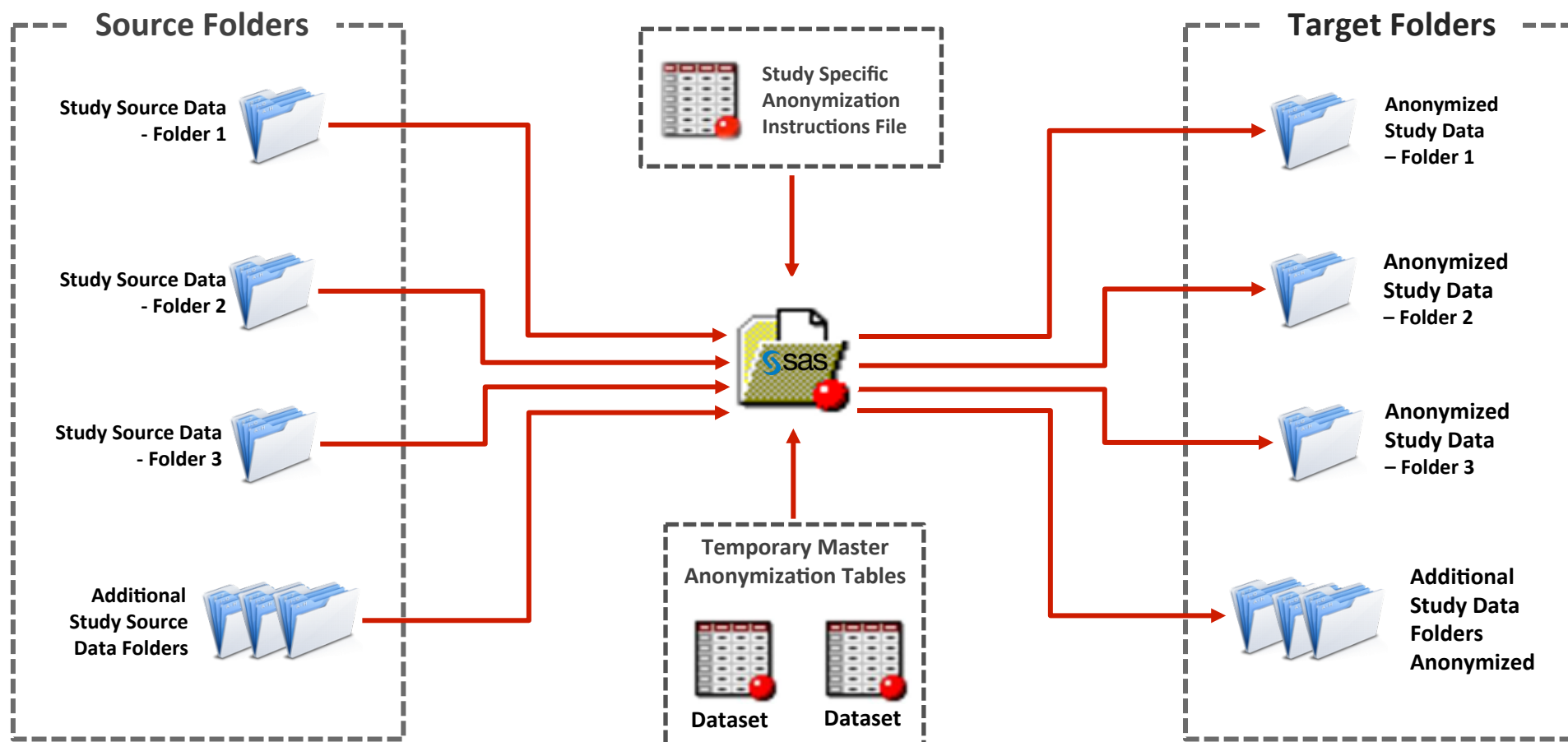
Function	Description
"DROP"	Remove from the dataset
"BLANK"	Content of the variable set to 'blank'
"VAR-REPLACE"	Variable will have named values replaced
"ID"	Subject numbers will be changed
"DATE"	Dates replaced with a newly computed date
"DOB"	Deals with the subjects Date of Birth
"AGE-CAT"	Creation of age categories

