GDPR is Coming

How Can You Identify Where Personal Data Resides in Your Oracle ERP and CRM Systems?

The EU's new rules on data protection (GDPR) come into force in May 2018. For those who may be unaware of them, they strengthen the rights of citizens over their own data and put more onus on the responsibilities of organisations to manage and protect that Personal Data.

Roland Bullivant,
Sales & Marketing Director,
Silwood Technology

The UK Government has also signalled that it will introduce fundamentally the same rules to apply before and after Brexit.

The UK Information Commissioner's Office provides a wealth of useful information including a 12-step guide to becoming compliant (https://ico.org.uk/media/1624219/preparing-for-the-gdpr-12-steps.pdf). One of these steps (number 2) says that you "should document what Personal Data you hold, where it came from and who you share it with. You may need to organise an information audit across the organisation or within particular business areas."

After all, if an organisation discovers that it has inaccurate Personal Data which it has shared with another organisation, it will need to inform the other organisation about the inaccuracy so it can amend its own records. In addition, if customers want to know what data is held about them, wish to have data erased or corrected or decide to withdraw consent to data being processed, it is necessary to respond to those requests quickly and effectively.

This will not be possible unless the organisation knows what Personal Data is being held and where it is. This article discusses the possible methods organisations might employ to find out where personal data exists in Oracle packages such as Oracle E-Business Suite, Siebel, PeopleSoft and JD Edwards. It offers an alternative software-driven approach using Safyr® from Silwood Technology Limited and uses an example based on JD Edwards as an illustration.

Needles in haystacks? Where is the Personal Data in your Oracle ERP and CRM packages?

As outlined above, one of the challenges organisations are likely to face in the race to GDPR compliance is knowing where personal data is stored across their information systems landscape. The basis for much of this information will be the metadata in source systems across the enterprise information landscape.

For many organisations the primary mechanism used to store this metadata information will be an Enterprise Data Catalogue, Data Dictionary or Data Governance platform. This will provide the basis for knowing where changes need to be made to enable compliance with the new Consent and Rights of Data Subjects rules.

Populating these with the right metadata from many systems is quite straightforward, as the metadata, data model or source data is easy to locate and understand. Additionally, there are many software products which can scan those systems and deliver that information relatively easily.

However, if an organisation is running enterprise Customer Resource Management (CRM) or Enterprise Resource Planning (ERP) applications from Oracle, Salesforce, SAP, Microsoft or others then finding metadata which relates to Personal Data will be more of a challenge, especially if its location is not already known.

This is because of the size, complexity and level of customisation of the data models (metadata) which underpin these systems. Also, in most cases, the metadata is very opaque because the database System Catalogue provides nothing useful in the form of business names for tables and fields and no information about table relationships.

So how does one go about finding Personal Data information quickly and accurately?

As an example, consider the methods one might employ to locate all the tables that store a particular Personal Data attribute in a JD Edwards system. These will be pretty much the same for any Oracle, SAP or other package.

In this case 'Date of Birth' is used as the piece of Personal Data to be located.

A typical JD Edwards system has well over 4,000 tables and is often embellished with a significant number of customisations.

Using documentation

The first question to ask when referring to documentation relating to metadata is: "Does it exist?" and, if it does exist, can it be accessed easily and does it reflect any changes that have been made to the data model during implementation?

If documentation is present, it can provide a good starting point for metadata discovery and JD Edwards is helpful to some degree because it delivers around 40 data models in document form. Most of the other packages are less friendly in this regard.

The challenge, however, is to identify and locate the Personal Data items in these models and also those which are not included in these documents. In addition, delivering data models in this static way means that they would have to have been updated to remain in line with any customisations made to the overall data model and, of course, any useful information cannot be shared easily with other tools.

Finally, recording that information for the Data Catalogue will require the information to be rekeyed or copied into that system.

Asking internal specialists

Internal technical specialists will have access to JD Edwards' tools for exploring the definitions of tables and field attributes within the system as implemented. It is likely that they will also have good knowledge of the system and the particular way in which it has been customised to meet specific requirements. They should then be able to go through a process to locate the Personal Data attributes for each table and record that information, perhaps in a spreadsheet or directly in the Data Catalogue tool.

