

Plataforma Transversal de Dades

Qualitat de la dada

Novembre 2024

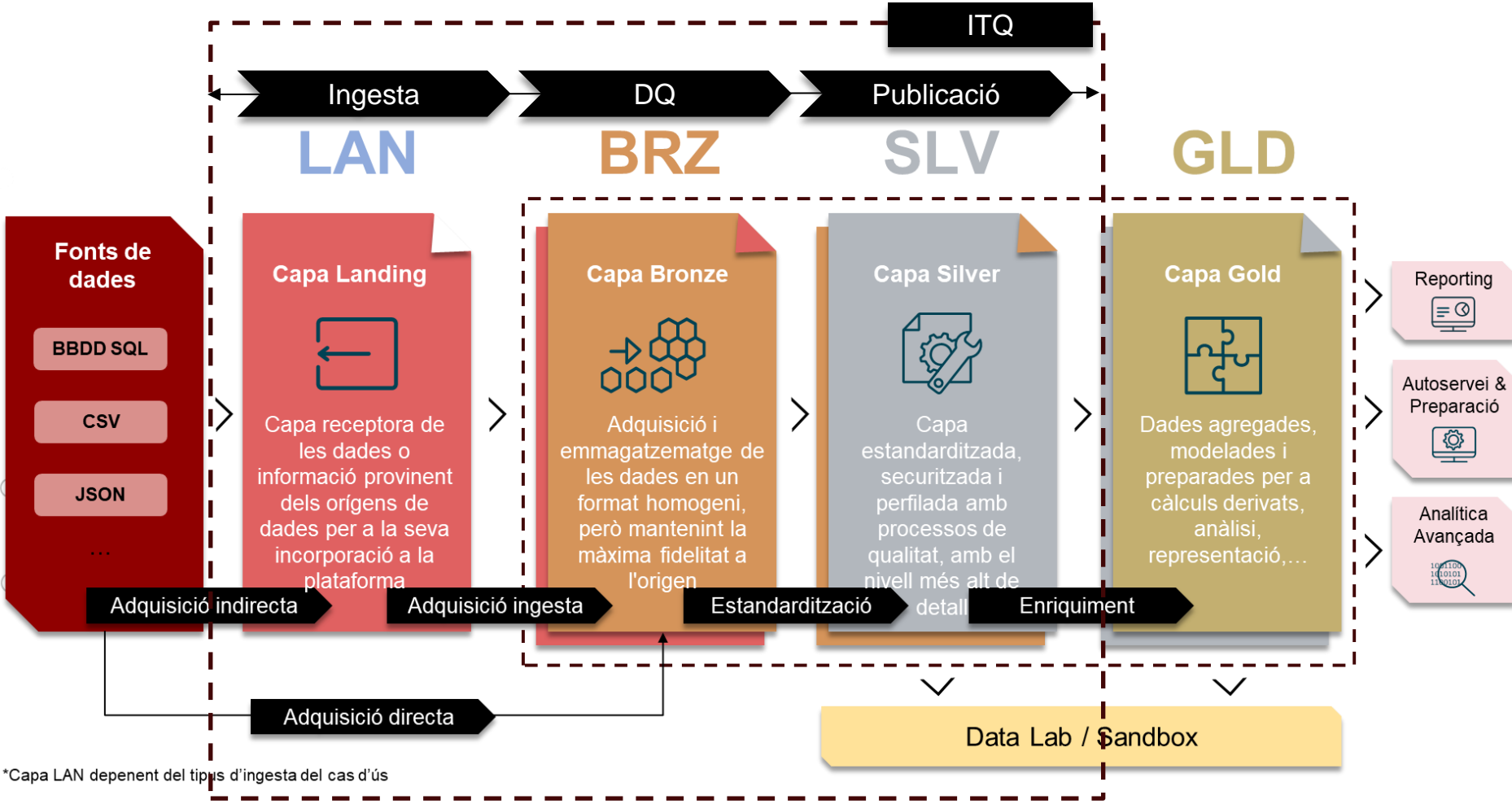


1

Capes lògiques i estàndards tècnics

Capes lògiques i estàndards tècnics

Distribució de la PTD per capes lògiques



*Capa LAN dependent del tipus d'ingesta del cas d'ús

Capes lògiques i estàndards tècnics

Nomenclatura

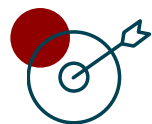
Per tal de poder utilitzar les dades per desenvolupar casos d'ús, és necessari que aquestes estiguin **estandarditzades** i compleixin una sèrie de requisits de nomenclatura (noms de taules i camps):