BDLS's Software Approach



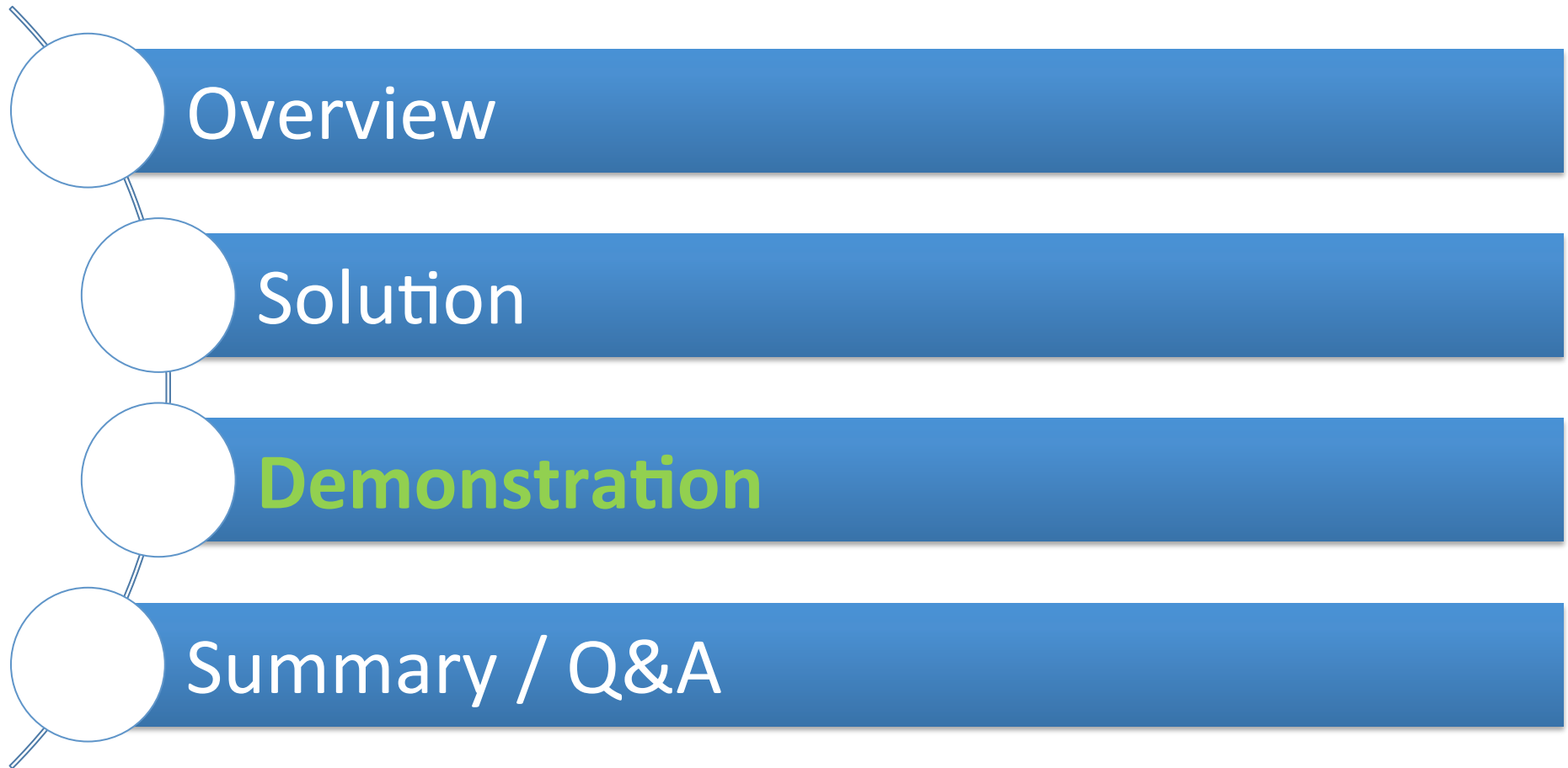
Step 2 – Generating master anonymization tables

