



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

UNIVERSITAT POLITÈCNICA DE VALÈNCIA  
PROGRAMA DE DOCTORADO EN TECNOLOGÍAS PARA LA SALUD Y EL BIENESTAR

**Diseño, construcción y evaluación de repositorios estandarizados con calidad de datos asegurada para la monitorización de la atención a la alimentación infantil**

TESIS DOCTORAL

Presentado por:

**Ricardo García de León Chocano**

Dirigida por:

Prof. Juan Miguel García Gómez

Dr. Carlos Sáez Silvestre

Valencia

Mayo 2022



# Agradecimientos

Esta tesis es el resultado de años de trabajo compaginando la investigación con el desarrollo diario de los sistemas y tecnologías de la información del Hospital Virgen del Castillo. Esto ha sido posible gracias al gran equipo de trabajo que me ha acompañado durante estos años.

En primer lugar, quiero agradecer a mis directores de tesis, Dr. Juan Miguel García-Gómez y Dr. Carlos Sáez, la confianza mostrada desde el primer momento en que compartimos los trabajos preliminares de calidad de datos que habíamos desarrollado en el hospital. Su entusiasmo, implicación y el trabajo compartido con ambos han sido el motor de esta tesis. Solo tengo palabras de admiración y afecto hacia ellos, estimables maestros y amigos que tanto me han enseñado durante estos años.

También quiero agradecer su apoyo y participación a los profesionales del entorno de investigación de mis directores de tesis, que colaboraron en artículos y proyectos relacionados con esta tesis, especialmente a la Dra. Montserrat Robles, Dr. David Moner y Dr. Jose Alberto Maldonado.

Las alianzas con otras instituciones y organismos han sido claves en el desarrollo de esta tesis, permitiendo aunar esfuerzos y compartir resultados en proyectos conjuntos. En este sentido, quiero agradecer a Pablo Serrano del Hospital 12 de Octubre y a Paloma Casado e Isabel Espiga del Ministerio de Sanidad, Servicios Sociales e Igualdad, su confianza y facilitación del desarrollo de estos proyectos. Y especialmente a Toni Oliver de la Universidad de Alicante por su inestimable ayuda, implicación y participación activa en los trabajos finales de esta tesis, por lo que le estoy profundamente agradecido. Sus contribuciones han enriquecido notablemente esta tesis.

Por parte del Hospital Virgen del Castillo y Área de Salud V - Altiplano, agradezco su compromiso a todos los directores gerentes que apostaron desde su inicio por los proyectos de lactancia materna y de calidad de datos, facilitando el desarrollo de los trabajos que han originado esta tesis. Particularmente al Dr. Francisco Soriano por su especial apoyo, afecto y confianza depositada en mí, propiciando mi desarrollo profesional en esta última etapa y haciéndome partícipe de su gestión. Y al Dr. Manuel Alcaraz, que tuvo un papel clave en el desarrollo del proyecto de lactancia materna ligado a esta tesis, su dilatada experiencia en la gestión y su amplio conocimiento de la organización de los servicios sanitarios nos supo guiar con sabiduría en el establecimiento de todos los proyectos relacionados con esta tesis. También agradezco a los pediatras, enfermeras y auxiliares del Servicio de Pediatría y tercera planta del hospital su participación en la definición de los modelos de datos y aspectos clínicos asociados a esta tesis. A mis compañeros del Servicio de Sistemas y Tecnologías de la Información, especialmente a Luis Clemente, por su ayuda en el diseño y construcción de repositorios. Y, por supuesto, a Verónica Muñoz, que participó activamente, desde su inicio, en los aspectos clínicos y técnicos de todos los trabajos de esta tesis.

Y quiero agradecer, muy especialmente, por su ilusión y su enorme apoyo a mi padre, Dr. Ricardo García de León, Jefe de Servicio de Pediatría del Área de Salud V - Altiplano durante la práctica totalidad del desarrollo de esta tesis y actualmente jubilado pero en incansable trabajo activo. Mis primeros pasos en informática fueron junto a él, aprendimos y comenzamos a programar juntos cuando yo todavía estaba en el colegio y ha sido un referente para mí durante toda mi formación y desarrollo profesional. He tenido la enorme suerte de compartir experiencia laboral y todo este proyecto de tesis doctoral con mi padre, sin él, esta tesis hubiese sido imposible. El entusiasmo e ilusión que compartíamos por la obtención y explotación de repositorios de datos clínicos de alimentación infantil fueron el germen de esta tesis que tantas horas de trabajo nos ha llevado a compartir. Su incommensurable trabajo para la salud y bienestar de los bebés y sus mamás desde todos los ámbitos: asistencial, calidad, investigación y enfoque comunitario, promovieron una impor-

tante cultura de mejora y aprendizaje en el hospital y todo el área de salud. Todos estos años de trabajo, ilusión y esfuerzo, han dejado una huella imborrable en el Área V y, por su puesto, en mí.

Y no puedo dejar de agradecer a mi madre, Conchi, su apoyo, su cariño y su confianza a lo largo de toda mi vida, que, sin duda alguna, me motivaron a estudiar una carrera desde pequeño, cuando pocos confiaban en que pudiese hacerlo, ayudándome a abrirme paso por este camino que hoy culmina con esta tesis. Y también, por su puesto, a mis hermanos, con los que he compartido buena parte de los momentos importantes de mi vida.

Finalmente, quiero dedicar esta tesis a mi mujer, Carmen, y a mis hijos, Eva y Ricardo, con todo mi cariño. Su apoyo y su amor incondicional son la más grata recompensa a todo este esfuerzo y el estímulo que me ha permitido avanzar con energía e ilusión en esta etapa, dedicándole todo el tiempo necesario; tiempo que les he tomado prestado y que ahora espero devolverles con la misma ilusión.

# Abstract

The establishment of the Electronic Health Record (EHR) as a source of knowledge in the clinical setting is a key issue for healthcare organizations, both for the development of research and teaching, and for the improvement of healthcare quality through the monitoring of care. The current deployment of health information systems means that EHR data are an excellent candidate data source with which to generate repositories for these purposes, due to their population coverage and continuity over time. Infant feeding care up to two years of age represents an important area of application, given the disparity in care and its relevance to maternal and infant health.

To ensure the validity and reliability of the reuse of EHR data for research and healthcare monitoring, it is necessary to assess and assure its data quality (DQ), which requires conceptual frameworks and software tools for its control. In addition, when different levels of care or different healthcare organizations are involved, it is necessary to support continuity of care so that healthcare providers are able to exchange information on the same subject. Different health information standards are intended to support this.

The purpose of this thesis is to design a standardized and DQ assured repository construction process to build an Integrated Data Repository (IDR) for infant feeding from multi-source EHR data. The goal of this IDR is to enable the reuse of data for monitoring infant feeding care up to two years of age, focusing on breastfeeding as one of the main determinants of maternal and child health.

For DQ management, a conceptual framework based on DQ dimensions has been used for a total data quality management of the process. As health information standards to support continuity of care, the UNE-EN ISO 13 940 standard was mainly used, together with the UNE-EN ISO 13 606 standard for a two-hospital repository integration experience. The "Indicators for assessing infant and young child feeding practices" (IYCF), defined by the World Health Organization and UNICEF, were used for monitoring. This set of indicators is the current reference for the evaluation of international infant feeding strategies such as the Baby Friendly Hospital Initiative (BFHI).

The construction process of repositories has been closely related to the "Hasta Que Tú Quieras (HQTQ)" strategy, developed in the Health Area V - Altiplano of the Servicio Murciano de Salud to improve the quality of care for infant feeding. The specifications and requirements of the infant feeding care process guided the definition of the perinatal and the infant feeding follow-up up to two years information models, implemented in the EHR of Health Area V.

The construction process has been applied on real world multi-source data from the EHR of the Health Area V - Altiplano. Specifically from the Hospital Information System (HIS) of its reference hospital, Hospital Virgen del Castillo, and from the Primary Care information system of the 3 health centers of the area, during the period 2009-2018.

The application of DQ dimensions on DQ assessment and assurance procedures, defined according to the specifications and requirements of the infant feeding care process, together with the use of concepts of the UNE-EN ISO 13 940 standard, have contributed to the design of the process of building standardized and DQ assured repositories.

The application of the methods of the repository construction process on perinatal EHR data from the HIS of the Hospital Virgen del Castillo has contributed to the creation of the perinatal repository of Health Area V - Altiplano, with DQ assured and standardized according to the UNE-EN ISO 13 940 standard. Subsequently, the application of the methods of the repository construction process on EHR data from the area's primary care information system for monitoring infant feeding up to two years of age contributed to the expansion of the repository to obtain an IDR for research and monitoring of infant feeding from birth to two years of age. The bundling of the methods of the construction process has contributed to the creation of a repository construction software with DQ assured.

The experience of integrating the mother-child repositories of the Virgen del Castillo Hospital and the 12 de Octubre Hospital in Madrid has contributed to the creation of a standardized IDR in accordance with the UNE-EN ISO 13 606 standard and a proposal of UNE-EN ISO 13 606 archetypes for the integration of mother-child repositories at the national level.

The application of methods for assessment and quality assurance of data on the longitudinal feeding record of the IDR, from birth to two years of age, have contributed to the robust estimation of infant feeding indicators and the implementation of a population-based and continuous monitoring system based on EHR data.

The quality information obtained from the IDR has contributed to obtaining the Seal of Good Practices of the Ministry of Health, Social Services and Equality (MSSSI) and to the scaling-up of the perinatal information model to the whole Servicio Murciano de Salud.

The results of this thesis have been published in 6 scientific contributions (indexed journals and conference papers). The results of this thesis have also contributed to the realization of a pilot project for the MSSSI.

# Resumen

El establecimiento de la Historia Clínica Electrónica (HCE) como fuente de conocimiento en el ámbito clínico es una cuestión clave para las organizaciones sanitarias, tanto para el desarrollo de la investigación y docencia, como para la mejora de la calidad asistencial a través de la monitorización de la asistencia. El actual despliegue de los sistemas de información sanitarios supone que los datos de HCE sean una excelente fuente de datos candidata con la que generar repositorios para estos propósitos, debido a su cobertura poblacional y continuidad en el tiempo. La atención a la alimentación infantil hasta los dos años representa una importante área de aplicación, dada la disparidad en la atención y la relevancia de la misma para la salud de la madre y el bebé.

Para garantizar la validez y fiabilidad de la reutilización de datos de HCE para investigación y monitorización de la asistencia sanitaria, es necesario evaluar y asegurar su calidad de datos (DQ, del inglés *Data Quality*). Para ello, es necesario disponer de marcos conceptuales y herramientas informáticas para su control. Así mismo, cuando se abarcan distintos niveles asistenciales o diferentes organizaciones sanitarias, es necesario dar soporte a la continuidad asistencial para que los proveedores de salud sean capaces de intercambiar la información sobre un mismo sujeto. Distintos estándares de información sanitaria tienen por objeto dar soporte a la misma.

Esta tesis tiene como propósito diseñar un proceso de construcción de repositorios, estandarizados y con calidad de datos asegurada, con el que construir un repositorio de datos integrados (IDR, del inglés *Integrated Data Repository*) de alimentación infantil a partir de datos multi-fuente de HCE. El objetivo de este IDR es habilitar el re-uso de datos para la monitorización de la atención a la alimentación infantil hasta los dos años, centrado en la lactancia materna como uno de los principales determinantes de la salud materno-infantil.

Para la gestión de la calidad de datos se ha utilizado un marco conceptual basado en dimensiones de calidad de datos para una gestión de calidad de datos total del proceso. Como estándares de información sanitaria para dar soporte a la continuidad asistencial, se ha utilizado principalmente el estándar UNE-EN ISO 13 940, junto con el estándar UNE-EN ISO 13 606 para una experiencia de integración de repositorios de dos hospitales. Para la monitorización se han utilizado los “indicadores para evaluar las prácticas de alimentación del lactante y del niño pequeño” (IYCF, del inglés *Indicators for assessing infant and young child feeding practices*), definidos por la Organización Mundial de la Salud y UNICEF. Este conjunto de indicadores es la referencia actual para la evaluación de estrategias internacionales de alimentación infantil como la Iniciativa para la Humanización de la Asistencia al Nacimiento y la Lactancia o IHAN (BFHI, del inglés *Baby Friendly Hospital Initiative*).