One challenge with this approach could be the amount of time it takes to achieve this task and whether all Personal Data attributes have been identified.

Other packages also have their own tools for looking at the data model; however, as with JD Edwards, they are designed for use by technical, development and administration staff rather than by data analysts.

Engaging software vendor staff or consultants

Some organisations may require the services of external consultants to achieve the same results as internal specialists. These may be specialists in the relevant package, in this case JD Edwards, or GDPR specialists, from a systems integrator or possibly from the Data Catalogue or Data Dictionary vendor.

Depending on their skillset they may be able use whatever application tools are available to them or provide some standard templates which can be used as a base for further investigation.

It may also be possible to automate some of the metadata ingestion processes – once the metadata discovery has been completed. In the example under consideration, it is necessary to identify the location of a specific Personal Data attribute; therefore, any software tools or templates need to be able to provide accurate pointers to this information in order to be of value.

In any event there will be work required to become familiar with, and cater for, the specific customisations which have been made to the original package data model in this context.

Internet search

When all else fails, here at Silwood we have encountered organisations whose staff search the internet for data models that represent specific business concepts in an ERP or CRM package.

This can be helpful but is fraught with problems, as the model, if found, may not be based on the same version of the software and is unlikely to have the same customisations as implemented by the customer.

Using dedicated metadata discovery software

As discussed earlier, many software vendors will provide facilities to import metadata that is easy to find and understand. All large

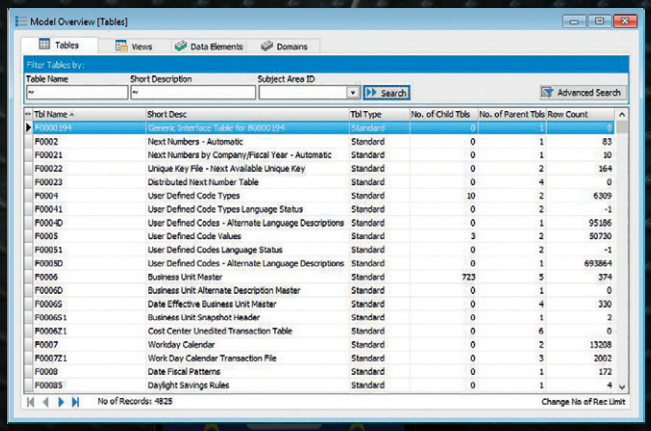

FIGURE 1: PARTIAL VIEW OF A COMPLETE LIST OF JD EDWARDS TABLES

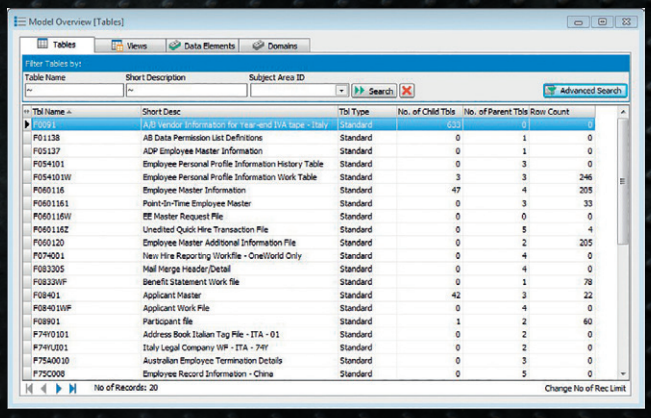

FIGURE 2: LIST OF JD EDWARDS TABLES WHICH HAVE THE 'DATE OF BIRTH' FIELD

complex packages present a more difficult challenge when it comes to metadata discovery.

In this example we will look at how it is possible to find all instances of an attribute which contains the string 'date of birth' across all tables in a JD Edwards system

Safyr from Silwood Technology offers an alternative approach and provides data analysts with unique insight into, and utilisation of the JD Edwards data model. It extracts the metadata from the application as implemented and stores it in a repository.

This means that customisations to the data model are included so that users can be confident that they are working with accurate metadata. Importantly, Safyr also surfaces business descriptions as well as physical information about tables and attributes and discovers the relationships between tables.