BDLS's Software Approach



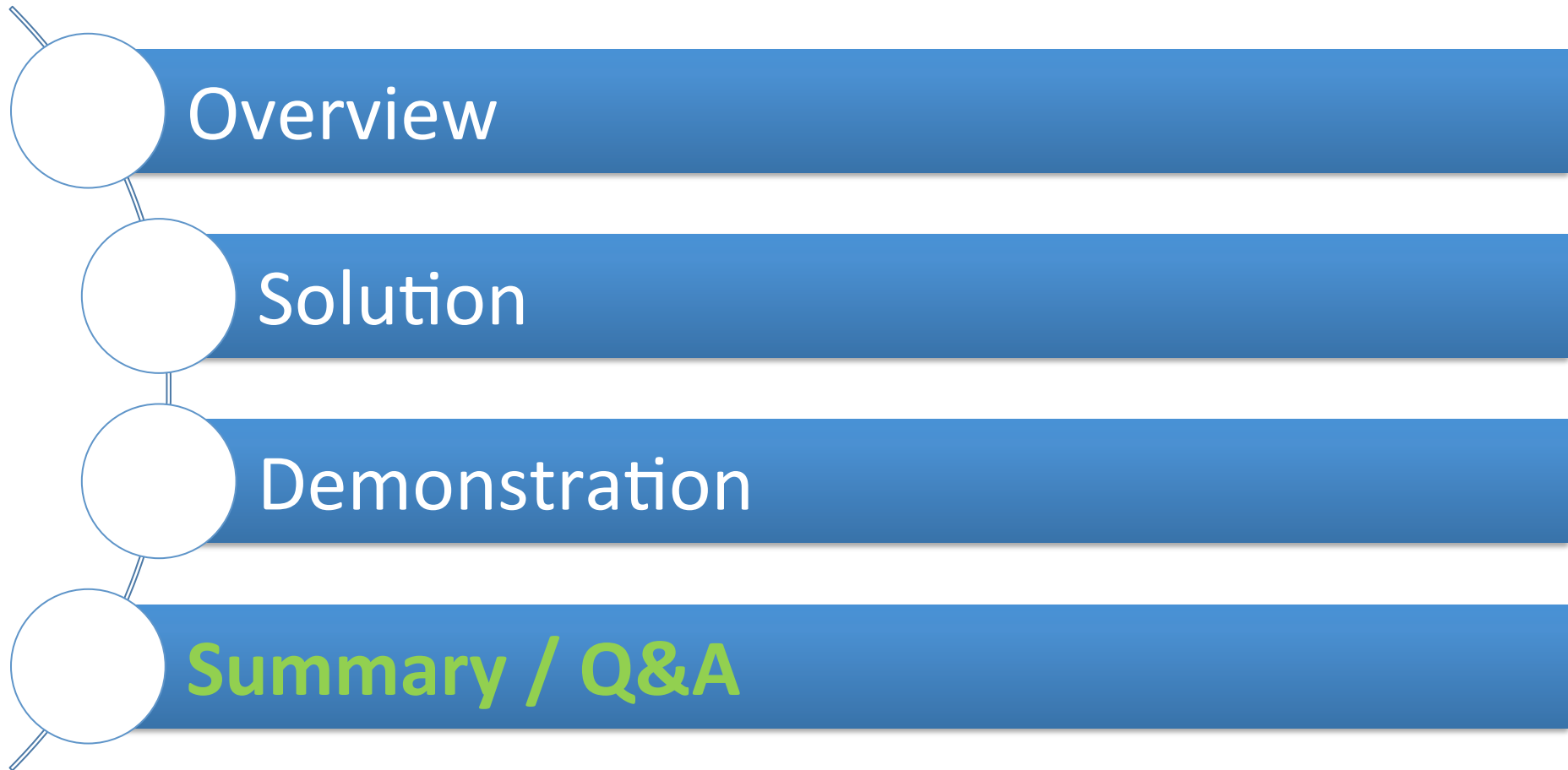
Step 3 – Generation of anonymized final datasets

