



Introduction

Anonymizing clinical data to ensure patient privacy and confidentiality remains a challenge, especially for **non-standard** or **legacy** structures (not CDISC like) and **SUPPQUAL CDISC** tables.

To aid the anonymization process, Keyrus Biopharma (KBP) has developed an easy-to-use, efficient tool which can anonymize data within any database structure. It can also be used as a final QC solution.

As with any other anonymization solutions the two most important challenges for data managers are:

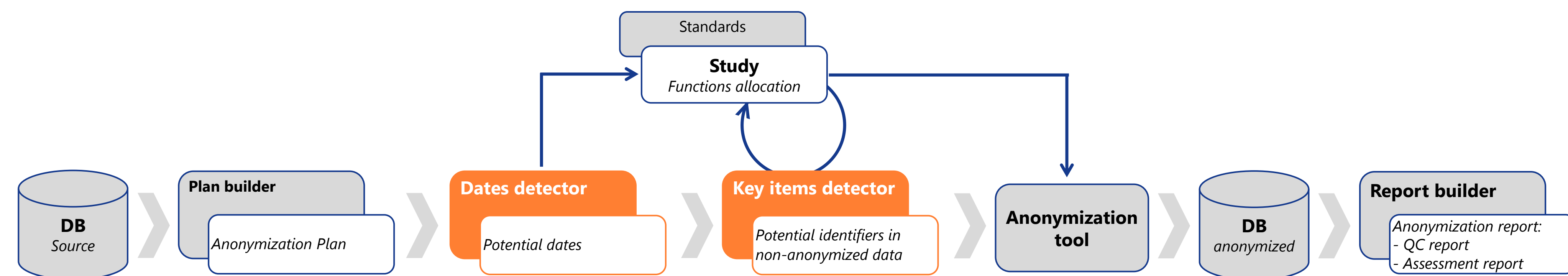
- Detecting all the variables to be anonymized;
- Associating the right anonymization functions.

KBP's tool scan the entire database to automatically detect:

- The variables that can be considered as dates;
- Non-anonymized variables that contain anonymized identifiers as a substring.

This poster presents the two solutions that we have implemented to **limit the risk of missing potential re-identifying information** by detecting the potential dates and potential identifiers.

Data anonymization process & KBP anonymization tool



Database to be anonymized

May contain multiple sources (raw and analysis datasets).

Plan builder

Automatically builds reports using DB anonymization parameters.

Dates detector

Detects all character variables (and count records) which are potentially dates (complete and incomplete) in the database.

Key items detector

Detects all variables (and count records) which contain other anonymized variable values as a substring.

Anonymization tool

Anonymizes database - including normalized datasets (vertical structure) - based on anonymization input parameters, allocated anonymization functions and standard anonymization functions allocation (if existing).

Report builder

Automatically builds report using DB and input anonymization parameters, including automatic QC and risk assessment reports.

Conclusion

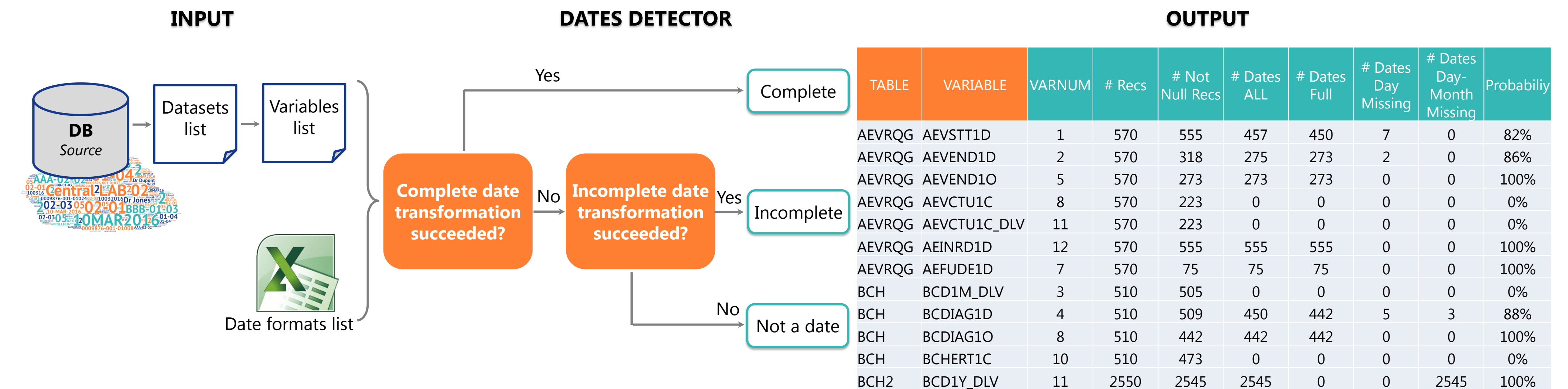
Reducing the risk of missing potential identifiable information in data which will be made available for external users - to maintain transparency and provide exploitable data - is a **difficult** and **time-consuming** obligation to comply with.

The solutions we have developed can help at the beginning of the anonymization process to identify:

- Which variables are at risk of being missed as "date" which should be anonymized
- Which variables are at risk of containing a value coming from another anonymized identifier

It can also be used as a QC tool at the end of the anonymization process.

Dates detector



Database to be anonymized

May contain multiple sources (raw and analysis datasets).