El proceso de construcción de repositorios ha estado estrechamente relacionado con la estrategia “Hasta Que Tú Quieras (HQTQ)”, desarrollada en el Área de Salud V – Altiplano del Servicio Murciano de Salud para mejorar la calidad de la atención a la alimentación al lactante. Las especificaciones y requerimientos del proceso asistencial de atención a la alimentación del lactante guiaron la definición de los modelos de información perinatal y de seguimiento de la alimentación hasta los dos años, implementados en la HCE del Área de Salud V.

El proceso de construcción ha sido aplicado sobre *real world data* multi-fuente proveniente de la HCE del Área de Salud V – Altiplano. En concreto del Sistema de Información Hospitalario (HIS, del inglés *Hospital Information System*) de su hospital de referencia, Hospital Virgen del Castillo y del sistema de información de Atención Primaria de los 3 centros de salud del área, durante el periodo 2009 - 2018.

La aplicación de dimensiones de calidad de datos sobre procedimientos de evaluación y aseguramiento de calidad de datos, definidos según las especificaciones y requerimientos del proceso asistencial de atención a la alimentación infantil, junto con el uso de conceptos

de la norma UNE-EN ISO 13 940, han contribuido al diseño del proceso de construcción de repositorios estandarizados con calidad de datos asegurada.

La aplicación de los métodos del proceso de construcción de repositorios sobre los datos de la HCE perinatal procedentes del HIS del Hospital Virgen del Castillo han contribuido a la creación del repositorio perinatal del Área de Salud V - Altiplano, con calidad de datos asegurada y estandarizado según la norma UNE-EN ISO 13 940. Posteriormente, la aplicación de los métodos del proceso de construcción de repositorios sobre los datos de HCE del seguimiento de la alimentación infantil hasta los dos años procedentes del sistema de información de atención primaria del área, ha contribuido a la ampliación del repositorio para la obtención de un IDR para la investigación y monitorización de la alimentación infantil desde el nacimiento hasta los dos años. La paquetización de los métodos del proceso de construcción ha contribuido a la creación de un software de construcción de repositorios con calidad de datos asegurada.

La experiencia de integración de los repositorios materno-infantiles del Hospital Virgen del Castillo y el Hospital 12 de Octubre de Madrid ha contribuido a la creación de un IDR estandarizado conforme a la norma UNE-EN ISO 13 606 y a una propuesta de arquetipos UNE-EN ISO 13 606 para integración de repositorios materno-infantiles a nivel nacional.

La aplicación de los métodos de evaluación y aseguramiento de calidad de datos sobre el registro longitudinal de alimentación del IDR, desde el nacimiento hasta los dos años, han contribuido a la estimación robusta de indicadores de alimentación infantil y la habilitación de un sistema de monitorización poblacional y continuo a partir de datos de HCE.

La información de calidad obtenida del IDR ha contribuido a la obtención del Sello de Buenas de Prácticas del Ministerio de Sanidad, Servicios Sociales e Igualdad (MSSSI) y a la ampliación a escala (*scaling-up*) del modelo de información perinatal a todo el Servicio Murciano de Salud.

Los resultados de esta tesis han sido publicados en 6 contribuciones científicas (revistas indexadas y artículos en congresos). Los resultados de esta tesis también han contribuido a la realización de un proyecto piloto para el MSSSI.

# Resum

L'establiment de la Història Clínica Electrònica (HCE) com a font de coneiximent a l'àmbit clínic és una qüestió clau per a les organitzacions sanitàries, tant per al desenvolupament de la investigació i docència, com per a la millora de la qualitat assistencial a través del monitoratge de l'assistència. L'actual desplegament dels sistemes de informació sanitaris suposa que les dades de HCE siguin una excel·lent font de informació amb la qual generar repositoris de dades per a aquests propòsits, a causa de la seua cobertura poblacional i continuïtat en el temps. L'atenció a l'alimentació infantil fins als dos anys representa una important àrea d'aplicació, donada la disparitat en l'atenció i la relevància de la mateixa per a la salut de la mare i el bebé.

Per garantir la validesa i fiabilitat de la reutilització de dades de HCE per a investigació i monitoratge de l'assistència sanitària, és necessari avaluar i assegurar la seua qualitat (DQ, de l'anglès Data Quality). Per aquesta raó, és necessari disposar de marcs conceptuals i eines informàtiques per al seu control. Així mateix, quan s'abasten diferents nivells assistencials o diferents organitzacions sanitàries, és necessari donar suport a la continuïtat assistencial perquè els proveïdors de salut siguin capaços d'intercanviar la informació sobre un mateix subjecte, on diferents estàndards d'informació sanitària tenen per objecte donar suport a aquest objectiu.

Aquesta tesi té com a propòsit dissenyar un procés de construcció de repositoris, estandarditzats i amb qualitat de dades assegurada, amb el qual construir un repositori de dades integrats (IDR, de l'anglès Integrated Data Repository) d'alimentació infantil a partir de dades multi-font de HCE. L'objectiu d'aquest IDR és habilitar la reutilització de dades per al monitoratge de l'atenció a l'alimentació infantil fins als dos anys, centrat en la lactància materna com un dels principals determinants de la salut matern-infantil.

Per a la gestió de la qualitat de dades s'ha utilitzat un marc conceptual basat en dimensions de qualitat de dades per a una gestió de qualitat de dades total del procés. Com a estàndards d'informació sanitària per a donar suport a la continuïtat assistencial, s'ha utilitzat principalment l'estàndard UNE-EN ISO 13940, juntament amb l'estàndard UNE-EN ISO 13606 per a una experiència d'integració de repositoris de dos hospitals. Per al monitoratge s'han empleat els "indicadors per a avaluar les pràctiques d'alimentació del lactant i del xiquet menut" (IYCF, de l'anglès Indicators for assessing infant and young child feeding practices), definits per l'Organització Mundial de la Salut i UNICEF. Aquest conjunt d'indicadors és la referència actual per a l'avaluació d'estratègies internacionals d'alimentació infantil com la Iniciativa per a la Humanització de la Assistència al Naixement i la Lactància o IHAN (BFHI, de l'anglès Baby Friendly Hospital Initiative).

El procés de construcció de repositoris ha estat estretament relacionat amb l'estratègia "Hasta Que Tú Quieras (HQTQ)" desenvolupada a l'Àrea de Salut V - Altiplano del *Servicio Murciano de Salud* per a millorar la qualitat de l'atenció a l'alimentació al lactant. Les especificacions i requeriments del procés assistencial d'atenció a l'alimentació del lactant van guiar la definició dels models d'informació perinatal i de seguiment de l'alimentació fins als dos anys, implementats en la HCE de l'Àrea de Salut V.

El procés de construcció ha sigut aplicat sobre *real world data* multi-font provinent de la HCE de l'Àrea de Salut V – Altiplano del *Servicio Murciano de Salud*. En concret del Sistema d'Informació Hospitalari (HIS, de l'anglès Hospital Information System) del seu hospital de referència, Hospital Virgen del Castillo i del sistema d'informació d'Atenció Primària dels tres centres de salut de l'àrea, durant el període 2009-2018.

L'aplicació de dimensions de qualitat de dades sobre procediments d'avaluació i assegurament de qualitat de dades, definits segons les especificacions i requeriments del procés assistencial d'atenció a l'alimentació infantil, juntament amb l'ús de conceptes de la norma UNE-EN ISO 13940, han contribuït al disseny del procés de construcció de repositoris estandarditzats amb qualitat de dades assegurada.

L'aplicació dels mètodes del procés de construcció de repositoris sobre les dades de la HCE perinatal procedents del HIS de l'Hospital Virgen del Castillo han contribuït a la creació del repositori perinatal de l'Àrea de Salut V - Altiplano, amb qualitat de dades assegurada i estandarditzat segons la norma UNE-EN ISO 13 940. Posteriorment, la aplicació dels mètodes del procés de construcció de repositoris sobre les dades de HCE del seguiment de l'alimentació infantil fins als dos anys procedents del sistema d'informació d'atenció primària de l'àrea, ha contribuït a l'ampliació del repositori per a l'obtenció d'un IDR per a la investigació i monitoratge de l'alimentació infantil des del naixement fins als dos anys. L'empaquetament dels mètodes del process de construcció ha contribuït a la creació d'un programari de construcció de repositoris amb qualitat de dades assegurada.

L'experiència d'integració dels repositoris matern-infantils de l'Hospital Virgen del Castillo i l'Hospital 12 d'Octubre de Madrid ha contribuït a la creació d'un IDR estandarditzat conforme a la norma UNE-EN ISO 13606 i a una proposta d'arquetips UNE-EN ISO 13606 per a integració de repositoris matern-infantils a nivell nacional.

L'aplicació dels mètodes d'avaluació i assegurament de qualitat de dades sobre el registre longitudinal d'alimentació del IDR, des del naixement fins als dos anys, han contribuït a l'estimació robusta d'indicadors d'alimentació infantil i l'habilitació d'un sistema de monitoratge poblacional i continu a partir de dades de HCE.

La informació de qualitat obtinguda del IDR ha contribuït a l'impacte del IDR en la Area de Salut V – Altiplano, específicament a l'obtenció del Segell de Bones de Pràctiques del Ministeri de Sanitat, Serveis Socials i Igualtat (MSSSI) i al "*scaling-up*" del model d'informació perinatal a tot el Servei Murcià de Salut.

Els resultats d'aquesta tesi han sigut publicats en sis contribucions científiques (revistes indexades i articles en congressos). Els resultats d'aquesta tesi també han contribuït a la realització d'un projecte pilot per al MSSSI.

# Acrónimos

<b>ABF</b>	Cualquier tipo de lactancia materna, (del inglés <i>Any breastfeeding</i> )
<b>BBDD</b>	Base de datos
<b>BBPP</b>	Buenas Prácticas, referido a Buenas Prácticas en el Sistema Nacional de Salud
<b>BDSLlab</b>	Biomedical Data Science Laboratory
<b>BFHI</b>	Iniciativa para la Humanización de la Asistencia al Nacimiento y la Lactancia o IHAN, del inglés <i>Baby Friendly Hospital Initiative</i> )
<b>CMDIC</b>	Conjunto Mínimo de Datos de Informes Clínicos
<b>CPD</b>	Centro de Proceso de Datos
<b>DMAIC</b>	Definir, Medir, Analizar, Mejorar, Controlar (del inglés <i>Define, Measure, Analyze, Improve, Control</i> )
<b>DQ</b>	Calidad de Datos (del inglés <i>Data Quality</i> )
<b>EBF</b>	Lactancia materna exclusiva, (del inglés <i>Exclusive breastfeeding</i> )
<b>EHR</b>	Electronic Health Record
<b>FBF</b>	Lactancia materna completa, (del inglés <i>Full breastfeeding</i> )
<b>ETL</b>	Extracción, Transformación y Carga, (del inglés <i>Extraction, Transformation and Load</i> )
<b>HdE</b>	Hecho de Estudio
<b>IIS</b>	Del inglés <i>Internet Information Server</i>
<b>MSSSI</b>	Ministerio de Sanidad, Servicios Sociales e Igualdad
<b>HCE</b>	Historia Clínica Electrónica
<b>HIS</b>	Sistema de Información Hospitalario (del inglés, <i>Hospital Information System</i> )
<b>HQTQ</b>	Experiencia “Hasta Que Tú Quieras” del Área V de Salud - Altiplano
<b>IDR</b>	Repositorio de datos integrados (del inglés <i>Integrated Data Repository</i> )
<b>IYCF</b>	Indicadores para evaluar las prácticas de alimentación del lactante y del niño pequeño, definidos por la OMS/UNICEF, (del inglés <i>Indicators for assessing infant and young child feeding practices</i> )

<b>IFS</b>	Estado de alimentación infantil, (del inglés <i>Infant Feeding Status</i> )
<b>OMS</b>	Organización Mundial de la Salud
<b>PAAL</b>	Proceso de atención a la alimentación infantil, definido en el proyecto HQTQ-2
<b>PAI</b>	Proceso Asistencial Integrado
<b>PANA</b>	Programa de Atención al Niño y Adolescente de la Región de Murcia ( <a href="https://www.murciasalud.es/pagina.php?id=110673&amp;idsec=182">https://www.murciasalud.es/pagina.php?id=110673&amp;idsec=182</a> )
<b>PDCA</b>	Planear, Hacer, Controlar, Actuar, (del inglés, <i>Plan, Do, Check, Act</i> )
<b>SNS</b>	Sistema Nacional de Salud
<b>TDQM</b>	Gestión Total de la Calidad de Datos, (del inglés <i>Total Data Quality Management</i> )
<b>UPV</b>	Universidad Politécnica de Valencia
<b>VPN</b>	Red Privada Virtual, (del inglés <i>Virtual Private Network</i> )