Agenda



DE-I.D. SOFTWARE DEMONSTRATION

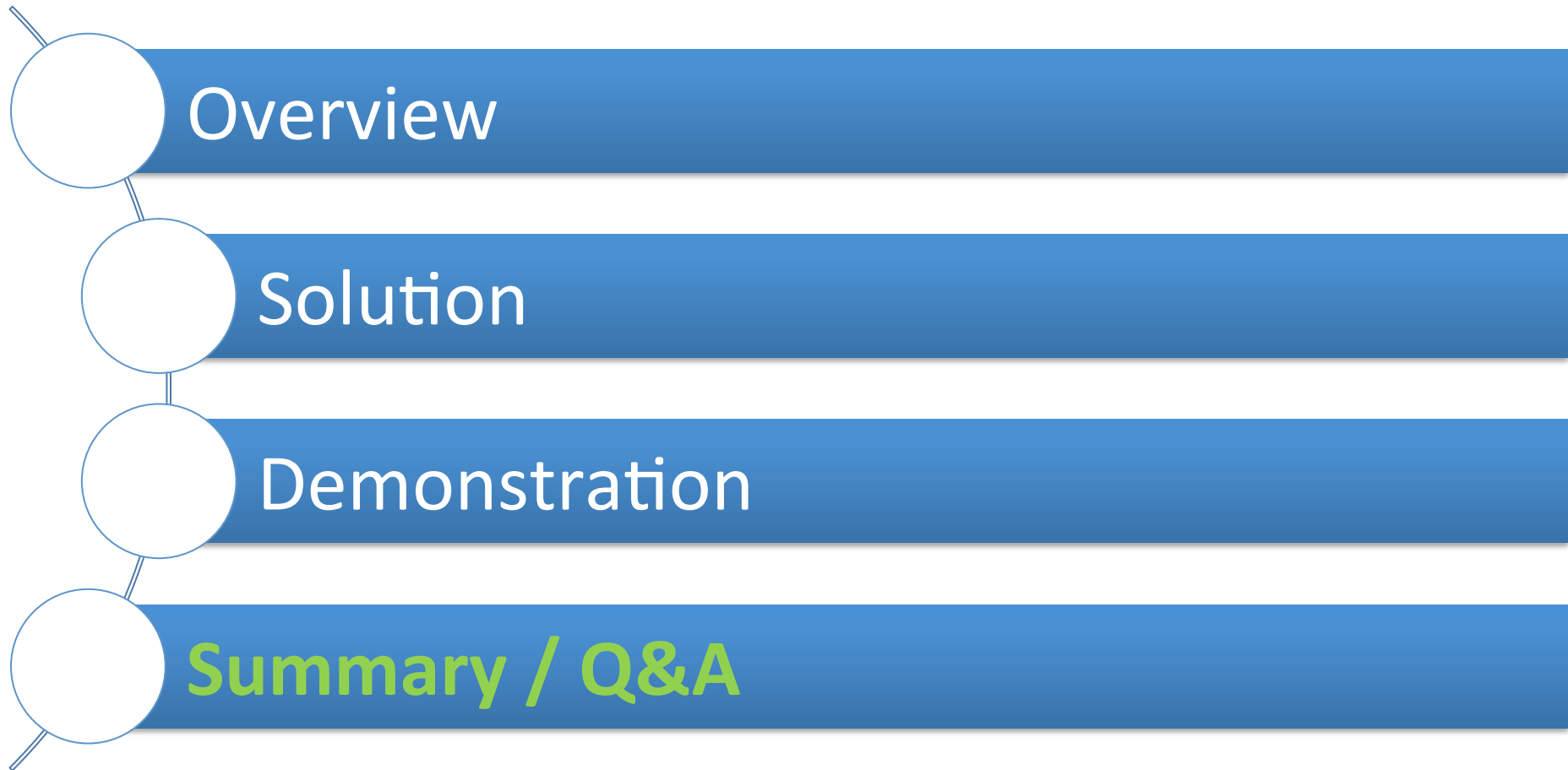
Agenda



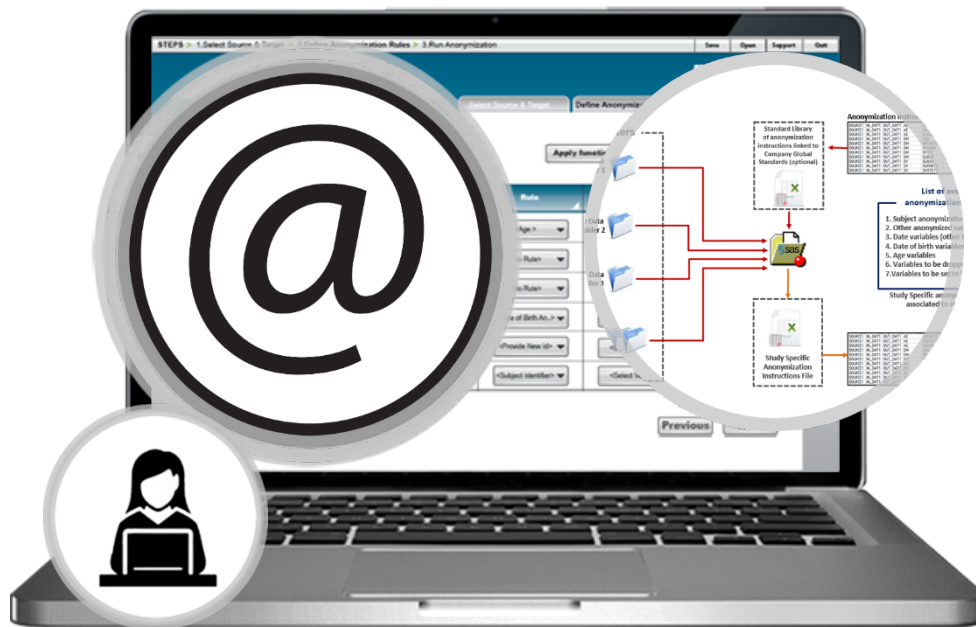
Summary of BDLS's De-I.D. Solution

- De-I.D. has been developed under the BDLS SDLC
- Addresses variability in source data
- Reduces effort, time and cost
- Delivers high-quality de-identified data
- Fully documents de-identification activities