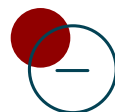
Noms **auto-explicatius**



Noms de camps escrits **en minúscules**



Notacions **clares i concises**



Utilitzar “_” **com a separador** a canvi de l'espai



Tot escrit **en català**



Utilitzar els **prefixos definits per identificar la naturalesa** de taules i camps



Noms de taules i camps escrits **en singular**



Evitar l'ús d'abreviatures comunes a diversos àmbits

Complir amb tots aquests requisits ens ajudarà a obtenir noms **representatius, funcionals i estandarditzats** que ens facilitaran la **comprensió dels nostres actius de dades**.

Capes lògiques i estàndards tècnics

Nomenclatura

El nom de les **taules** ha d'anar precedit per un dels següents prefixos definits que determinen la tipologia de la taula:

- | | |
|---|--|
| <ul style="list-style-type: none">• “DIM”: Taula dimensional que serveix per a classificar o categoritzar un fet.• “FAC”: Taula <i>fact o</i> de fets que conté els valors de les mesures o indicadors de negoci.• “AGG”: Taula agregada d'una taula de fets respecte alguna de les seves dimensions d'anàlisi.• “REL”: Taula relacional que s'utilitza per establir el vincle entre dos dimensions.• “LKP”: Taula lookup o diccionari que s'utilitza per a la traducció o descodificació de valors.• “SUK”: Taula per identificar claus subrogades. | <ul style="list-style-type: none">• “AUX”: Taula auxiliar, per necessitat d'un pas intermedi dins d'un procés tècnic.• “REF”: Taula que conté dades de referència.• “MES”: Taules mestres.• “OUT”: Taules de sortida que contenen els resultats finals d'un procés d'anàlisi de dades o d'aplicació d'un model d'aprenentatge automàtic.• “FTR”: Taules que contenen les característiques (features) que s'utilitzaran com a variables independents en un model d'aprenentatge automàtic. |
|---|--|

Capes lògiques i estàndards tècnics

Nomenclatura

El nom dels **camps** ha d'anar precedit per un dels següents prefixos definits que determinen la seva tipologia:

- | | |
|---|---|
| <ul style="list-style-type: none">• “IDE”: Identificador. Es recomana utilitzar identificadors del tipus “integer” excepte en casos on pugui ser necessari tipus “varchar”.• “DAT”: Data i hora. Per defecte tindrà una precisió nivell dia i hora.• “HRS”: Data i hora. Per defecte tindrà una precisió a nivell de dia i hora.• “TMS”: “Timestamp” o Marca de temps. Utilitzat per indicar moments precisos amb precisió fins al segon, per exemple, per indicar quan s’ha executat una càrrega.• “DES”: Descriptor. | <ul style="list-style-type: none">• “FLG”: “Flag” o Booleà (boolean).• “IMP”: Import. Els imports podran tenir associats el sufix <code>_{TIPO MONEDA}</code> on s’especifica si el valor de la moneda es local (“LC”) o l’identificador propi de la moneda d’extensió fixe 3 caràcters (ISO 4217).• “NUM”: Indicador numèric sempre i quant no sigui un import o percentatge.• “PER”: Indicador numèric per percentatges.• “URL”: Utilitzat per anomenar camps que contenen una direcció web (URL). |
|---|---|

Capes lògiques i estàndards tècnics

Eina de govern de les dades Cloud de la Generalitat (IDMC)

Disponibilitat i entorns

Entorn únic de prova que passarà a ser l'entorn productiu. **No esta disponible** pels usuaris.