# Índice

1	Introducción.....	1
1.1	Motivación .....	1
1.2	Justificación y fundamentos de la tesis.....	2
1.2.1	La alimentación infantil durante los 1000 primeros días de vida y la estrategia internacional Baby Friendly Hospital Initiative (BFHI) .....	3
1.2.2	La calidad de datos biomédicos .....	4
1.2.3	Estándares de información sanitaria para dar soporte a la continuidad asistencial .....	5
1.2.4	La experiencia “Hasta Que Tú Quieras” del Área de Salud V – Altiplano del Servicio Murciano de Salud .....	6
1.3	Preguntas de investigación y objetivos.....	8
1.4	Contribuciones.....	9
1.4.1	Principales contribuciones.....	9
1.4.2	Publicaciones científicas.....	10
1.5	Proyectos de investigación.....	11
1.6	Esquema de la tesis .....	12
2	Diseño del proceso de construcción de repositorios de alimentación infantil .....	15
2.1	Introducción .....	15
2.2	Antecedentes.....	16
2.3	Metodología.....	18
2.3.1	El modelo de información del PAAL.....	18
2.3.2	Marco de calidad de datos .....	20
2.3.3	Definición del proceso de construcción de repositorios del PAAL.....	22
2.4	Resultados.....	23
2.4.1	Fase I. Definición de los parámetros del estudio y selección de variables informativas .....	25
2.4.2	Fases II y III. Obtención de una estructura homogénea y estable .....	25
2.4.3	Fases IV y V. Abordaje de los criterios de integridad semántica de los datos a nivel univariante y preparación de la estructura canónica del repositorio.....	26
2.4.4	Fase VI. Evaluación de la integridad semántica multivariante de los conceptos clínicos.....	27
2.5	Discusión .....	28
2.5.1	Significancia.....	28
2.5.2	Limitaciones .....	29
2.5.3	Trabajos futuros.....	30
2.6	Conclusiones .....	30
3	Construcción del repositorio perinatal .....	33

3.1	Introducción y antecedentes .....	33
3.2	Metodología.....	35
3.2.1	Materiales .....	35
3.2.2	Métodos .....	36
3.3	Resultados.....	38
3.4	Discusión .....	44
3.5	Conclusiones .....	47
4	Integración de repositorios materno-infantiles basada en estándares de información sanitaria .....	49
4.1	Introducción .....	50
4.2	Material y métodos .....	50
4.3	Resultados.....	53
4.4	Discusión .....	55
4.5	Conclusiones .....	56
5	Ampliación del repositorio con el seguimiento de la alimentación hasta los dos años .....	57
5.1	Introducción .....	57
5.2	Metodología.....	58
5.2.1	Materiales .....	59
5.2.2	Métodos .....	60
5.3	Resultados.....	63
5.3.1	Resultados de la paquetización.....	65
5.4	Discusión .....	67
6	Estimación robusta de indicadores de alimentación infantil.....	69
6.1	Introducción y antecedentes .....	69
6.2	Materiales .....	71
6.3	Métodos.....	73
6.3.1	Fase 1: Evaluación de DQ de contactos individuales.....	74
6.3.2	Fase 2: Evaluación longitudinal de DQ de la serie de contactos.....	75
6.3.3	Fase 3: Estimación robusta de indicadores .....	78
6.4	Resultados.....	79
6.4.1	Resultados de la evaluación de DQ de los contactos individuales.....	79
6.4.2	Resultados de la evaluación longitudinal de DQ.....	80
6.4.3	Resultados de aseguramiento de DQ para la estimación robusta de indicadores .....	80
6.5	Discusión .....	82
6.5.1	Limitaciones .....	84
6.5.2	Trabajos futuros.....	84
6.6	Conclusiones .....	85
7	El valor del repositorio en el Área de Salud V - Altiplano.....	87

7.1	Introducción .....	87
7.2	Metodología.....	88
7.3	Resultados.....	89
7.4	Discusión .....	90
8	Conclusiones y trabajo futuro .....	93
8.1	Conclusiones .....	93
8.2	Trabajos futuros.....	97
	Bibliografía.....	99
	Apéndice A .....	115
	Apéndice B .....	121
	Apéndice C .....	129



# 1 Introducción

Este capítulo presenta el esquema y marco de la tesis. En primer lugar, se describe la motivación para la realización de esta tesis, seguidamente se presentan la justificación y fundamentos de la misma. A continuación, se definen las preguntas de investigación y objetivos, las contribuciones derivadas de la investigación realizada en esta tesis y los proyectos de investigación relacionados con la tesis en los que se ha participado. Por último, se ofrece un esquema de la estructura de la tesis.

## 1.1 Motivación

Los 1000 primeros días de vida, etapa que comprende desde la concepción hasta los dos años, constituyen un período único de oportunidad donde se establecen las bases de una salud, crecimiento y desarrollo neurológico óptimos a lo largo de la vida ([United Nations Children's Fund, World Health Organization, World Bank, Early Childhood Development Action Network, The Partnership for Maternal, Newborn and Child Health, 2018](#)). Se trata, por tanto, de un período estratégico en términos de prevención y salud pública. Uno de los aspectos clave de este período es una nutrición adecuada, donde la lactancia materna es la norma biológica, proporcionando numerosos beneficios para la salud de la madre y el bebé ([Cattaneo, 2012; Victora et al., 2016](#)). La Organización Mundial de la Salud (OMS) recomienda lactancia materna exclusiva durante los primeros 6 meses de vida y con alimentos complementarios adecuados hasta, al menos, los 2 años de edad ([World Health Organization, 2003](#)). Sin embargo, menos de la mitad de los bebés son amamantados de acuerdo con estas recomendaciones.

La Iniciativa “*Baby Friendly Hospital Initiative*” (BFHI), conocida en castellano como *Iniciativa para la Humanización de la Asistencia al Nacimiento y la Lactancia o IHAN* es una estrategia internacional de la OMS / UNICEF destinada a mejorar las prácticas de lactancia materna en los servicios de salud ([Rollins et al., 2016](#)), con el objetivo brindar a cada bebé el mejor comienzo en la vida mediante la creación de un entorno de atención sanitaria donde la lactancia materna sea la norma. En este sentido, OMS y UNICEF también definen un conjunto de indicadores de proceso y resultado para la monitorización de la estrategia y, en general, de la alimentación infantil. La BFHI es una de las intervenciones más eficaces ([Pérez-Escamilla et al., 2016](#)) y rentables ([Horton et al., 1996](#)) para la mejora general de las tasas de lactancia materna. Un factor importante para la implementación y el mantenimiento exitosos de la BFHI es la monitorización regular de indicadores de proceso y resultado ([Pérez-Escamilla, 2007; Pérez-Escamilla & Chapman, 2012; Zakarija-Grković et al., 2018](#)). A pesar de ello, en la actualidad, una minoría de países lleva a cabo una monitorización regular de indicadores de alimentación infantil ([Pérez-Escamilla & Moran, 2016; UNICEF, 2020](#)). Una de las principales barreras para establecer un sistema de monitorización regular es la obtención de datos poblacionales fiables.

En este escenario, el actual despliegue de los sistemas de información sanitaria permite que los datos procedentes de Historia Clínica Electrónica (HCE) sean una potencial fuente de datos para generar repositorios para su re-uso en investigación y monitorización de la asistencia sanitaria, debido a su cobertura poblacional y continuidad en el tiempo. Sin embargo, la re-utilización de los datos de HCE para estos propósitos plantea importantes desafíos. En primer lugar, se debe evaluar y asegurar su calidad de datos (DQ, del inglés *Data Quality*), dado que la falta de DQ puede afectar a la interpretación de resultados,

dando lugar a decisiones sub-óptimas que pueden afectar a la investigación o a la gestión de la calidad asistencial. Para ello, es necesario disponer de marcos conceptuales de DQ y herramientas informáticas para su control. Por otro lado, los repositorios pueden integrar datos multi-fuente, procedentes de distintos niveles asistenciales o incluso de distintas organizaciones sanitarias, por lo que se debe asegurar la interoperabilidad de los datos mediante el uso de estándares de información sanitaria. Distintos estándares de información sanitaria tienen por objeto dar soporte a la continuidad asistencial y se debe valorar su incorporación al proceso de diseño y construcción de repositorios.

El Área de Salud V – Altiplano del Servicio Murciano de Salud está fuertemente comprometida con la protección, promoción y apoyo a la lactancia materna desde el año 2000. La incorporación del autor de esta tesis al Servicio de Informática del Hospital Virgen del Castillo de Yecla, en septiembre de 2002, permitió iniciar una línea de trabajo para la generación de repositorios de datos clínicos a partir de datos de HCE de alimentación infantil, para su re-uso en la mejora de la calidad asistencial a través de la monitorización de la atención sanitaria. Esta línea de trabajo se inició junto con el Servicio de Pediatría del hospital, que se encontraba dentro de un proceso de mejora de la calidad asistencial, con el Dr. García de León, padre del autor de esta tesis, como responsable clínico. Las necesidades del Servicio de Pediatría estaban centradas en la atención perinatal y, en concreto, en el adecuado registro y explotación de todos los aspectos relacionados con el establecimiento, protección, promoción y apoyo a la Lactancia Materna, dado que se estaba preparando la acreditación BFHI del hospital.

Evaluaciones preliminares de DQ de los datos clínicos extraídos de la HCE pusieron de manifiesto la necesidad de un abordaje sistemático para la construcción de repositorios y aseguramiento de su DQ, con objeto de garantizar la fiabilidad de los resultados de la monitorización de la alimentación infantil.

En este sentido, el inicio de trabajos conjuntos con el Grupo IBIME, actual Biomedical Data Science Laboratory (BDSLlab) de la Universidad Politécnica de Valencia (UPV), supuso un punto de inflexión en la forma de abordar la problemática de la DQ, estableciendo un marco formal para su evaluación y sistematización.

La relevancia del re-uso de datos de HCE para la investigación y monitorización de la asistencia sanitaria está ampliamente aceptada, así como la necesidad de evaluar su calidad de datos para garantizar su fiabilidad. Sin embargo, a pesar de que existen numerosas publicaciones relacionadas con la evaluación de DQ, se ha prestado poca atención al problema del aseguramiento de la DQ y al abordaje de la misma desde un enfoque centrado en indicadores. Específicamente, en el ámbito de la monitorización de la alimentación infantil a partir de datos de HCE, las publicaciones son muy escasas y no reportan información de DQ.

La motivación principal de esta tesis es dar respuesta a los retos que supone la generación de repositorios de datos extraídos de HCE, estandarizados y con calidad de datos asegurada, para la monitorización de la atención sanitaria; y su aplicación práctica en el ámbito específico de la alimentación infantil.

## **1.2 Justificación y fundamentos de la tesis**

En este apartado se describe la justificación y fundamentos de la tesis desde los cuatro ejes que la definen. En el apartado 1.2.1 se describe el ámbito clínico de aplicación, la alimentación infantil durante los primeros años de vida. Seguidamente, en el apartado 1.2.2 se describen los fundamentos de la DQ, como elemento imprescindible para asegu-

rar la fiabilidad del re-uso de datos de HCE en investigación y monitorización. En el apartado 1.2.3 se presentan los principales estándares de información sanitaria orientados a dar soporte a la continuidad asistencial en entornos con distintos proveedores sanitarios, como son los procesos asistenciales que cubren los dos niveles asistenciales (primaria y especializada) de un mismo área de salud o la compartición de información de salud entre distintas organizaciones sanitarias. Finalmente, en el apartado 1.2.4 se describen los antecedentes, objetivos y enfoque del re-uso de datos de HCE para la investigación y monitorización de la alimentación infantil en el Área de Salud V – Altiplano, enmarcado en una estrategia para implementar las recomendaciones de la BFHI en toda un área de salud.

### **1.2.1 La alimentación infantil durante los 1000 primeros días de vida y la estrategia internacional Baby Friendly Hospital Initiative (BFHI)**

La OMS y UNICEF establecen los 1000 primeros días de vida (desde la concepción hasta los dos años de edad) como un periodo crítico para el cuidado y la atención. Se trata de una ventana única de oportunidad donde se forman las bases para una salud, crecimiento y neurodesarrollo óptimos para el resto de la vida ([United Nations Children's Fund, World Health Organization, World Bank, Early Childhood Development Action Network, The Partnership for Maternal, Newborn and Child Health, 2018](#)).

