GDPR Retention Policy Work

Data Dictionary Enhancement

The GDPR advance on DPA with respect to data-retention policies is more on the complete inclusion of all structured data. A design approach using whole-data Entity-Life-Histories is preferred over ad-hoc policy outlines for small parts of the data-set. (GDPR: Data protection by design and by default)

All database, table and field additions and modifications must be controlled; this applies for both Anonymise and Retention policies. This means that entries in the 'datadict' tables must be maintained and a separate schema control management system must be used. This schema control management system (SCMS) will need to define the retention date/time of the database/table as well as the person responsible for its use and deletion. The current automatic restrictions of database schema changes for Snowdon database is to be extended to all production databases; with required entries in the Data-Dictionary.

The 'Data Dictionary' (DD) requires expanding to be inclusive of our entire dataset. This comprises two expansions, the first being the content of the DD must reference every database/schema/table/column for the production Databases. Secondly the DD structure needs to handle more complex control situations for the 'Schema Control Management System' (SCMS).

It is a pre-requisite to have an entry in this sub-system before alterations are made to any database schemas.

Every dataset needs to have an entry in the 'Model' table; this includes 'Access' databases as well as the database 'SnapShots'. Although 'Excel' spreadsheets may contain structured data, they are only required to be noted in the SCMS. Some Word documents may also be included in this table, if they contain personal data in a structured form; for example mail-merged documents, but only a reference to the containing directory and 'template' document need be stated here. Each entry is required to state the GDPR-Retention policy implementation process for the dataset.

An entry for anonymisation rules in the data-dictionary implies a corresponding automatic retention action.

Handle Retention delete failures in 'Retain_Queue' table better and include dormant accounts in investigation.

GDPR 'Privacy by Design' governs SCMS.

Support for multiple retention policies, For example; working data, back-up data, 'Trace' data and 'Audit' data.

Support for different retention policies for different clients.

Schema Control Management System (SCMS)

A proposal. To ensure we include every database/table/column in the GDPR, we need to add to the data-dictionary (DD) to enable this to happen. At the highest level, this will only be a 'Statement' of inclusion to allow for the support of transitory datasets; this will also allow for short-term temporary data-sets but with more complete DD entries. The 3 levels of control envisioned are Complete (to column detail), Partial (to table detail) and Outline (environment, scope and database detail only). For temporally limited datasets, a retention period needs to be assigned as well as a controlling corporate entity (person or department). Production data will need to be defined to the 'Complete' level of detail, whilst control and operational data need only be to the 'Partial' level of detail. However, every dataset must as a minimum have 'Outline' level of detail.

New table to support 'Outline' level of detail, named 'DataDictionaryModel' or 'Model', as part of the [datadict] schema.

Points not followed in SharePoint 'Data Retention' document

2.2.2 Data Categorisation Workshops - Defining the last Debtor Contact Date

This is determined from the 'last Trans_Date Transactions date' and 'last Notes entry date', contrary to the document.

2.2.3 Retention Framework Data Model - Retention Schedule

The 'Retention Action – remove row' is the only process to be used.

We are not following the design document as far as field granularity is concerned; instead we use 'Our-Ref' as the only filter.

To meet requirement 2.2.3 (Retention Action) in the design document, we need to address the thinning of indirect personal data. This can be achieved by replacing the 'OurRef' fields with zero or null values when the records cannot be deleted. The apparent requirement to retain the records of the 'Account' table in 'Snowdon' database when the customer data is deleted is not supported in the current system.

Not covered by GDPR: the following cases are not covered by the regulation:

- Statistical and scientific analysis
- Deceased persons are subject to national legislation
- There is a dedicated law on employer-employee relationships and data handling (HR).

Data Entry and Cleaning

This list may be modified after investigations are finished into Database removal per Environment.