Ús actual

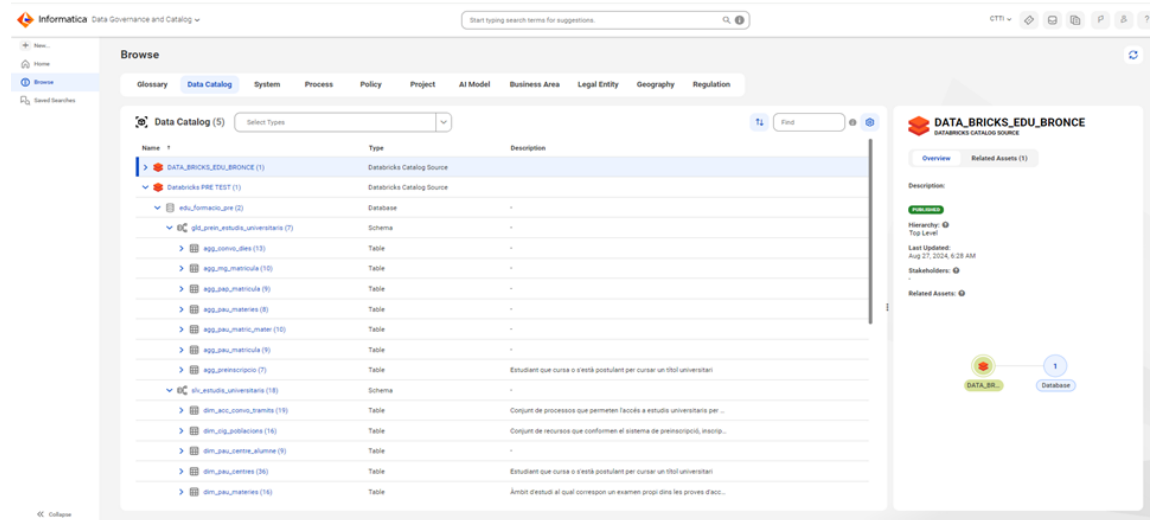
El glossaris de termes de negoci s'introdueixen a l'eina mitjançant plantilles (de la mateixa manera que amb l'eina On-Premise), mentre que els actius tècnics s'introdueixen a partir d'establir una connexió amb la BBDD i dur a terme escanejjs per tal d'extreure les metadades.

Contingut

- Dominis i subdominis (Glossari)
- Cas d'ús de REU, sobre el que s'estan fent totes les proves:
 - Taules i camps dels esquemes de BRZ, SLV i GLD
 - Glossari de termes
 - Perfilats de les taules (Distinct, Non-Distinct i Null)

Proves en curs

- Rols, permisos i responsabilitats
- Càrrega d'actius tècnics mitjançant escanejjs a la BBDD
- Càrrega d'actius tècnics mitjançant plantilles excel
- Perfilats i qualitat de la dada
- Representació del llinatge tècnic