Uno de los aspectos clave de este periodo es una nutrición adecuada. La lactancia materna es la norma biológica para la alimentación infantil durante los primeros años de vida y, por tanto, proporciona numerosos beneficios para la salud de la madre y del bebé ([Cattaneo, 2012; Victora et al., 2016](#)).

La OMS recomienda que se inicie la lactancia materna durante la primera hora de vida, que se mantenga de forma exclusiva durante los primeros 6 meses, y continúe junto con alimentos complementarios sanos y adecuados hasta los 2 años de edad o más ([World Health Organization, 2003](#)). La lactancia materna se considera uno de los principales determinantes de la salud materno-infantil y está relacionada con la prevención de factores de riesgo de las enfermedades crónicas de la infancia (obesidad, enfermedades cardiovasculares o autoinmunes), así como de la carga de enfermedad evitable ([Verschuuren et al., 2012](#)). A pesar de ello, menos de la mitad de los bebés son amamantados de acuerdo con estas recomendaciones por lo que mejorar esta realidad es fundamental para alcanzar los objetivos mundiales de nutrición, salud y supervivencia, crecimiento económico y sostenibilidad medioambiental ([Rollins et al., 2016](#)).

La OMS estima que: *“cada año podría evitarse la muerte de más de 820 000 niños menores de cinco años a escala mundial, si todos los niños fueran amamantados correctamente. La lactancia materna promueve el desarrollo del cerebro, reduce el riesgo de obesidad infantil y protege a la mujer contra el cáncer de mama, el cáncer de ovario y la diabetes. Los establecimientos que prestan servicios de maternidad y neonatología desempeñan una función excepcional al proporcionar a las nuevas madres y a los bebés una asistencia y el estímulo oportunos y adecuados que necesitan para amamantar a sus hijos satisfactoriamente, lo que permite ahorrar fondos públicos y salvar vidas humanas”* ([WHO, UNICEF, 2018](#)).

Para hacer frente al complejo problema del abandono del amamantamiento, la OMS y UNICEF publicaron en 1989 el documento *Diez Pasos hacia una feliz lactancia natural (los Diez Pasos)* en el marco de un conjunto de políticas y procedimientos que las organizaciones que prestan servicios de maternidad y neonatología deberían aplicar para apoyar la lactancia materna. En 1991, la OMS y el UNICEF pusieron en marcha la estrategia internacional *“Baby Friendly Hospital Initiative” (BFHI)*, conocida en castellano como *Iniciativa para la Humanización de la Asistencia al Nacimiento y la Lactancia o IHAN*, a fin de alentar a las organizaciones que prestaban servicios de maternidad y neonatología de todo el

mundo a que adoptaran los Diez Pasos. Los establecimientos que documentaron y acreditaron su plena adhesión a los Diez Pasos, así como la adopción del Código Internacional de Comercialización de Sucedáneos de la Leche Materna y las resoluciones pertinentes de la Asamblea Mundial de la Salud, podían designarse como hospitales «*Baby-friendly*».

Desde entonces, la BFHI se ha demostrado como una de las intervenciones más eficaces y rentables para la mejora general de las tasas de lactancia materna (Horton et al., 1996; Pérez-Escamilla et al., 2016). Sin embargo, en la actualidad, sólo el 10% de los nacimientos en el mundo tienen lugar en un hospital BFHI (WHO, 2017, p.). Es, por tanto, necesario mejorar esta realidad para lograr los objetivos mundiales de nutrición, salud y supervivencia, crecimiento económico y sostenibilidad ambiental (Rollins et al., 2016; Vitoria et al., 2016; WHO, UNICEF, 2018).

La OMS/UNICEF también ha definido el conjunto de "Indicadores para evaluar las prácticas de alimentación del lactante y del niño pequeño" (WHO, UNICEF, 2007). Incluye la definición de indicadores para medir la iniciación, la duración y la exclusividad de la lactancia materna, que son los principales resultados de los estudios sobre lactancia materna (Wood & Woods, 2018). Este conjunto de indicadores constituye también el conjunto de indicadores de resultado recomendados para el seguimiento de la BFHI.

Tanto el establecimiento, como el posterior mantenimiento de la BFHI requieren un importante esfuerzo y compromiso por parte del hospital. La aplicación de los 10 pasos incluye tanto procedimientos de gestión como prácticas clínicas. Dentro de los procedimientos de gestión se incluye la creación de sistemas de seguimiento permanente y de gestión de datos (paso 1c). De hecho, un factor importante para el éxito de la aplicación y el mantenimiento de la BFHI es el seguimiento regular de los indicadores de proceso y de resultados (Pérez-Escamilla, 2007; Pérez-Escamilla & Chapman, 2012; Zakarija-Grković et al., 2018).

Sin embargo, hoy en día, una minoría de países realiza una monitorización regular de indicadores de alimentación infantil (Pérez-Escamilla & Moran, 2016; UNICEF, 2020). Una de las principales razones es la dificultad de obtener datos poblacionales fiables (Chapman & Pérez-Escamilla, 2009; European Centre for Disease Control Prevention, 2014; Nickel et al., 2017; Paul et al., 2019; Pérez-Escamilla & Chapman, 2012).

### **1.2.2 La calidad de datos biomédicos**

La obtención de datos fiables es un aspecto esencial para cualquier investigación, estudio o monitorización de la práctica clínica. Los datos biomédicos provienen principalmente de registros realizados por profesionales sanitarios o por dispositivos de electromedicina, dentro de entorno altamente cambiante de la salud. Es de vital importancia que estos datos muestren una visión precisa de la situación clínica real de los pacientes. Sin embargo, dichos datos pueden estar sujetos a errores, omisiones y cambios en la aplicación de protocolos sanitarios, por lo que su calidad no puede ser obviada. No evaluar la DQ puede suponer un desequilibrio entre la realidad del paciente y el conocimiento que se obtiene del mismo a partir de los datos, lo que puede llevar a tomar decisiones erróneas y realizar procesos subóptimos. En definitiva, un obstáculo insalvable para la explotación eficaz de la información.

Respecto a la fuente para la obtención de dichos datos, frente a las costosas encuestas ad-hoc, lo que actualmente se conoce como "datos del mundo real" o en inglés "*real world data*" y, *en concreto*, los registros de HCE tienen el potencial de constituirse en una fuente de datos idónea para la investigación, la calidad de la atención sanitaria o la monitorización y vigilancia de la salud, incluyendo la salud infantil, debido a su cobertura poblacional y su continuidad en el tiempo (Ajetunmobi et al., 2014; American Academy of Pediatrics et al., 2008; Euro-Peristat project with SCPE and Eurocat, 2010; Halvorsen et al., 2015; Nickel

et al., 2017; Paul et al., 2019; Weng et al., 2012). Sin embargo, el establecimiento de la HCE como fuente de conocimiento en el ámbito clínico plantea importantes retos a las organizaciones sanitarias. La variedad de problemas de DQ que aparecen al habilitar este re-uso de la HCE están relacionados con dos causas principales (Cruz-Correia et al., 2010): 1) La HCE original está diseñada para su uso primario en la atención sanitaria, sin tener en cuenta el futuro re-uso de la información clínica, que puede requerir diferentes grados de calidad, y 2) los registros de HCE no están diseñados pensando en la prevención de problemas de DQ. Por tanto, es imprescindible evaluar la DQ para garantizar la validez y fiabilidad del re-uso secundario de los datos (Khare et al., 2017; Reimer et al., 2016; Verheij et al., 2018; Weiskopf & Weng, 2013; Weng et al., 2012).

Para habilitar un control sistemático de la DQ es necesario disponer de marcos conceptuales fiables que establezcan la definición y medición de los aspectos relevantes de la DQ, que habitualmente denominados dimensiones de la DQ (Wang, 1996). Además, se necesitan herramientas informáticas para el control de la DQ en la asistencia sanitaria y la investigación en salud (Weiskopf & Weng, 2013).

La DQ suele definirse en términos de idoneidad para cumplir los requerimientos del usuario (Karr et al., 2006; Madnick et al., 2009) o en términos de objetivos de DQ (Jeusfeld et al., 1998). Con tal propósito, muchos estudios definieron algunas dimensiones de la DQ (Cruz-Correia et al., 2010; Karr et al., 2006; Sáez et al., 2012; Wang, 1996; Weiskopf & Weng, 2013) que debían ser medidas: atributos que representan un aspecto o constructo único de la DQ (Wang, 1996). Las dimensiones pueden ajustarse a las especificaciones de los datos o a las expectativas de los usuarios (Lee et al., 2002).

Weiskopf & Weng (2013) realizaron una revisión sistemática de las dimensiones de la DQ y de los métodos de evaluación de la DQ de la HCE (Weiskopf & Weng, 2013). A partir de 27 términos únicos que describen dimensiones en la literatura, derivaron empíricamente cinco dimensiones de alto nivel: completitud, corrección, concordancia, verosimilitud y vigencia.

Dado el carácter de “*fitness for use*” o ajuste al propósito intrínseco a la gestión de calidad, a lo largo de todo el desarrollo de la tesis, utilizaremos el marco de DQ publicado por la UPV (Sáez et al., 2012) basado en 9 dimensiones para medir la DQ: valor predictivo, corrección, unicidad, consistencia, completitud, contextualización, estabilidad temporal, estabilidad espacial y confiabilidad. Este marco de DQ ayuda a definir los requisitos del proyecto de DQ y proporciona un conjunto completo de dimensiones de DQ en concordancia con la literatura.

### **1.2.3 Estándares de información sanitaria para dar soporte a la continuidad asistencial**

La atención sanitaria de un proceso asistencial suele involucrar a distintos niveles asistenciales (atención primaria y atención especializada) y, en ocasiones, a distintas organizaciones sanitarias.

En el caso específico de la atención a la alimentación infantil hasta los dos años, comienza en el hospital (atención especializada) con el parto-nacimiento. Tras el alta hospitalaria, la atención continúa con revisiones de salud en atención primaria.

La continuidad asistencial es una visión continua y compartida del trabajo asistencial en el que intervienen múltiples profesionales en centros de trabajo diferentes, que actúan en tiempos distintos, con el objetivo común de mejorar la salud del ciudadano. Constituye, por tanto, un elemento clave para garantizar la calidad asistencial.

Desde el punto de vista de la compartición o transferencia de información, la continuidad asistencial se puede entender como un principio organizativo que requiere que los provee-

dores de salud sean capaces de intercambiar la información sobre un mismo sujeto (Instituto de Salud Carlos III, 2013). En este sentido, distintos estándares de información sanitaria tienen como objetivo dar soporte a la misma.

En concreto, la norma UNE-EN ISO 13 940 (European Committee for Standardization Health Informatics, 2007), también denominada ContSys, establece un sistema de conceptos y relaciones con significado unívoco que sustente la continuidad asistencial. El sistema de conceptos intenta ser el nivel más abstracto para asegurar el intercambio de información entre organizaciones sanitarias, denominado interoperabilidad organizativa. Habilita que las distintas organizaciones puedan describir sus procesos de salud de forma universalmente comprensible, estableciendo así un marco común para realizar un intercambio seguro y completo de la información, tanto en su contenido como en su contexto. El sistema de conceptos definido en la norma UNE-EN ISO 13 940 es un primer paso necesario para la interoperabilidad.

Por otro lado, la norma UNE-EN ISO 13 606 (ISO, 2008, p. 13606) tiene como objetivo la transferencia de historia clínica de pacientes o de partes de ella, ofreciendo una arquitectura o representación genérica para la comunicación de registros de historias clínicas entre distintos sistemas, conservando su significado original, el contexto en que fue creada y la confidencialidad, con capacidad de adaptación a lo largo del tiempo. Para ello, utiliza el llamado “modelo dual”, mediante el cual se separa la parte que no varía con el tiempo, la información, representada a través del “modelo de referencia”, de la parte que sí evoluciona, el conocimiento, representada mediante el “modelo de arquetipos”. El modelo de referencia define el conjunto de clases que forman los bloques constituyentes del registro. Las instancias del modelo de referencia son los datos. Por otro lado, un arquetipo es la definición de una combinación jerarquizada de componentes del modelo de referencia para modelar conceptos clínicos, para ello, impone restricciones al modelo de referencia (estableciendo nombres, tipos de datos posibles, valores por defecto, cardinalidad, etc.). La representación formal de los arquetipos permite su procesamiento automático.

