

RAE_CMDB: Anonimización de Fecha de Nacimiento y Validación de Restricciones

Equipo de Datos

16 de octubre de 2025

1. Resumen

Se ha incorporado anonimización de **FECHA_NACIMIENTO** en **DIEGO.RAE_CMDB** sin modificar los mecanismos previos. La fecha queda desplazada de forma determinista y, para mayores de 90 años, se generaliza al año. Se han validado restricciones no dependientes de tokenizaciones y se han corregido datos para cumplir reglas UCI.

2. Objetivo

- Reducir riesgo de reidentificación manteniendo utilidad analítica.
- No depender de identificadores previamente anonimizados.
- Dejar el modelo con restricciones *ENABLE VALIDATE* cuando sea posible.

3. Diseño de anonimización

Técnica aplicada

- **Desplazamiento determinista** $d \in [-15, +15]$ días por fila.
- **Clave estable no sensible**: NUM_REGISTRO_ANUAL (PK).
- **Generalización para ≥ 90 años**: se fija a 01/07/AAAA.
- **Acotaciones**: no fechas futuras ni anteriores a 1900-01-01.

Implementación

- **Paquete nuevo**: DIEGO.ANON_DATE con funciones `anon_shift_days` y `anon_fecha_nacimiento`.
- **Trigger adicional**: DIEGO.ANON_RAE_CMDB_NAC_BIU (*FOLLOWS* al existente). Sólo toca FECHA_NACIMIENTO.
- **Backfill controlado**: actualización masiva con triggers deshabilitados y reactivación posterior.

4. Procedimiento técnico ejecutado

1. Despliegue

1. Crear/actualizar paquete DIEGO.ANON_DATE.
2. Crear trigger nuevo:

```
CREATE OR REPLACE TRIGGER DIEGO.ANON_RAE_CMDB_NAC_BIU
BEFORE INSERT OR UPDATE OF FECHA_NACIMIENTO ON DIEGO.RAE_CMDB
FOR EACH ROW
```

```

FOLLOWS DIEGO.RAE_CMDB_ANON_BIU
BEGIN
    :NEW.FECHA_NACIMIENTO :=
        DIEGO.ANON_DATE.anon_fecha_nacimiento(
            :NEW.FECHA_NACIMIENTO,
            TO_CHAR(:NEW.NUM_REGISTRO_ANUAL));
END;
/

```

3. Backfill (deshabilitar/rehabilitar triggers alrededor del UPDATE masivo).

2. Normalización para checks UCI

- **INGRESO_UCI** normalizado al dominio esperado por las *checks* ('1' ingresó, '0'/'N' no ingresó, '9' desconocido).
- **DIAS_UCI** ajustado: ≥ 1 si **INGRESO_UCI**='1'; 0 o NULL si no ingresó.

3. Validación de constraints

- **ENABLE VALIDATE** de PK y CHECKs no relacionadas con tokenización.
- **ENABLE VALIDATE** de FKs seguras. **Excluida** la FK de centro (RAE_CMDB_CENTRO_FK) por depender de claves tokenizadas en HOSPITALES.
- Para FKs de diagnóstico secundario: limpieza de códigos inválidos o nulos antes de validar.

5. Comprobaciones realizadas

Anonimización

1. **Formato visual**: ALTER SESSION SET NLS_DATE_FORMAT='DD/MM/YYYY'.
2. **Determinismo por PK**: mismo NUM_REGISTRO_ANUAL \Rightarrow misma fecha anon.
3. **Rango**: sin fechas futuras ni < 1900 .
4. **Regla ≥ 90** : 01/07/AAAA.

Restricciones

1. Listado de constraints en estado VALIDATED vs NOT VALIDATED.
2. Tabla de excepciones DIEGO.CONSTR_EXC vacía tras la validación.
3. Para DIAGNOSTICO_2..20: nulos o códigos corregidos para que existan en DIEGO.DIAGNOSTICO(CODIGO).

6. Resultados

- **FECHA_NACIMIENTO** anonimizadas para datos históricos y futuras cargas por trigger.
- **Checks y FKs seguras** pasadas a **ENABLE VALIDATE**.
- **UCI**: datos coherentes con reglas de negocio.

7. Riesgos y pendientes

- **FK de centro**: si se desea validar, crear HOSPITALES.CCN_TOKEN y apuntar la FK a esa columna.
- **Procedimientos y diagnósticos**: asegurar catálogo maestro completo para futuras validaciones.

8. Snippets de verificación (SQL)

Fechas

```
ALTER SESSION SET NLS_DATE_FORMAT='DD/MM/YYYY';

-- Muestra
SELECT NUM_REGISTRO_ANUAL, FECHA_NACIMIENTO
FROM   DIEGO.RAE_CMDB FETCH FIRST 20 ROWS ONLY;

-- >=90 años a 01/07
SELECT COUNT(*) may_90_con_01jul
FROM   DIEGO.RAE_CMDB
WHERE  MONTHS_BETWEEN(TRUNC(SYSDATE), FECHA_NACIMIENTO)/12 >= 90
AND    TO_CHAR(FECHA_NACIMIENTO,'DD/MM') = '01/07';

-- Sin futuro ni <1900
SELECT
    SUM(CASE WHEN FECHA_NACIMIENTO > TRUNC(SYSDATE) THEN 1 ELSE 0 END) futuro,
    SUM(CASE WHEN FECHA_NACIMIENTO < DATE '1900-01-01' THEN 1 ELSE 0 END) muy_antiguo
FROM DIEGO.RAE_CMDB;
```

Constraints

```
-- Estado
SELECT constraint_name, constraint_type, status, validated
FROM   user_constraints
WHERE  table_name='RAE_CMDB'
ORDER BY constraint_name;

-- Excepciones
SELECT * FROM DIEGO.CONSTR_EXC ORDER BY constraint, row_id;

-- Violaciones UCI (si existieran)
SELECT ROWID, INGRESO_UCI, DIAS_UCI
FROM   DIEGO.RAE_CMDB
WHERE  NOT (
    (INGRESO_UCI IN ('N','O') AND (DIAS_UCI IS NULL OR DIAS_UCI = 0)) OR
    (INGRESO_UCI = '1'          AND DIAS_UCI >= 1) OR
    (INGRESO_UCI NOT IN ('S','N','1','O') OR INGRESO_UCI IS NULL)
);
```

9. Conclusión

La tabla RAE_CMDB queda con la fecha de nacimiento protegida por desplazamiento determinista con generalización para edades avanzadas, y con las restricciones de integridad validadas en los ámbitos no afectados por tokenización. Se mantienen todas las reglas previas.