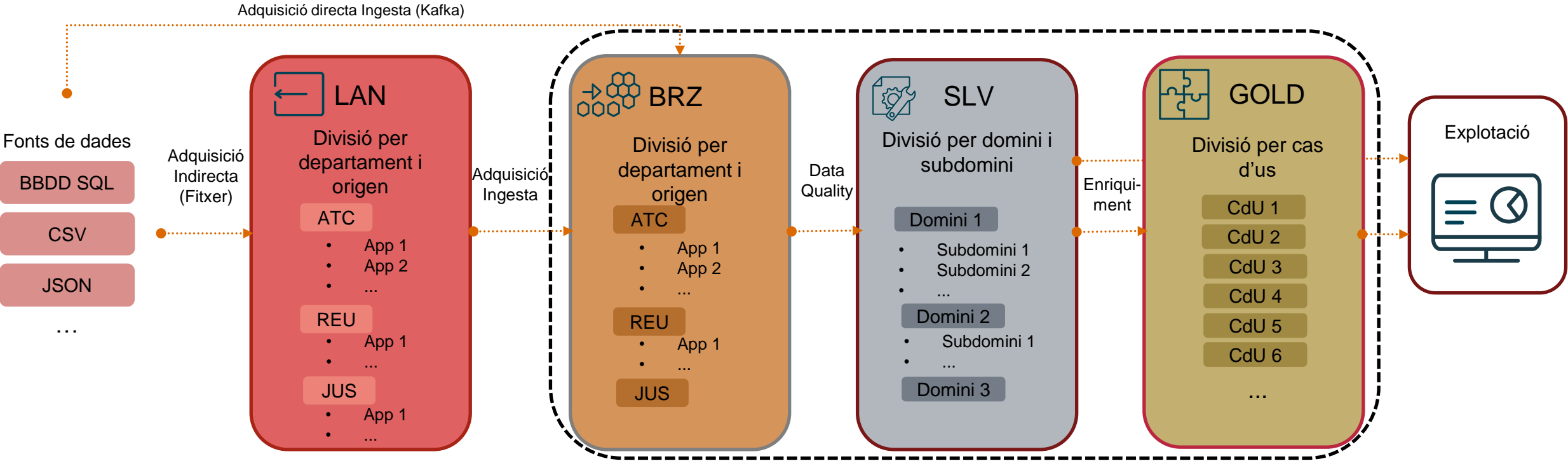
2

Aspectes de seguretat

Seguretat

Distribució funcional de la PTD per capes lògiques

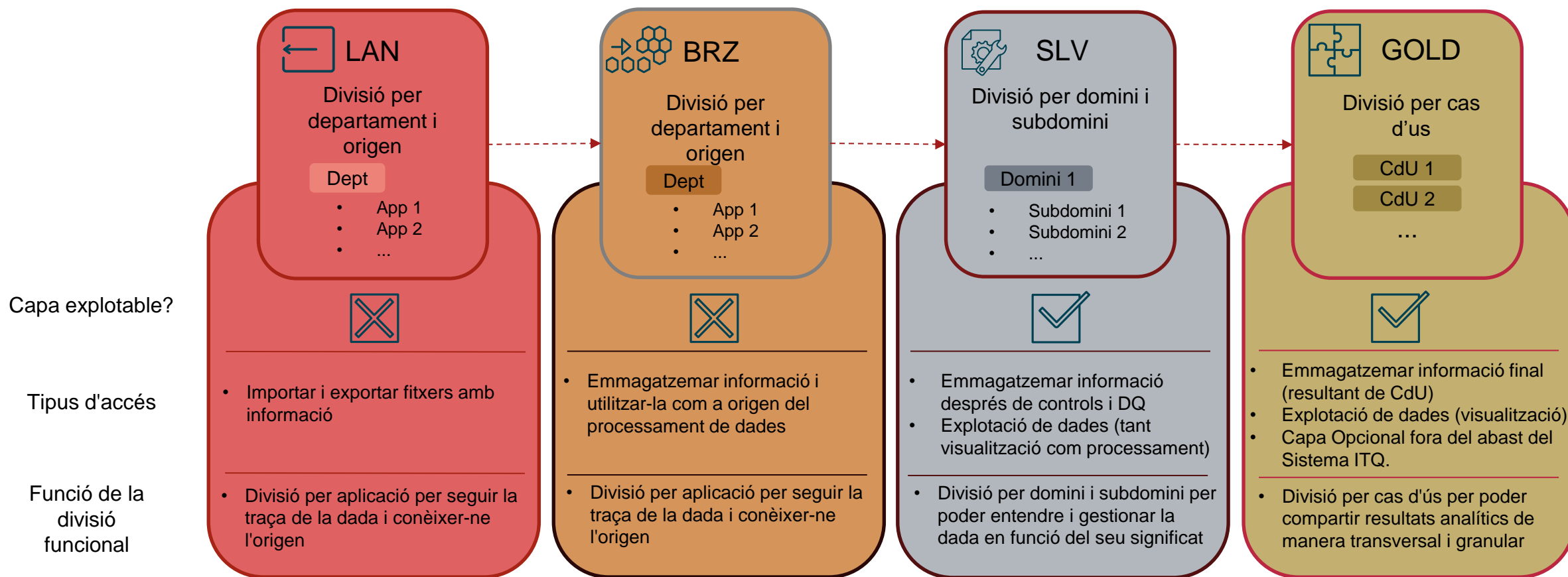
Cada capa de la plataforma tindrà una separació funcional depenent de la seva finalitat i l'estat de la dada. Landing i Bronze es distribuïran per origen de dades, Silver per subdomini funcional i Gold per Cas d'ús:



Seguretat

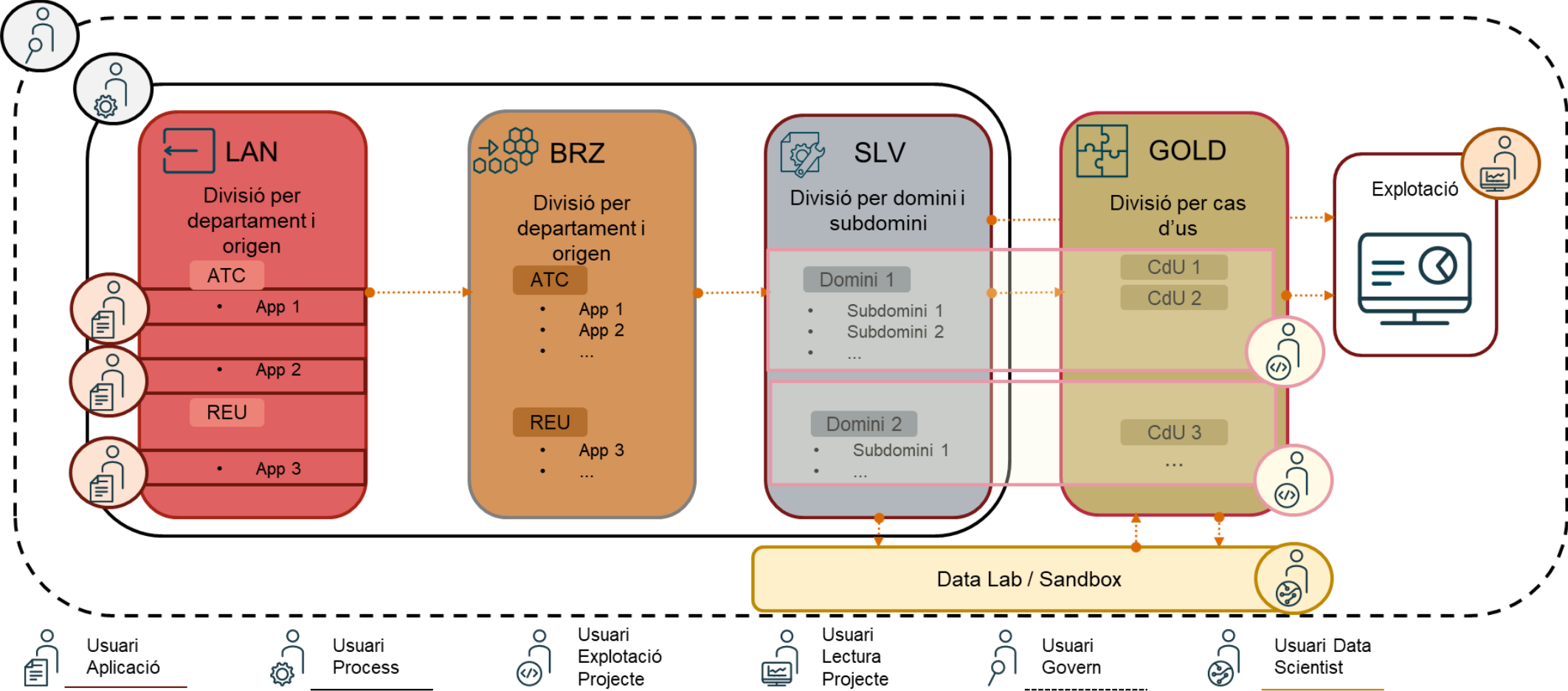
Distribució funcional de la PTD per capes lògiques

Cada capa tindrà una política d'accés depenent de la finalitat amb què ha estat definida i les subdivisions funcionals dins d'elles permetran la gestió granular de permisos:



Seguretat

Esquema de permisos d'usuaris al llarg de la plataforma



3

Aspectes de privacitat

Privacitat

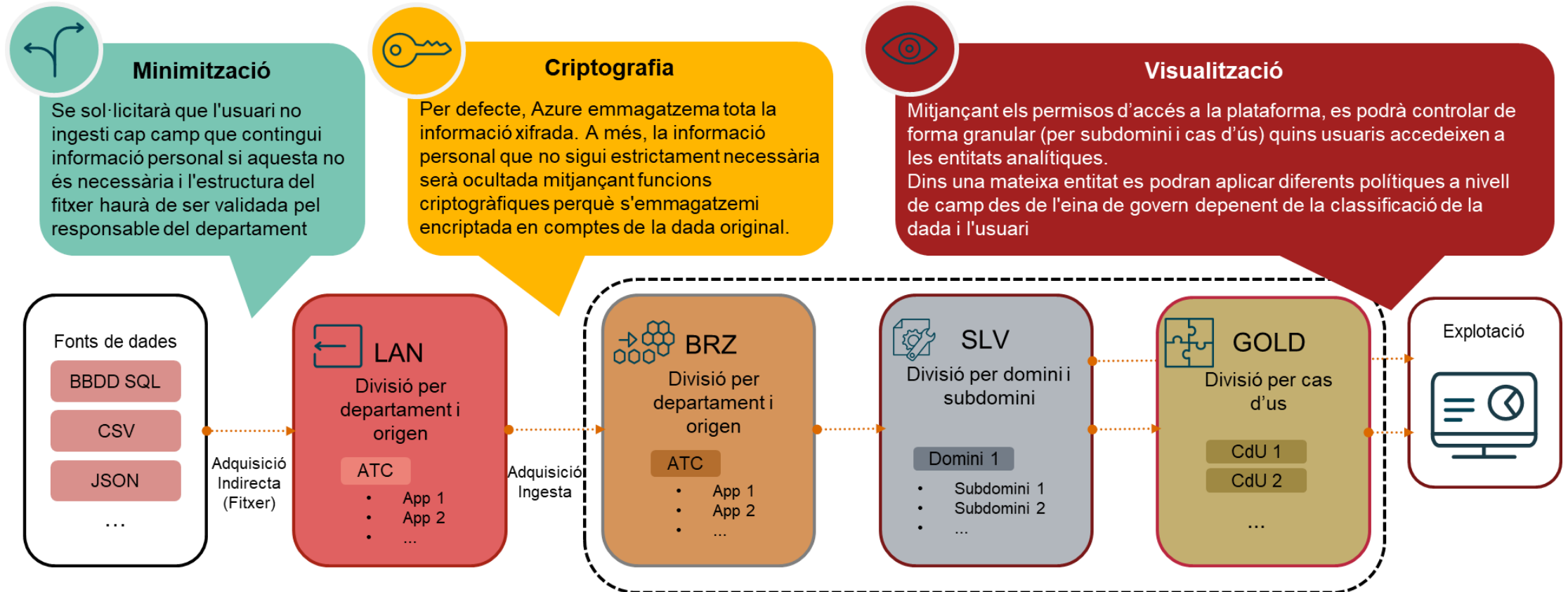
Els aspectes de privacitat es basen en 3 grans pilars:



Privacitat

Protecció a l'accés a les dades sensibles de tota la plataforma

Cada mesura per garantir la privadesa s'aplicarà en una fase diferent del cicle de la dada per assegurar-ne l'ús correcte en tot moment:

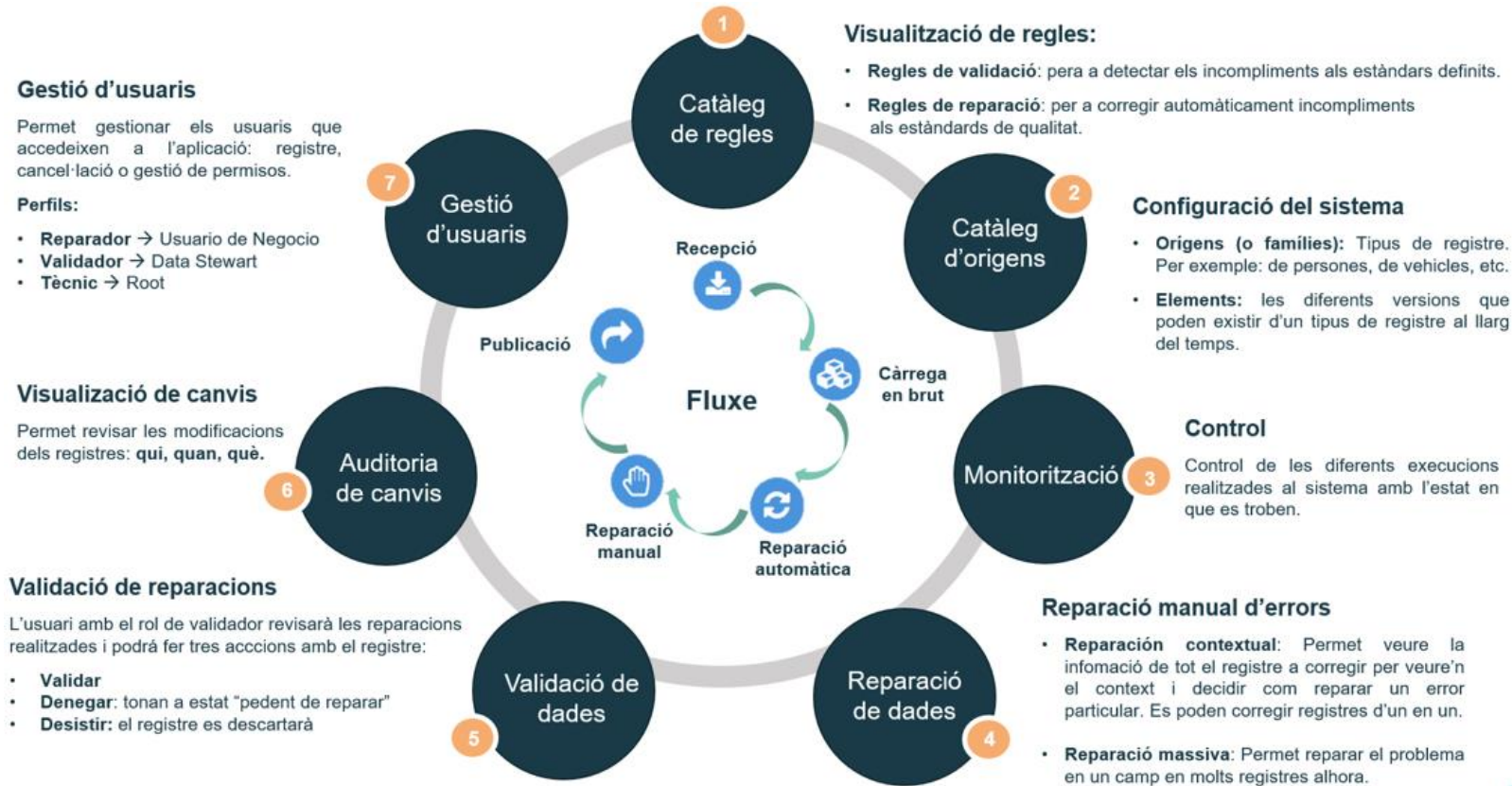


4

Mòdul de qualitat

Mòdul de qualitat

Solució configurable mitjançant l'aplicació de regles de validació, remeiació i enriquiment



Regles de Validació

Criteria aplicats a les dades per assegurar que compleixen uns estàndards abans de ser processades.



Regles de remeiació

Processos dissenyats per corregir problemes en les dades identificades durant les validacions.