1. Copy Data-Dict changes to live environment, initialise 'Scope' data, and add current databases to new 'Model' table; by scripts or manually.
   1.1. Add entries to DD 'entities' table.
   1.2. Add entries to DD 'attributes' table.
   1.3. Add new table DD 'tabDataDictionaryModel'.
   1.4. Add new table DD 'tabDataDictionaryModelScope'.
   1.5. Add new columns to DD 'entities' table.
   1.6. Add new columns to DD 'retain_queue' table.
   1.7. Add current DD Databases to DD-Model table.
   1.8. Add new relationship links to DD tables.
2. Add remaining identified databases to SCSM with personal data.
   2.1. Call-Credit: add 8 'ourref's
   2.2. Customer-Contact: add 4 'ourref's
   2.3. Data-Services: add 1 'ourref'
   2.4. Scanner: add 3 'ourref's
   2.5. Snowdon-Migrated: ignore
   2.6. Snowdon-Temp: ignore
3. Add missing tables from production Databases to SCSM, fully controlled; volume of columns to add 37K columns, by scripts run on each DB.
4. Add missing other database's and tables to SCSM, columns not required; volume of tables to add 3200 tables, by scripts run on each DB.
5. Add all other databases to SCSM, tables and columns not required; by manual data-entry, approx. 20 databases.
6. Add other structured datasets to SCSM (Access files, Excel files and Word-Merge Templates and directories). Use dedicated column to contain the Filespec, and fill in the Implementation paradigm used to follow the relevant Anonymisation and Retention Policies; by manual data-entry.
7. Remove 'orphan' records from databases keyed on Erased or Deleted 'OurRef' accounts, by script run on each DB.
8. Unstructured datasets are out-of-scope for this work, if they cannot be included in point-6 above. They should use SharePoint pages to explain how they are controlled for GDPR-Policy purposes.

Extra Tables

These tables will handle every database in the dataset, even if they are not under our control (e.g. 3rd party software), with different levels of control and detail.

Data Dictionary Model Table

Column Name	Field Type	Size	Is Key	Comment
DataDictModel_DB	sysname		Primary	
DDM_Environment	nvarchar	250		List of environments applicable to, blank=all
DDM_Controller	nvarchar	80		Owner, or blank for production
DDM_Scope	nchar	32	Foreign	retention policy facet, database level
DDM_RetentionExpire	date/time			delete by DTS, required by all transitory datasets
DDM_FileSpec	nvarchar	250		Directory and filename of entry or directory
DDM_Implementation	nvarchar	MAX		The processes used to implement retention policy
DDM_Purpose	nvarchar	MAX		comments or other directives

Every database/table/column must be owned by someone (department or entity), the 'Controller'.
The 'Scope' is used to determine the inclusion of objects in a anonymisation and retention policies.
Permanent changes require in-depth details, whereas temporary additions/changes follow a simpler paradigm but require a 'Retention-Expire' delete-by DTS.

Data Dictionary Model Scope Table

Column Name	Field Type	Size	Key field	Comment
DDModelScopeIdent	nchar	32	Primary	retention policy scope-codes
DDMS_DBTab	ncar	10		switch for Database or Table scopes
DDMS_Description	nvarchar	MAX		description of scope for information only

This table allows the support of variant scope of the control policy facets and level of detail. Database List: 'Full Control', 'Ignore - DB', 'Ignore - Table', 'Ignore - Column' or 'Ignore - Other'; Table List: 'Controlled' or 'Ignore'.

Extra Columns

These are for Retention policy use; other columns will be required by Data-Anonymisation processes. These columns for the 'Entity' table follow the same general usage as those from the 'Model' table above.

Data Dictionary Entities Table

Column Name	Field Type	Size	Key field	Comment
DDE_Controller	nvarchar	80		Owner, if different from Model
DDE_Scope	nchar	32		retention policy facet, table level
DDE_RetentionExpire	date/time			delete by DTS, required by all temporary tables

Data Dictionary retain_queue Table

Column Name	Field Type	Size	Key field	Comment
RQ_CriteriaNoteDate	date/time			Useful for investigation purposes
RQ_CriteriaTransDate	date/time			Useful for investigation purposes

Retention Paradigm Processes

Current

The current system generally relies on two Stored Procedures to be run at intervals combined with the DD tables' data. We are waiting for confirmation of the exact paradigm used to enact DPA retention policy from 'Data-Services'.

Add Accounts to Processing Queue

Stored procedure [retain_grab_accounts], a 'Database Job' queued to run weekly or daily, that adds account to the table 'retain_queue' for erasure processing.

Process Account entries in Erasure Queue

Stored procedure [retain_execution_add_sb_queue] used in TEMP database, a 'Database Job' queued to run every 15-30 minutes, to run 1000 Account erasures each pass.