Typically the first view from Safyr of the JD Edwards metadata would be to look at a complete list of tables (see Figure 1). In this instance of JD Edwards there are a total of 4,825 tables. The physical name for each table is accompanied by its Short Description as defined in the system. Other information presented here includes the Table Type, Number of Child and Parent relationships and a Row count.

There are many ways to use Safyr to search and analyse metadata; however, in this example it is necessary to find specific fields which include the Personal Data attribute 'date of birth'.

Therefore the Advanced Search feature has been used to look across all tables in the JD Edwards repository in Safyr and locate each instance where the string 'date of birth' appears in a field. In the system used here this delivers a total of 20 tables which have fields that meet that criteria (see Figure 2).

There are several options for sharing these results with other tools and technologies, such as metadata management or data governance products, data catalogues or dictionaries and data quality solutions in order to support the data readiness or information audit phase of a GDPR initiative. In this example it has been exported to automatically create an Excel spreadsheet which contains the table and field names which have come from the original search (see Figure 3).

Safyr Tables

Table Name	Short Desc.	Type	Phys. Name	No. of Parent Tables	No. of Child Tables	Row Count
F054101	Employee Personal Profile Information History Table	Standard	JDEDATA910.F054101	0	0	0
F054101W	Employee Personal Profile Information Work Table	Standard	JDEDATA910.F054101W	3	0	246
F060116	Employee Master Information	Standard	JDEDATA910.F060116	47	0	205
F0601161	Point-In-Time Employee Master	Standard	JDEDATA910.F0601161	0	0	33
F060116W	EE Master Request File	Standard	JDEDATA910.F060116W	0	0	0
F060116Z	Unedited Quick Hire Transaction File	Standard	JDEDATA910.F060116Z	0	0	4
F060120	Employee Master Additional Information File	Standard	JDEDATA910.F060120	0	0	205
F074001	New Hire Reporting Workfile - OneWorld Only	Standard	JDEDATA910.F074001	0	0	0
F083305	Mail Merge Header/Detail	Standard	JDEDATA910.F083305	0	0	0
F0833WF	Benefit Statement Work file	Standard	JDEDATA910.F0833WF	0	0	78
F08401	Applicant Master	Standard	JDEDATA910.F08401	42	0	22
F08401WF	Applicant Work File	Standard	JDEDATA910.F08401WF	0	0	0
F08901	Participant file	Standard	JDEDATA910.F08901	1	0	60
F74Y0101	Address Book Italian Tag File - ITA - 01	Standard	JDEDATA910.F74Y0101	0	0	0
F74YUI01	Italy Legal Company WF - ITA - 74Y	Standard	JDEDATA910.F74YUI01	0	0	0
F75A0010	Australian Employee Termination Details	Standard	JDEDATA910.F75A0010	0	0	0
F75C008	Employee Record Information - China	Standard	JDEDATA910.F75C008	0	0	0

Metadata Export | Tables | Table Fields | Table Relationships | Table Relationship Members

FIGURE 3: USING SAFYR METADATA REPORTING TO AUTOMATICALLY CREATE AN EXCEL SPREADSHEET WITH DATE OF BIRTH TABLES AND FIELDS FROM JD EDWARDS.

In contrast to manual or more resource-intensive methods this whole process can be accomplished in a few minutes. This means that Safyr can be used to accelerate understanding and exploitation of the JD Edwards data model and therefore the delivery of a wide range of data management projects, including those designed to achieve GDPR compliance. It does this by providing the user with a range of facilities to navigate and create subsets of the data model, as implemented, quickly and easily and in an accessible format. As a result it delivers an effective data-driven software alternative to traditional methods of metadata discovery.

Safyr provides the same capabilities for other Oracle packages along with SAP E-Business Suite, Salesforce and Microsoft Dynamics. ■

ABOUT THE AUTHOR

Roland Bullivant
Sales & Marketing Director, Silwood Technology

Roland Bullivant is a veteran of the software industry and for most of his career he has worked with enterprise customers to increase the value they realise from their data. He has specialised in Data and Information Management, Data Warehouse, Business Intelligence and Data Governance. He has worked for organisations such as SAP/ Business Objects, SeeBeyond and Siebel (both now Oracle) and joined Silwood Technology in 2010.