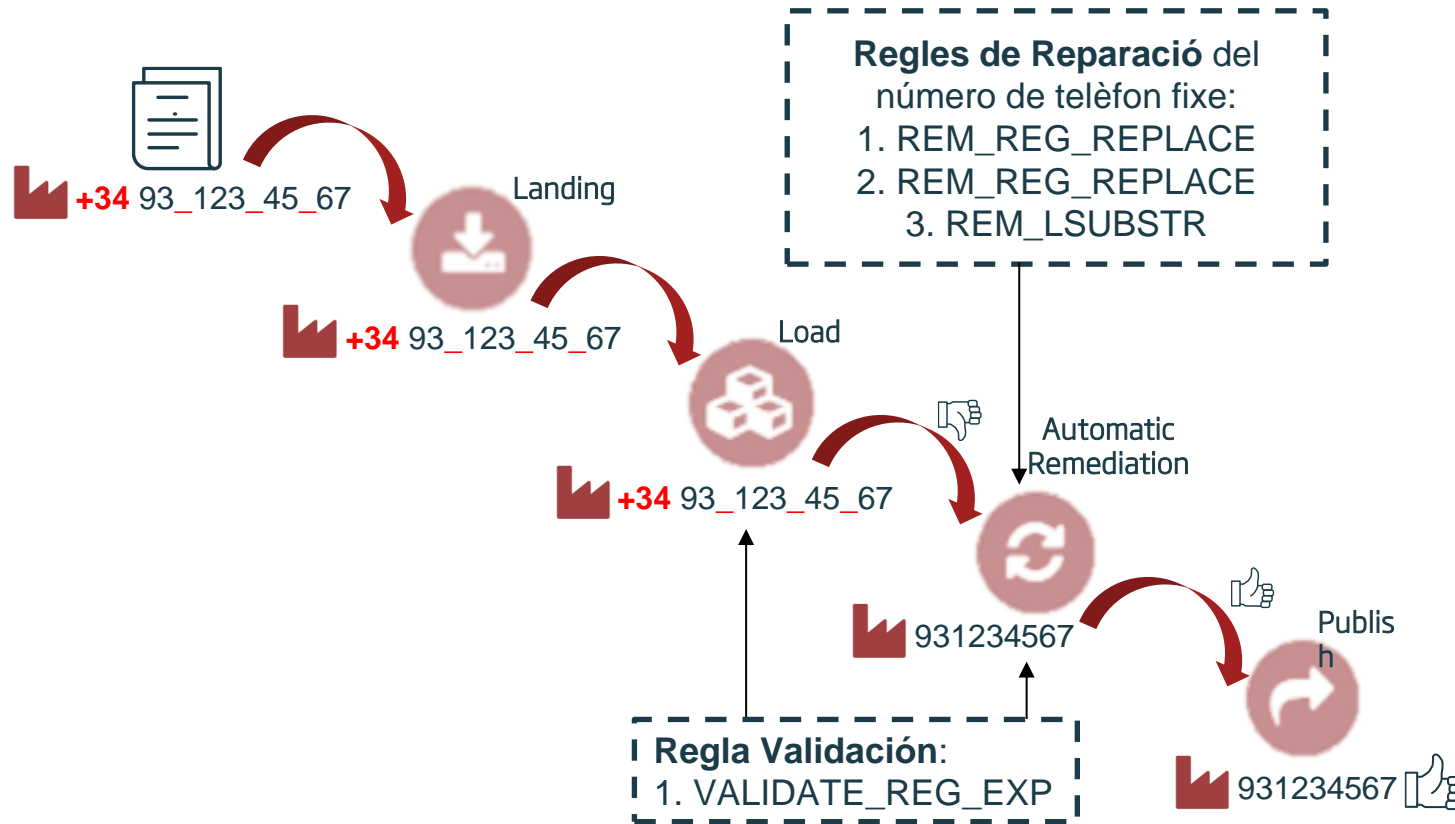
Regles d'enriquiment

Processos utilitzats per a millorar les dades, afegint informació addicional o context per a fer-les més útils i completes

Mòdul de qualitat

Exemple de funcionament a alt nivell del mòdul de Qualitat del sistema ITQ

Validem el format d'un camp telèfon que ha de complir aquest format 9XYNNNNNN / 8XYNNNNNN on X és sempre diferent de zero:



L'algorisme de la **Regla de Validació** verifica que un número de telèfon fixe es vàlid si i només si:

1. VALIDATE_REG_EXP:
 - Té longitud de 9 dígit.
 - Està format per caràcters només numèrics.
 - Comença per 9 o 8, el següent dígit no és 0 i a continuació venen 6 dígit més.

Quan un camp no passa una regla de validació l'algorisme executa una o més **Regles de Reparació** automàtica. En aquest cas:

1. REM_REG_REPLACE: Eliminem qualsevol caràcter que no sigui numèric del número de telèfon fixe.
2. REM_REG_REPLACE: Eliminem el prefix 34 del número de telèfon fixe si escau.
3. REM_LSUBSTR: Ens quedem amb els nou primers dígit com a número de telèfon fixe vàlid.

5

Estudi d'un fitxer d'exemple



Generalitat de Catalunya

Centre de Telecomunicacions i Tecnologies de la Informació (CTTI)

www.gencat.cat

ptd.ctti@gencat.cat