#### **1.2.4 La experiencia “Hasta Que Tú Quieras” del Área de Salud V – Altiplano del Servicio Murciano de Salud**

El Área de Salud V – Altiplano de la Región de Murcia, es un área de salud perteneciente al Servicio Murciano de Salud, tiene una superficie de 1580 km<sup>2</sup> y abarca una población de 60 828 habitantes (INE 2020), representando el 4% de la población total de la región. Dispone de tres centros de salud (dos en la localidad de Yecla y uno en Jumilla), tres consultorios periféricos, dos servicios de urgencia extra-hospitalaria y un hospital comarcal, Hospital Virgen del Castillo, en la localidad de Yecla, con una capacidad de 98 camas y una unidad neonatal de segundo nivel.

La experiencia “Hasta Que Tú Quieras” (HQTQ) desarrollada en el Área de Salud V – Altiplano del Servicio Murciano de Salud es una estrategia para implementar las recomendaciones de la BFHI en toda un área de salud, con objeto de mejorar la calidad de la atención materno-infantil durante los primeros 1000 días de vida. Está basada en la información, válida y fiable, para la monitorización continua, realización de feedback personalizado a los profesionales sanitarios participantes y formación y capacitación orientada al proceso. Se diseñó como una estrategia “*bottom-up*” y fue promovida desde 2009 por un grupo de mejora multidisciplinar líder en el Área de Salud, formado por expertos clínicos y de tecnologías de la información.

La experiencia HQTQ, se enmarca en el ámbito de las buenas prácticas (BBPP) basadas en la evidencia, que son aquellas intervenciones o experiencias en el marco de líneas estratégicas de salud basadas en el mejor conocimiento científico disponible y que han de-

mostrado ser efectivas, que puedan ser transferibles y representen un elemento innovador para el sistema sanitario.

Los antecedentes de la experiencia HQTQ se remontan al año 2000; cuando los profesionales del Área de Salud V se plantearon una cuestión ética fundamental: *¿Cuál es nuestra responsabilidad en el fenómeno social del abandono prematuro del amamantamiento como forma habitual de alimentación del lactante?* Tras conocer la BFHI, se inició el proceso de adaptación e implantación en el Hospital Virgen del Castillo, con un enfoque centrado en el cumplimiento de los estándares de calidad propuestos por la BFHI. Se trató de una apuesta firme por proteger la primera etapa de la vida que transformó el servicio de maternidad y neonatología del hospital, obteniendo importantes resultados en salud. El hospital fue acreditado por Unicef como hospital BFHI en 2004.

Una vez superada la acreditación BFHI del hospital, el proyecto HQTQ ([García-De-León-González et al., 2005](#)) surgió para dar respuesta a una cuestión que estuvo latente durante todo el proceso de implantación de la BFHI *¿cómo afrontar los problemas y dificultades de las madres con su lactancia?* El nombre “Hasta Que Tú Quieras” (HQTQ) surgió espontáneamente como respuesta a la incertidumbre que referían muchas madres al preguntarles *¿Cuánto tiempo querían amamantar a sus bebés?* y donde una respuesta habitual era “hasta que pueda”. El respeto y el compromiso para adecuar las prácticas a sus necesidades y expectativas supusieron un nuevo horizonte de mejora más allá del cumplimiento de estándares de calidad de la BFHI. El marco de la calidad total, basado en unos principios que enfatizan la importancia de orientar la organización a las necesidades de los clientes (madres, bebés y familia), la gestión por procesos y su orientación hacia los resultados, respondía a este nuevo reto. Bajo estas premisas, se diseñó y desplegó el Proceso de Atención a la Alimentación del Lactante (PAAL) ([García-De-León-González et al., 2008](#)) siguiendo la metodología de gestión por procesos, considerada como la herramienta más eficaz para desplegar los principios de la calidad total. La misión del PAAL era disminuir la variabilidad injustificada y adecuar las prácticas y recomendaciones a las necesidades de la madre para facilitar que pudiese amamantar a su bebé todo el tiempo que decidiese. Se trataba de un proceso de continuidad asistencial, cuyo ámbito se extendía desde el hospital hasta Atención Primaria, pero que además incluía un importante componente comunitario.

La concepción y diseño del PAAL coinciden con lo que actualmente se denominan Procesos Asistenciales Integrados (PAI), pues tienen como objetivo aumentar la efectividad de las actuaciones clínicas a través de una mayor coordinación y garantía de continuidad asistencial, incorporando la visión del paciente en el conjunto del equipo asistencial y desarrollando la práctica clínica en base a la mejor evidencia científica disponible, proporcionando una guía que permita a los profesionales sanitarios disminuir la variabilidad en la práctica clínica.

Para la implementación del PAAL, se diseñó una intervención de calidad asistencial con dos ejes vertebrales: 1) la ética deontológica y 2) la generación de información de calidad a partir de datos de Historia Clínica Electrónica (HCE) para una monitorización continua del proceso asistencial (PAAL), realización de feedback personalizado a los profesionales y capacitación orientada al proceso. En este contexto, los primeros análisis de datos procedentes de HCE en 2009 constataron problemas de calidad de datos (DQ) y los sucesivos cambios en los modelos de información de la HCE infantil ampliaron esta problemática.

El análisis, todavía manual y exploratorio, de los datos extraídos de la HCE identificó distintos tipos de problemas de DQ a pesar de no seguir una metodología específica para su evaluación. El gran esfuerzo invertido en el análisis manual y la amplitud del problema motivaron la búsqueda de metodologías que nos permitiesen abordar la DQ, punto en el que nos encontramos con el grupo IBIME, actualmente denominado Biomedical Data Science Laboratory (BDSLab) de la UPV y su marco de DQ basado en 9 dimensiones. La confluencia de intereses en este ámbito facilitó el inicio de trabajos y proyectos comunes en el ámbito de la DQ de la alimentación infantil.

Bajo estas premisas, para habilitar la generación de información de calidad para la monitorización continua del PAAL, se identificó la necesidad de diseñar e implementar un proceso de construcción de repositorios con DQ asegurada, a partir de datos multi-fuente de los distintos sistemas de HCE donde se registra información relacionada con la alimentación infantil. A partir del cual se pudiese construir un repositorio de datos integrados procedentes de la recolección rutinaria de HCE, que tras el proceso de aseguramiento de DQ, permitiese la estimación robusta de indicadores internacionales de alimentación infantil, para su monitorización continua.

### 1.3 Preguntas de investigación y objetivos

La obtención de información válida y fiable para la investigación y monitorización de la alimentación infantil depende en gran medida de la calidad de datos. La variedad de fuentes de HCE, la problemática de calidad de datos inherente al re-uso de datos de HCE, dentro del entorno siempre cambiante de la salud y la falta de metodologías específicas para construir repositorios, orientadas a asegurar su DQ, son problemas con los que se encuentran las organizaciones sanitarias cuando pretender habilitar el re-uso de HCE. A esta problemática se une la necesidad de asegurar la correcta compartición o transferencia de información en entornos con distintos proveedores sanitarios, para lo que se debe valorar el uso de estándares de información sanitaria, como UNE-EN ISO13 940 y UNE-EN ISO13 606. Esta tesis se ha concebido con el propósito de ayudar a dar respuesta a estos problemas con metodologías, herramientas y casos de uso, con el fin último de habilitar una monitorización poblacional y continua de indicadores de alimentación infantil a partir de datos de HCE. En consecuencia, nos planteamos las siguientes preguntas:

- PR1 ¿Es posible confeccionar un diseño y metodología común para el proceso de construcción de repositorios para investigación y monitorización de la asistencia a partir de datos de HCE, indistintamente del ámbito clínico y la fuente de datos?, ¿Qué elementos deben guiar el proceso de construcción y como debemos organizarlo?
- PR2 ¿Qué calidad tienen los registros de HCE de alimentación infantil? ¿Es posible construir repositorios de investigación a partir de datos multi-fuente de HCE de alimentación infantil, que tengan un nivel de calidad de datos adecuado para su re-uso en investigación y monitorización de la asistencia sanitaria?
- PR3 ¿Cómo podemos integrar repositorios de distintos centros sanitarios para obtener un único repositorio de datos integrados?, ¿Qué estándares debemos emplear para su integración?
- PR4 ¿Es posible asegurar una estimación robusta de indicadores de alimentación infantil a partir de datos procedentes de recolección rutinaria de HCE?
- PR5 ¿Qué impacto tienen estos repositorios en las organizaciones?

El trabajo de investigación llevado a cabo en esta tesis pretende responder a estas preguntas, al tiempo que define y construye los procesos y repositorios necesarios para esta investigación y consecución final de una monitorización poblacional y continua de la alimentación infantil a partir de datos de HCE. Para este fin, se definieron los siguientes objetivos de la tesis:

- O1 - Diseñar un proceso de construcción de repositorios poblacionales con calidad asegurada, a partir de datos multi-fuente de HCE de alimentación infantil desde el

nacimiento hasta los dos años, para su re-uso en investigación y monitorización de la asistencia clínica.

- O2 – Aplicar el proceso de construcción diseñado en O1 para construir un repositorio para investigación y monitorización del proceso de alimentación infantil en su etapa perinatal (atención hospitalaria).
- O3 - Ampliar el repositorio perinatal (ámbito hospitalario) construido en O2 con la información de seguimiento de la alimentación hasta los dos años (atención primaria) para la obtención de un IDR para la investigación y monitorización del proceso completo de atención a la alimentación del lactante.
- O4 - Integrar repositorios materno-infantiles de distintos hospitales en base a modelos de información clínica UNE-EN ISO 13 606 y uso concurrente de UNE-EN ISO13 940 - Consys para construir un único repositorio integrado e interoperable donde se asegure la calidad de sus datos.
- O5 – Construcción de indicadores de alimentación infantil a partir de los elementos de los modelos de información clínica estandarizados (O4), con lo que poder evaluar buenas prácticas y el grado de implantación de las estrategias internacionales de alimentación infantil.
- O6 – Diseñar e implementar una estimación robusta de indicadores de alimentación infantil a partir de los datos longitudinales y transversales de alimentación presentes en el repositorio de datos integrados (IDR).

Los mencionados objetivos confluyen en el objetivo global de la tesis: implementar un sistema de monitorización poblacional de alimentación infantil con calidad de datos asegurada a partir de *real world data*, procedente de recolección rutinaria de HCE.

## 1.4 Contribuciones

### 1.4.1 Principales contribuciones

Las principales contribuciones de esta tesis son las siguientes:

- **C1 – Proceso de construcción de repositorios con calidad de datos asegurada**  
Proceso computacional para obtener conjuntos de datos de alimentación infantil con calidad asegurada, para su re-uso en investigación y monitorización de la asistencia sanitaria dentro del ámbito de la alimentación infantil.
- **C2 –Repositorio perinatal con calidad de datos asegurada del Área de Salud V**  
Repositorio con calidad de datos asegurada a partir de los datos de la HCE perinatal del Hospital Virgen del Castillo, construido utilizando el proceso de construcción definido en C1.
- **C3 - IDR materno-infantil de los hospitales Virgen del Castillo de Yecla (Murcia) y 12 de Octubre (Madrid), basado en estándares de información sanitaria**  
Propuesta de un IDR que integra datos de gestación, nacimiento y alimentación del lactante de dos áreas de salud (Murcia y Madrid) utilizando el estándar UNE-EN ISO 13 606 y uso concurrente de UNE-EN ISO 13 940, con evaluación de calidad de datos y una propuesta de indicadores basada en los elementos del IDR.
- **C4 – Propuesta de arquetipos UNE-EN ISO 13 606 al Ministerio de Sanidad, Servicios Sociales e Igualdad (MSSSI) para integración a nivel nacional**

Dos arquetipos principales para la gestación-nacimiento y alimentación del lactante, respectivamente, basados en la norma UNE-EN ISO 13 606 en sintaxis ADL 1.4, compuestos por 39 subarquetipos, presentados al MSSSI.