Agenda



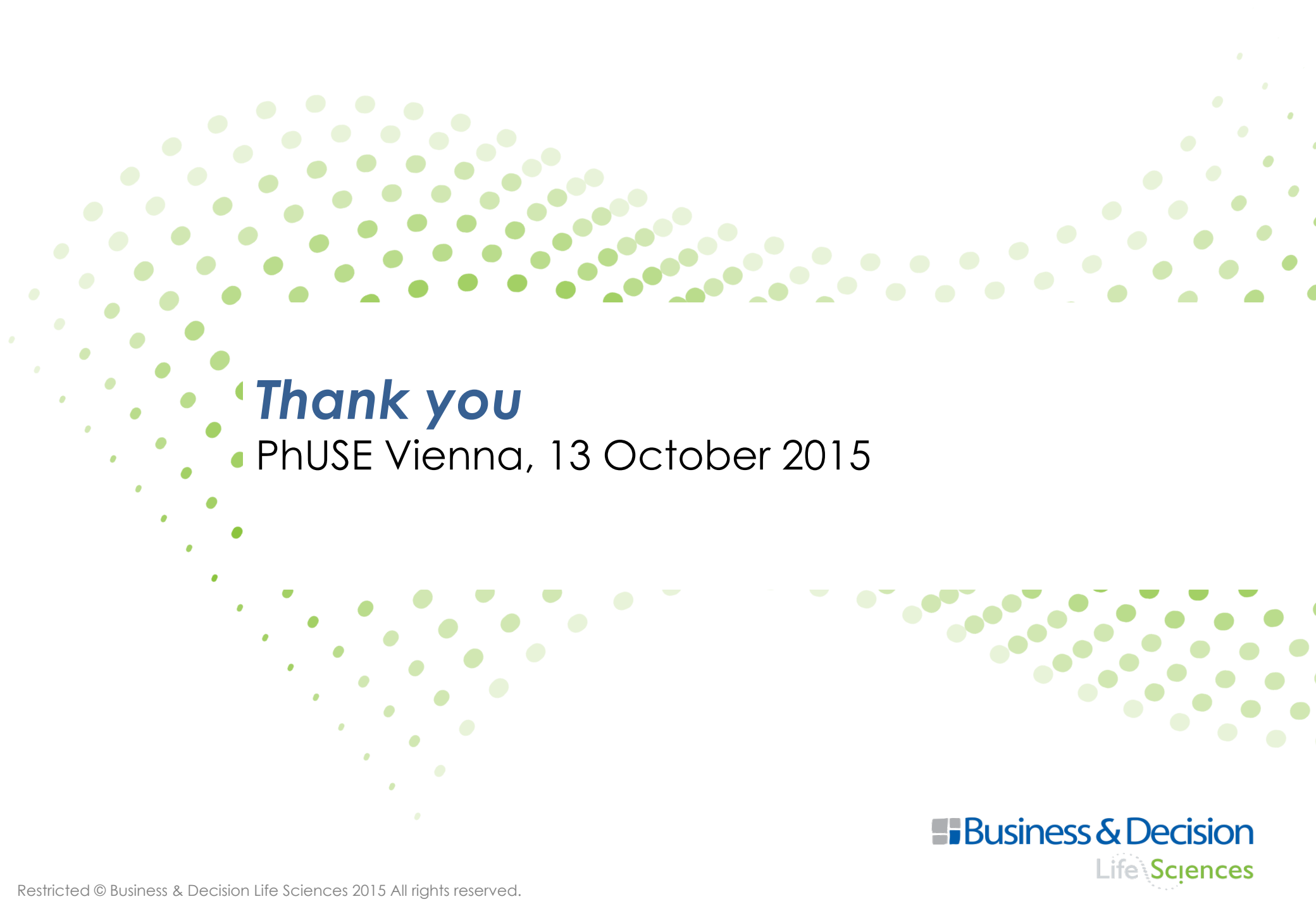
De-I.D. available in 3 different models



1. A private instance of De-I.D. securely hosted by BDLS
2. A in-house instance of De-I.D. with implementation support as required
3. BDLS De-identification as a service

Questions & Answers





Thank you

PhUSE Vienna, 13 October 2015



Stephen Bamford | Director of Biometrics Consultancy | T. +44 75544 18 667 | stephen.bamford@businessdecision.com

Business & Decision Life Sciences

141 rue Saint-Lambert

B-1200 Brussels

T: +32 2 774 11 00

F: +32 2 774 11 99

lifesciences@businessdecision.com

<http://www.businessdecision-lifesciences.com/>