Other Processes

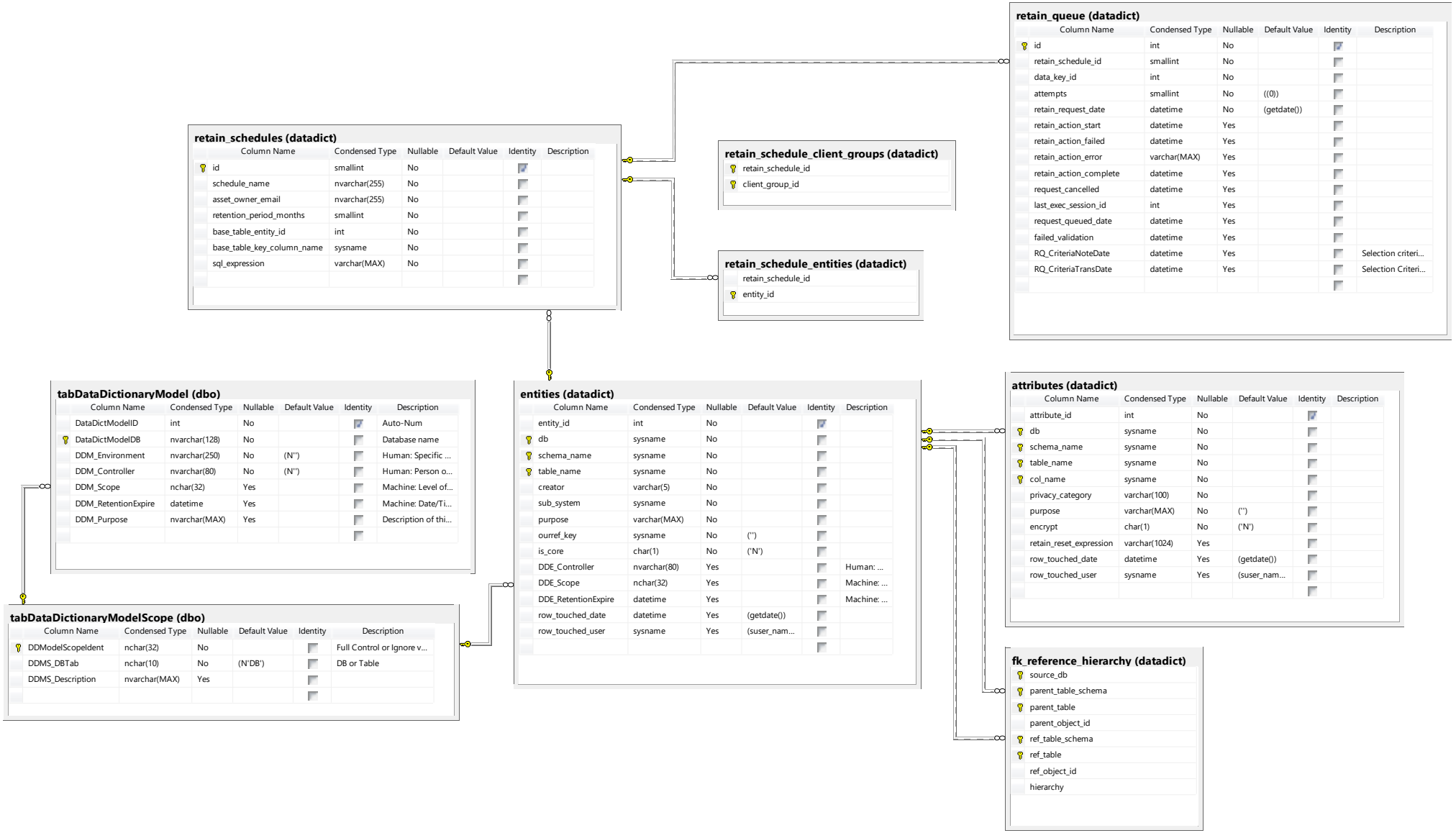
There are other stored-procedures that do other parts of the process, but I have been unable to ascertain exactly what they are and what they do.

- [fk_reference_hierarchy_populate] auto-create table
- [retain_execution_rsid_1] list of table deletes statements
- [retain_generate_procs] used to produce above SP with FK_Reference table
- [retain_execution_all] calls 'retain_execution_rsid_1' for each account in queue

Proposed

The proposed system will use same SPs, but with the extra tables and columns to enact the SCMS, which will allow for complete inclusion and control of the dataset.

Data Dictionary Data Diagram



Proposed data structure: 12 April 2018, with extra link between ‘Model’ and ‘Entity’ tables on the ‘DB’ column.