- **C5 – Repositorio de datos integrados (IDR) de alimentación infantil hasta los dos años, con calidad de datos asegurada, del Área de Salud V**  
IDR de alimentación infantil que integra los datos perinatales con los datos de seguimiento de alimentación hasta los dos años provenientes de la HCE de atención primaria, construido como una ampliación de repositorio C2 y utilizando de nuevo el proceso de construcción C1.
- **C6. – Software para la automatización del proceso de construcción de repositorios con calidad asegurada**  
Aplicación web para la paquetización de todas las herramientas e interfaces necesarias de las fases del proceso de construcción de repositorios C1.
- **C7 – Estimación robusta de indicadores de alimentación infantil y sistema de monitorización, poblacional y continuo con calidad de datos asegurada**  
Sistema de monitorización realizado en base a la estimación robusta de indicadores de alimentación infantil a partir del registro longitudinal de alimentación que conforma el repositorio C5.
- **C8 – Sello de Buena Práctica del Sistema Nacional de Salud (SNS) a la estrategia HQTQ**  
La estrategia HQTQ del Área V de Salud del Servicio Murciano de Salud obtuvo el sello de reconocimiento de Buenas Prácticas del SNS en la convocatoria 2014 en el área de atención al parto y salud reproductiva, debido, entre otros factores, a su novedoso enfoque basado en la información obtenida de los repositorios C2 y C5.
- **C9 – Scaling – up del modelo de información perinatal al resto de hospitales de Servicio Murciano de Salud**  
Contribución extraordinaria de la tesis. A partir de los resultados de la experiencia HQTQ y el Sello de Buena Práctica (C5), el Servicio Murciano de Salud decidió extender la experiencia HQTQ al resto de hospitales murcianos, específicamente su modelo de información perinatal, para disponer de un modelo de información homogéneo para poder evaluar y explotar de forma exhaustiva la información poblacional sobre la salud de los recién nacidos.

#### 1.4.2 Publicaciones científicas

Las contribuciones de esta tesis han llevado a realizar 3 publicaciones en revistas científicas y 3 participaciones en congresos:

- Pu1. **Ricardo García-de-León-Chocano**, Carlos Sáez, Verónica Muñoz-Soler, Ricardo García-de-León-González, Juan Miguel García-Gómez. “*Construction of quality-assured infant feeding process of care data repositories: definition and design (Part 1)*”. *Comput Biol Med.* 2015 Dec 1;67:95-103
- Pu2. **Ricardo García-de-León-Chocano**, Verónica Muñoz-Soler, Carlos Sáez, Ricardo García-de-León-González, Juan Miguel García-Gómez. “*Construction of quality-assured infant feeding process of care data repositories: Construction of the perinatal repository (Part 2)*”. *Comput Biol Med.* 2016 Apr 1;71:214-222
- Pu3. **Ricardo García-de-León-Chocano**, Verónica Muñoz-Soler, Carlos Sáez, Juan Miguel García-Gómez, Jose Alberto Maldonado, David Moner, Luis Clemente, María Ángeles Gómez, Ricardo García-de-León-González. “*El valor de la informa-*

ción para desarrollar e implementar Buenas Prácticas (BBPP)". XIX Congreso Nacional de Informática de la Salud (Inforsalud 2016)

- Pu4. Carlos Sáez, David Moner, **Ricardo García-de-León-Chocano**, Verónica Muñoz-Soler, Ricardo García-de-León-González, José Alberto Maldonado, Diego Boscá, Salvador Tortajada, Montserrat Robles, Juan Miguel García-Gómez, Manuel Alcaraz, Pablo Serrano, José Luis Bernal, Jesús Rodríguez, Gerardo Bustos, and Miguel Esparza. "A Standardized and Data Quality Assessed Maternal-Child Care Integrated Data Repository for Research and Monitoring of Best Practices: A Pilot Project in Spain" (Informatics for Health 2017). Stud Health Technol Inform. 2017; 235: 539-43.
- Pu5. **Ricardo García-de-León-Chocano**, Carlos Sáez, Ricardo García-de-León-González, Juan Miguel García-Gómez. "An information-based best practice for mother and child care outcomes improvement. A successful bottom-up strategy" (i-HD Hospital Network Workshop. 2018)
- Pu6. **Ricardo García-de-León-Chocano**, Carlos Sáez, Verónica Muñoz-Soler, Antonio Oliver-Roig, Ricardo García-de-León-González, Juan Miguel García-Gómez. "Robust estimation of infant feeding indicators by data quality assessment of longitudinal electronic health records from birth up to 18 months of life". Comput Methods Programs Biomed, 207:106147.

## 1.5 Proyectos de investigación

Buena parte del trabajo desarrollado en esta tesis ha sido posible gracias a la participación en proyectos de investigación y otros proyectos de mejora de la calidad asistencial relacionados con la alimentación infantil, donde la generación de información de calidad, ha sido uno de los ejes vertebrales de dichos proyectos.

**HQTQ-1.** "Hasta Que Tú Quieras (HQTQ)". Ref: PIEMCA05-13

**Objetivos:** Conocer la situación de la lactancia materna en el Área de Salud V - Altiplano, estudiar las necesidades y expectativas de las madres y profesionales, desarrollar un registro de datos prospectivo de alimentación infantil, de ámbito poblacional, para la continuidad y mejora de los cuidados.

**Organización financiadora:** Fundación para la Formación e Investigación Sanitarias de la Región de Murcia (FFIS). Programa EMCA

**Participantes:** Área de Salud V – Altiplano y Facultad de Ciencias de la Salud de la Universidad de Alicante.

**URL:** [https://www.murciasalud.es/recursos/ficheros/302353-PIENCA05\\_13.pdf](https://www.murciasalud.es/recursos/ficheros/302353-PIENCA05_13.pdf)

**HQTQ-2.** "Diseño de la atención integral al proceso de la alimentación del lactante (HQTQ2)". Ref: PIEMCA08/13

**Objetivos:** Formalizar la experiencia y aprendizajes alcanzados durante HQTQ-1 en un modelo teórico que incluya todos los elementos y relaciones que conforman el proceso de atención a la alimentación infantil. Adaptación de los sistemas de información del Área de Salud V – Altiplano para responder a las necesidades de información del proceso. Generación de información de calidad a partir de Historia Clínica Electrónica (HCE) para una monitorización continua del proceso asistencial.

**Organización financiadora:** Fundación para la Formación e Investigación Sanitarias de la Región de Murcia (FFIS). Programa EMCA.

**Participantes:** Área de Salud V – Altiplano y Facultad de Ciencias de la Salud de la Universidad de Alicante.

**URL:** [https://www.murciasalud.es/recursos/ficheros/302355-Informe\\_Final\\_Marzo.pdf](https://www.murciasalud.es/recursos/ficheros/302355-Informe_Final_Marzo.pdf)

**P. Piloto MSSSI.** *“Proyecto piloto para mejorar la calidad de la información en la atención perinatal, basada en buenas prácticas del SNS, en el marco de la estrategia de atención al parto y salud reproductiva”*. Expediente nº 2015/07PN0010

**Objetivos:** Obtención de una fuente de datos estandarizada y con calidad asegurada sobre la que construir una propuesta de indicadores para la monitorización de la atención materno-infantil y dar respuesta a la necesidad de evaluación de las estrategias de salud materno-infantiles, con el fin último de difundir sus buenas prácticas.

**Organización financiadora:** Ministerio de Sanidad, Servicios Sociales e Igualdad (MSSSI). Subdirección General de Calidad y Cohesión.

**Colaboradores:** Hospital Virgen del Castillo, Hospital 12 de Octubre de Madrid y la empresa Veratech for Health, spin-off de la UPV.

## 1.6 Esquema de la tesis

La estructura de esta tesis refleja las diferentes etapas de investigación llevadas a cabo para su desarrollo. El capítulo 1 presenta la motivación, justificación y fundamentos de la tesis, así como sus objetivos, contribuciones y proyectos. El capítulo 2 presenta el proceso de construcción de repositorios con calidad asegurada. El capítulo 3 presenta la utilización del proceso de construcción de repositorio para la construcción del primer repositorio con calidad asegurada, a partir de datos de HCE del ámbito hospitalario perinatal, con sus resultados de DQ. El capítulo 4 presenta la experiencia de integración de repositorios materno-infantiles de dos hospitales: Hospital Virgen del Castillo de Yecla y Hospital 12 de Octubre de Madrid, basada en el uso de estándares de información sanitaria, así como la construcción de indicadores a partir de elementos de los modelos clínicos estandarizados. El capítulo 5 presenta la construcción del IDR de alimentación infantil del Área de Salud V – Altiplano, aplicando el proceso de construcción de repositorios para ampliar el repositorio perinatal con los datos del seguimiento de la alimentación hasta los dos años procedentes de la HCE de atención primaria. El capítulo 6 presenta la estimación robusta de indicadores de alimentación mediante el aseguramiento de DQ del registro longitudinal de los registros de HCE desde el nacimiento hasta los 18 meses, para habilitar, finalmente, una monitorización poblacional y continua de indicadores de alimentación infantil. El capítulo 7 describe el valor del repositorio en el Área de Salud V – Altiplano. Por último, el capítulo 8 describe las conclusiones finales y trabajos futuros.

La Figura 1.1 ilustra el esquema de las contribuciones de la tesis, mostrando las relaciones entre sus capítulos, contribuciones, publicaciones y proyectos.

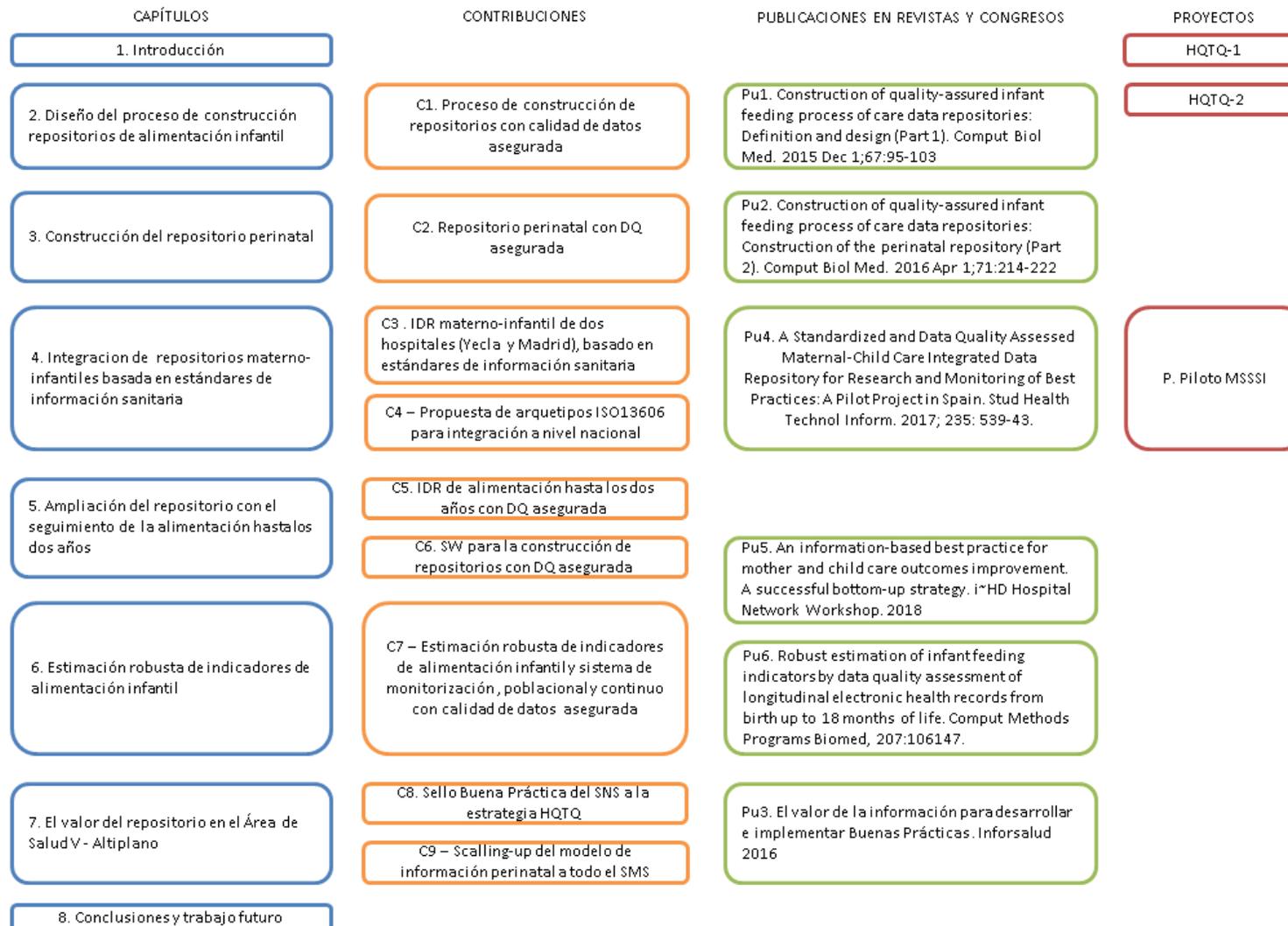


Figura 1.1 Esquema de la tesis



## 2 Diseño del proceso de construcción de repositorios de alimentación infantil

En este capítulo, se presenta el diseño del proceso de construcción de repositorios de alimentación infantil, desde el nacimiento hasta los dos años, a partir de datos poblacionales extraídos de múltiples fuentes de HCE. Se propone un proceso computacional para obtener conjuntos de datos de alimentación infantil con calidad asegurada, para su re-uso en el seguimiento y monitorización de la asistencia sanitaria, dentro del ámbito del establecimiento y monitorización de estrategias internacionales como la BFHI, así como del estudio de cuestiones relevantes sobre la alimentación infantil.