Date formats list

SAS formats and informats

Dates detector

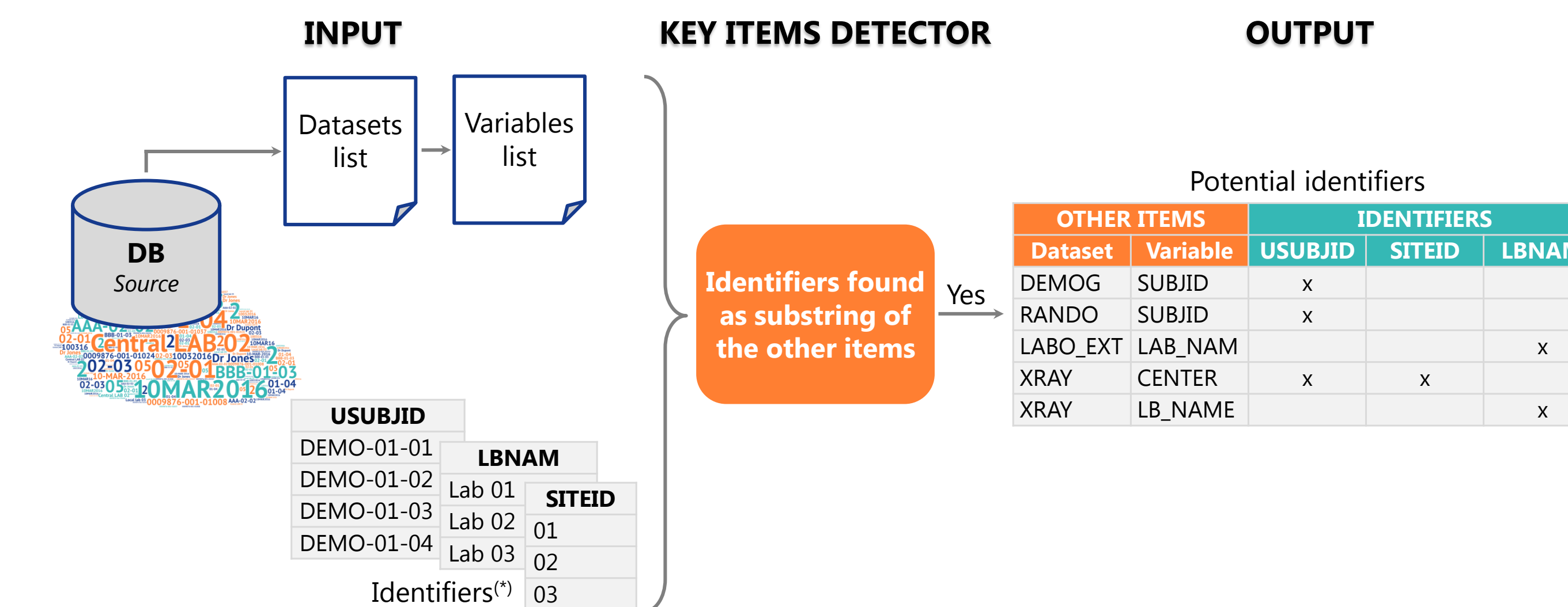
The SAS program scans all tables & variables of the DB. Then, checks if the value is a complete or incomplete "date" (missing day or missing day-month).

A report by table-variable is generated with:

- Numbers of records that are detected as a "date" (complete or incomplete);
- Numbers of records by formats.

Missing day-month symbols (UNK, NK ...) can be used as an input parameter. Additional detailed reports (on records position in the dataset when detected as date) are available, and mixed date formats within variable are detected as well.

Key items detector



The key items detector tool scans the records in the database for variables (character or numeric) which are not anonymized.

If any record looks like a key identifier value (contains as a substring), then the corresponding variable is highlighted.

Those records are at risk for re-identification and the variable should probably be anonymized with a new identifier as well.

This tool will **reduce the risk of missing variables requiring anonymization**.

(*) Variables that need to be anonymized with new identifiers - such as key identifiers - are defined during the functions allocation step.

References

Code of Federal Regulations - Title 45; Public Welfare, Subtitle A \$164.514: Available at http://www.ecfr.gov/cgi-bin/text-idx?tpl=/ecfrbrowse/Title45/45cfr164_main_02.tpl
Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule – http://privacyruleandresearch.nih.gov/pr_08.asp#8a
Council Regulation (EC) 45/2001 of the European Parliament and of the Council of 18 Dec 2000 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data [2008] OJ L193/7.
Council Directive (EC) 95/46 of the European parliament and of the council of 24 Oct 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L281.
Regulation (EU) 536/2014 of the European parliament and of the council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC [2014] OJ L158/1.
European Medicines Agency: EMA/240810/2013 - Publication of clinical data for medicinal products for human use. http://www.ema.europa.eu/docs/en_GB/document_library/Other/2014/10/WC500174796.pdf
Article 29 Data Protection Working Party: 0829/14/EN WP216 - Opinion 05/2014 on Anonymisation Techniques. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf
European Medicines Agency policy on publication of clinical data for medicinal products for human use: EMA/240810/2013– 02Oct2014 http://www.ema.europa.eu/docs/en_GB/document_library/Other/2014/10/WC500174796.pdf
Data De-identification and Anonymization of Individual Patient Data in Clinical Studies – A Model Approach - 2013 Transcelerate BioPharma Inc. All rights reserved – <http://www.transceleratebiopharmainc.com/wp-content/uploads/2015/04/CDT-Data-Anonymization-Paper-FINAL.pdf>