El proceso de construcción consta de trece pasos que garantizan la armonización, estandarización, completitud, de-duplicación y consistencia del conjunto de datos. Se controla la calidad de datos (DQ) de entrada y salida de cada uno de estos pasos de acuerdo con ocho dimensiones de DQ: valor predictivo, corrección, duplicación, consistencia, completitud, contextualización, estabilidad temporal y estabilidad espacial. La secuencia de pasos del proceso está orientada a: 1) la obtención de una estructura de datos canónica, donde no solo se asegura la integridad estructural de los datos, sino que se estandarizan los conceptos según la norma UNE-EN ISO 13 940 (Consys); y 2) el aseguramiento de la integridad semántica de los datos.

A partir del proceso computacional que se define en este capítulo, en capítulos subsiguientes se describirá la obtención de repositorios con calidad asegurada para la monitorización de la práctica clínica, tanto en el ámbito de la atención especializada perinatal (proceso del parto-nacimiento del bebé realizado en el hospital) como su ampliación al ámbito de la atención primaria para cubrir todo el proceso de seguimiento de la alimentación infantil hasta los dos años.

*El contenido de este capítulo es una adaptación para la tesis del artículo publicado (García-de-León-Chocano et al., 2015) – contribución Pu1. Como resultado, quedó definido e implantado en la Gerencia del Área de Salud V – Altiplano el “Proceso de construcción de repositorios con calidad de datos asegurada”, contribución C1. La publicación del artículo Pu1 y la implantación de la contribución C1 fueron posibles gracias a los trabajos previos desarrollados en los proyectos de investigación HQTQ-1 y HQTQ-2.*

### 2.1 Introducción

La Organización Mundial de la Salud define la lactancia materna como la forma más natural y saludable de alimentar a lactantes y niños pequeños. Se recomienda lactancia materna exclusiva durante los seis primeros meses de vida y con una alimentación complementaria adecuada hasta los dos años o más (World Health Organization, 2003). Actualmente, la lactancia materna se considera uno de los principales determinantes de la salud materno-infantil. Está relacionada con la prevención de factores de riesgo de las enfermedades crónicas de la infancia y de la carga de enfermedad evitable (Verschuuren et al., 2012). Sin embargo, los datos europeos muestran que menos del 20% de los bebés son alimentados siguiendo estas recomendaciones (Cattaneo et al., 2010).

Dadas las importantes implicaciones para la salud de estas bajas tasas, la estrategia europea para la protección, promoción y apoyo de la lactancia materna (E. U. [Project on Promotion of Breastfeeding in Europe, 2004](#)) propone, entre otras acciones, el despliegue generalizado de la BFHI, mediante acciones globales y coordinadas, junto con el desarrollo de mecanismos de control que faciliten su implantación. Sin embargo, la ausencia de estos mecanismos de control es la principal causa de la brecha existente entre la evidencia y la práctica clínica diaria ([Grol & Grimshaw, 2003](#)).

En el Hospital "Virgen del Castillo", del Área de Salud V – Altiplano, las directrices internacionales descritas anteriormente se materializaron a través de una intervención de mejora de la calidad de la atención materno-infantil, que permitió su acreditación como hospital BFHI y produjo mejoras significativas en las tasas de lactancia materna ([García-de-León-González et al., 2011](#)). Sin embargo, la sostenibilidad de este tipo de intervenciones depende de la adaptación continua al contexto local y de su inclusión sistemática en los flujos de trabajo rutinarios ([Scott et al., 2011](#)). Por ello, el grupo de mejora diseñó el proceso de atención a la alimentación infantil (PAAL), ampliando el alcance a la atención primaria para adecuar la práctica clínica a las recomendaciones de la evidencia, desde el nacimiento hasta los dos años de edad. El diseño del PAAL ([García-De-León-González et al., 2008](#)) se basa en los principios de lo que actualmente se denominan Procesos Asistenciales Integrados (PAI) y estuvo enmarcado dentro de la experiencia HQTQ, "Hasta Que Tú Quieras", desarrollada a partir de los proyectos de investigación HQTQ-1 ([García-De-León-González et al., 2005](#)) y HQTQ-2 ([García-De-León-González et al., 2008](#)), para implementar las recomendaciones de la BFHI en toda el área de salud, con objeto de mejorar la calidad de la atención materno-infantil durante los primeros 1000 días de vida.

Para la monitorización del proceso asistencial y la activación de estudios de investigación sobre el PAAL, se decidió utilizar datos poblacionales multi-fuente de HCE, mediante la construcción de un repositorio de datos con calidad asegurada. Así, el objetivo del presente trabajo es diseñar el proceso de construcción de repositorios sobre el nacimiento y la alimentación infantil a partir de los datos poblacionales de HCE. El proceso de construcción se ha implementado mediante trece procedimientos de calidad de datos para evaluar, recuperar o filtrar datos clínicos que permitan su re-uso en el seguimiento y monitorización del PAAL, así como del estudio de cuestiones relevantes sobre la alimentación infantil.

## 2.2 Antecedentes

El actual despliegue de los sistemas de información sanitaria e HCE supone una fuente de datos valiosa para la monitorización sistemática de la práctica clínica, tanto para la gestión como para la investigación. Sin embargo, a diferencia de lo que ocurre en otros ámbitos sanitarios, hasta donde sabemos no existe ninguna experiencia en el uso de los datos de la HCE sobre alimentación infantil para la implementación y control de la BFHI, a pesar de que la evidencia muestra una relación entre la conveniencia de utilizar datos poblacionales para el seguimiento de prácticas hospitalarias y la viabilidad del desarrollo y fortalecimiento de la BFHI, un tipo de estudio que sólo se realiza mediante complejas y aisladas recopilaciones de datos poblacionales ([Pérez-Escamilla, 2007](#); [Pérez-Escamilla & Chapman, 2012](#)).

Además, importantes foros del ámbito de las ciencias de la salud avalan la pertinencia de basar la investigación clínica y las intervenciones de mejora de la calidad asistencial en estudios poblacionales ([American Academy of Pediatrics et al., 2008](#); [Klazinga et al., 2011](#); [World Health Organization, 2002](#)). En este sentido, existen experiencias que demuestran el potencial de la HCE como fuente de estudios poblacionales cuando los datos son sometidos a procesos adecuados para asegurar la calidad de los registros ([Yih et al., 2011](#)).

Sin embargo, el establecimiento de la HCE como fuente de conocimiento en el ámbito clínico plantea importantes retos a las organizaciones sanitarias. Es de gran importancia que

los datos de la HCE muestren una visión precisa de la situación clínica real de los pacientes. En este sentido, la DQ está directamente relacionada con la calidad asistencial (Liaw et al., 2013). Además, existen posibles errores producidos en el proceso de registro por parte de los usuarios del sistema de HCE (Burnum, 1989; Weng et al., 2012), así como errores sistemáticos asociados a la estructura del sistema de información (Weng et al., 2012).

La falta de mecanismos de control sistemático de DQ de la HCE supone un serio obstáculo para asegurar la recogida de información válida y fiable para la reutilización de datos (Weiskopf & Weng, 2013; Weng et al., 2012). Son múltiples los factores que determinan la falta de control del DQ. Entre ellos destaca la necesidad de disponer de marcos conceptuales fiables que establezcan la definición y medición de los aspectos relevantes de la DQ, que habitualmente denominados dimensiones de la DQ (Wang, 1996). Además, se necesitan herramientas informáticas para el control de la DQ en la asistencia sanitaria y la investigación en salud (Weiskopf & Weng, 2013).

Existe un acuerdo general para definir la DQ en términos de idoneidad para cumplir los requerimientos del usuario (Karr et al., 2006; Madnick et al., 2009) o en términos de objetivos de DQ (Jeusfeld et al., 1998). Con tal propósito, muchos estudios definieron algunas dimensiones de la DQ (Cruz-Correia et al., 2010; Karr et al., 2006; Sáez et al., 2012; Wang, 1996; Weiskopf & Weng, 2013) que debían ser medidas: atributos que representan un aspecto o constructo único de la DQ (Wang, 1996). Las dimensiones pueden ajustarse a las especificaciones de los datos o a las expectativas de los usuarios (Lee et al., 2002).

Recientemente, (Weiskopf & Weng, 2013) realizaron una revisión sistemática de las dimensiones de la DQ y de los métodos de evaluación de la DQ de la HCE. A partir de 27 términos únicos que describen dimensiones en la literatura, derivaron empíricamente cinco dimensiones de alto nivel: completitud, corrección, concordancia, verosimilitud y vigencia. Además de estas dimensiones, (Oliveira et al., 2005) formalizaron un conjunto de problemas de calidad de datos a lo largo de diferentes niveles de granularidad de las bases de datos organizativas, a saber, fuentes múltiples, relaciones múltiples, relaciones simples y atributos/tuplas. Por lo tanto, las dimensiones pueden asociarse a estos problemas y medirse en los diferentes niveles de granularidad.

En este estudio, utilizamos ocho de las dimensiones previamente definidas por los autores en su marco de DQ (Sáez et al., 2012) para medir la DQ de entrada y salida de cada paso de nuestro proceso de construcción de repositorios, incluyendo: valor predictivo, corrección, unicidad, consistencia, completitud, contextualización, estabilidad temporal y estabilidad espacial. Las razones para seleccionar este marco de DQ fueron: 1) ayuda a definir los requisitos del proyecto de DQ teniendo en cuenta las circunstancias en las que se adquirieron los datos, 2) proporciona un conjunto completo de dimensiones de DQ, en concordancia con las de la literatura, que sirven de base para definir qué aspectos de DQ que deben medirse y 3) constituye un caso de estudio interesante en el que se puede explorar la aplicabilidad del marco teórico.

Por otro lado, existen diferencias en la forma en que los distintos sistemas de HCE organizan los datos a nivel local, que pueden superarse utilizando estándares. Varios estándares se centran en la interoperabilidad, tanto a nivel organizativo (por ejemplo, UNE-EN ISO 13 940 - CONTsys) como a nivel semántico (por ejemplo, UNE-EN ISO 13 606 y Open EHR). La norma UNE-EN ISO 13 940 Parte1 (European Committee for Standardization Health Informatics, 2007), también conocida como CONTsys, ofrece un sistema de conceptos de referencia para la continuidad asistencial. Esta norma proporciona la terminología necesaria para planificar, aplicar y supervisar las actividades y el estado de salud de los procesos clínicos. Por otro lado, las normas de interoperabilidad semántica (como la UNE-EN ISO 13 606 y open EHR) han demostrado su utilidad para lograr la interoperabilidad semántica entre sistemas (Martínez-Costa & Schulz, 2014; Menárguez-Tortosa et al., 2012). Sin embargo, en el ámbito materno-infantil se ha encontrado poco desarrollo de modelos clínicos siguiendo estos estándares.

## 2.3 Metodología

El grupo de mejora del PAAL promovió la construcción de repositorios de datos con calidad asegurada a partir de HCE, para permitir la monitorización y la evaluación de la intervención. Para ello, se creó un grupo de expertos formado por los líderes clínicos responsables del PAAL y expertos en tecnologías de la información para definir las especificaciones y requisitos del repositorio para la monitorización del PAAL y su modelo de información. Posteriormente, este grupo diseñó la representación visual y las interfaces de usuario para configurar la construcción de los repositorios. A continuación, se definieron los procedimientos de extracción de datos de la HCE. En nuestro caso, los datos del PAAL están distribuidos en dos sistemas de HCE: los datos perinatales están alojados en el Sistema de Información Hospitalario (HIS) Selene y los datos de las revisiones de salud del Programa de Atención al Niño y al Adolescente (PANA) están alojadas en el Sistema de Información de Atención Primaria OMI-AP. Para la construcción de los repositorios de datos se adoptó un enfoque de Gestión de la Calidad de Datos Total (TDQM, del inglés *Total Data Quality Management*), donde la gestión de la calidad de datos es un proceso de mejora continua, que considera los datos como un producto, el cual sustentamos con la definición de dimensiones de DQ medibles y controlables. Este enfoque permite evitar posteriores decisiones subóptimas basadas en datos inexactos.

En esta sección, primero describimos la definición del modelo de información del PAAL utilizando conceptos de la norma CONTsys ([European Committee for Standardization Health Informatics, 2007](#)). A continuación, explicamos el marco de DQ adoptado para garantizar la calidad total del proceso. Por último, definimos el proceso de construcción de repositorios con calidad asegurada.

### 2.3.1 El modelo de información del PAAL

El contexto organizativo del modelo de información del PAAL se definió mediante los conceptos de la norma CONTsys. La norma CONTsys define un conjunto de conceptos que deben utilizarse para modelar los procesos de continuidad asistencial. Esta norma garantiza el mismo significado para los conceptos fundamentales que rigen la organización de la continuidad asistencial en entornos heterogéneos. La tabla 2.1 muestra los conceptos de la CONTsys que hemos utilizado en el PAAL para definir su contexto organizativo.

**Tabla 2.1. Conceptos del proceso de atención a la alimentación infantil (PAAL) usando la terminología CONTsys para definir su contexto organizativo**

Concepto de la CONTsys	Descripción	Correspondencia en el PAAL
Proceso asistencial	Proceso en el que un "sujeto de la asistencia" y un "actor de salud" ("organización sanitaria" o "profesional sanitario") interactúan con el objetivo de influir, directa o indirectamente, en el "estado de salud" del "sujeto de la asistencia"	El proceso de atención a la alimentación infantil (PAAL)
Organización sanitaria	Organización involucrada en la provisión del cuidado asistencial	Los centros involucrados en el PAAL pertenecientes a los dos niveles asistenciales: Hospital (Atención Especializada) y Centros de Salud (Atención Primaria)
Profesional sanitario	Profesional de la asistencia sanitaria con un derecho/titulación profesional reconocido en una determinada jurisdicción	Los profesionales que realizan la atención y registran la información
Sujeto de la asistencia	Paciente o usuario que recibe la asistencia.	El recién nacido y niño pequeño hasta los 2 años

Episodio Asistencial	Intervalo de tiempo durante el cual las actividades sanitarias se llevan a cabo para hacer frente a un tema relativo a la salud	Los dos episodios asistenciales que comprenden el PAAL: Episodio asistencial perinatal y el Programa de Atención al Niño y Adolescente (PANA), hasta los dos años de edad
Contacto	Período de actividad asistencial durante la cual el "sujeto de la asistencia" interacciona, directa o indirectamente, con uno a más profesionales de salud. Durante un contacto se pueden tratar uno o más "temas de salud"	El episodio asistencial perinatal contiene dos contactos: Partorio/Quirófano y Planta (Maternidad y/o Neonatología) El contacto paritorio/quirófano comienza con el nacimiento que puede ocurrir en el paritorio y/o en quirófano (cesáreas) y termina con el traslado a planta del recién nacido, momento en el que comienza el contacto en planta. El contacto en planta puede ocurrir en maternidad o en neonatología; termina con el alta hospitalaria del recién nacido. El episodio asistencial del PANA (hasta los dos años) contiene 8 contactos: visita a los 15 días, visita a los 2 meses, 4 meses, 6 meses, entre los 9-12 meses, 15 meses, 18 meses y 2 años
Actividad Sanitaria	Conjunto de actividades realizadas para un "sujeto de la asistencia" con la intención de mejorar o mantener, directa o indirectamente, el estado de salud de dicho sujeto de la asistencia	Para cada contacto, definimos todas las actividades sanitarias que realizan los profesionales sanitarios. Las actividades sanitarias junto con el estado de salud del recién nacido son el objeto de la monitorización del PAAL, constituyendo el conjunto de conceptos clínicos de los que se debe asegurar la calidad

Adicionalmente, utilizamos el término concepto clínico con el significado definido por la norma UNE-EN ISO 13 606 ("unidad de conocimiento, expresada mediante características pertinentes a su uso en el cuidado de la salud"). Ambas normas, CONTsys e UNE-EN ISO 13 606, se complementan entre sí y se conectan a través del concepto "extracto de HCE" (parte o la totalidad de la HCE de un sujeto de la asistencia que se pretende comunicar).

La figura 2.1 muestra el modelo de información del PAAL utilizando los conceptos de la norma CONTsys vinculados a las fuentes de datos de nuestra HCE. Cada Sujeto de la Asistencia cuenta con dos Episodios Asistenciales, definidos como los intervalos durante los cuales se realizan las actividades asistenciales, correspondientes al Episodio Asistencial Perinatal y Episodio Asistencial de las revisiones de salud del PANA hasta los dos años. Por un lado, el Episodio Asistencial Perinatal consta de dos Contactos (Contacto en Partorio/Quirófano, correspondiente al parto-nacimiento y el contacto en Planta, correspondiente al periodo en planta de hospitalización tras el nacimiento hasta el momento del alta hospitalaria). Por otro lado, el Episodio Asistencial del PANA consta de 8 contactos pertenecientes a las revisiones de las siguientes edades: antes del primer mes, a los 2 meses, a los 4 meses, a los 6 meses, entre los 9 y los 12 meses, entre los 15, a los 18 meses y a los dos años. Durante los distintos contactos, se realizan una serie de actividades sanitarias. La realización de las actividades sanitarias implica el uso de una serie de conceptos clínicos (contacto precoz piel con piel, alimentación al alta hospitalaria, somatometría, etc.), cuya información es registrada a través de un conjunto de variables. Como se puede observar en la figura, determinadas actividades sanitarias y conceptos clínicos pueden ser comunes a diferentes contactos y episodios. Del mismo modo, en un concepto clínico  $n$ , que representaremos por un conjunto de variables  $V_i, \dots, V_k$ , también podemos encontrar que algunas de éstas variables pueden formar parte de otro concepto clínico (bien misma la misma variable dentro de un sistema de información sanitario, bien variables conceptualmente equivalentes entre distintos sistemas de información).

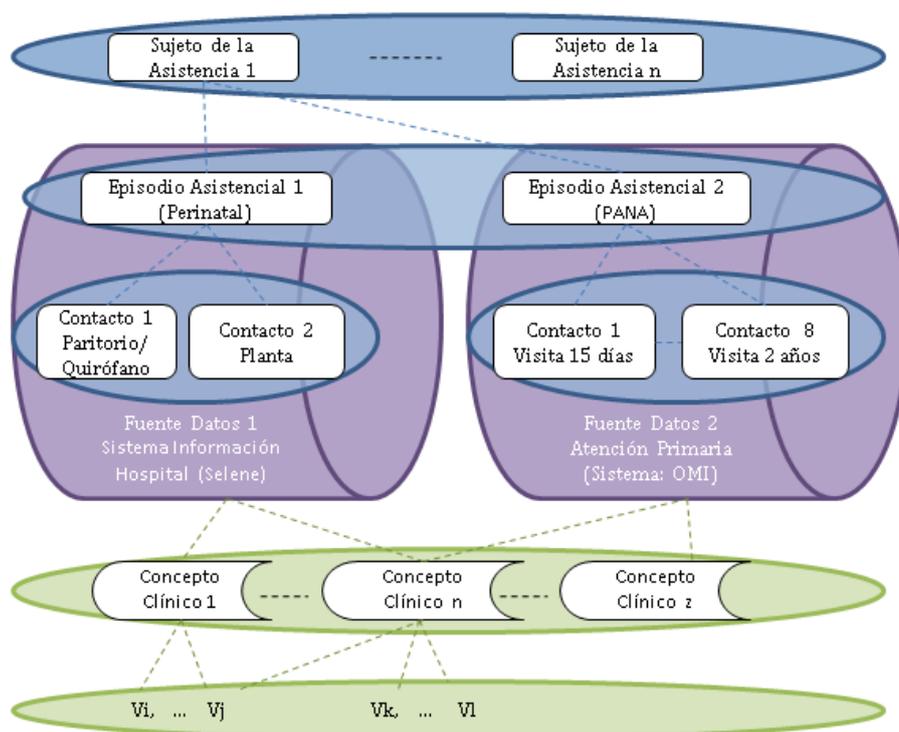


Figura 2.1. Modelo de información del proceso de atención a la alimentación infantil (PAAL) utilizando conceptos del estándar CONTsys mapeados a las fuentes de datos del hospital y atención primaria. En azul, los principales conceptos de CONTsys empleados en el modelo y su equivalente en el PAAL, donde PANA refiere al Programa de Atención al Niño y Adolescente de la Región de Murcia. En magenta, las fuentes de datos del modelo. En verde, los conceptos clínicos y sus variables.

### 2.3.2 Marco de calidad de datos

La gestión total de la DQ en la experiencia HQTQ del Área de Salud V - Altiplano guió el diseño del proceso de construcción repositorios con calidad asegurada. Dado que el proceso incluirá diferentes acciones que influyen directamente en la DQ, decidimos incluir un amplio control de calidad del proceso midiendo una métrica de calidad antes y después de cada procedimiento que afecta a los datos del repositorio. Esto permitirá al diseñador controlar el efecto del proceso sobre el repositorio al nivel más detallado.

De este modo, proponemos utilizar ocho dimensiones intrínsecas a los datos (véase la tabla 2.2) del marco de DQ definido por (Sáez et al., 2012). La definición de estas dimensiones se derivó empíricamente de la literatura (Cruz-Correia et al., 2010; Karr et al., 2006; Lee et al., 2002; Pipino et al., 2002; Shankaranarayanan & Cai, 2006; Wang, 1996; Weiskopf & Weng, 2013) y se basó en la experiencia de los autores en el análisis de datos bio-médicos.

**Tabla 2.2. Las ocho dimensiones de calidad de datos (DQ) utilizadas para nuestra evaluación de la calidad de los procedimientos de construcción del depósito.**

Dimensión	Definición
Complejidad	El grado con que se registran datos relevantes
Consistencia	El grado con que los datos satisfacen restricciones y reglas especificadas
Unicidad	El grado en que los datos registrados contienen duplicados que representan la misma entidad
Corrección	El grado de exactitud y precisión con que se representan los datos con respecto a su estado en el mundo real
Estabilidad temporal	Grado de estabilidad de los datos a lo largo de un período de tiempo
Estabilidad espacial	Grado en el que los datos son estables entre diferentes fuentes de datos
Contextualización	Grado en el que los datos se anotan correcta / óptimamente en el contexto en el que se adquirieron
Valor predictivo	Grado en el que los datos contienen información adecuada para fines específicos de toma de decisiones

Aunque las dimensiones de completitud, consistencia, corrección y unicidad se utilizan generalmente en la literatura sobre DQ, su uso se solapa con frecuencia o las dimensiones están contenidas entre sí. Las definiciones proporcionadas hacen que las dimensiones sean complementarias en el sentido de que se refieren a un aspecto específico de la calidad de datos que no está incluido en otras dimensiones. Por ejemplo, un registro de pacientes está incompleto si contiene una observación omitida; una observación es inconsistente si está fuera del rango de su variable, o es incorrecta si su valor, aun siendo coherente, es improbable. Del mismo modo, la estabilidad (temporal y espacial) representan en positivo la variabilidad o heterogeneidad, indicando hasta qué punto el contenido de los datos permanece estable respecto a cambios en los sistemas de información o práctica médica, tanto en el tiempo (estabilidad temporal) como entre distintas fuentes (estabilidad espacial), referidas a distintos sistemas de información, profesionales, etc (Sáez et al., 2016, 2020, 2021).

Además, en este estudio adaptamos el modelo de Oliveira et al (2005) (Oliveira et al., 2005) guiándonos por la semántica de los conceptos clínicos del PAAL. Así, identificamos tres niveles de granularidad en relación con los problemas de DQ (véase la Figura 2.1): i) el Hecho de estudio (HdE, del inglés *Fact of Study*), es el nivel global de análisis que puede corresponder al "Sujeto de la Asistencia", un "Episodio asistencial" o un "Contacto"; ii) los conceptos clínicos del PAAL (como "Contacto precoz piel con piel", "alimentación del lactante al alta", etc.); y iii) los atributos individuales de los conceptos clínicos. Esta granularidad orienta el diseño del proceso de construcción del repositorio para garantizar la DQ.

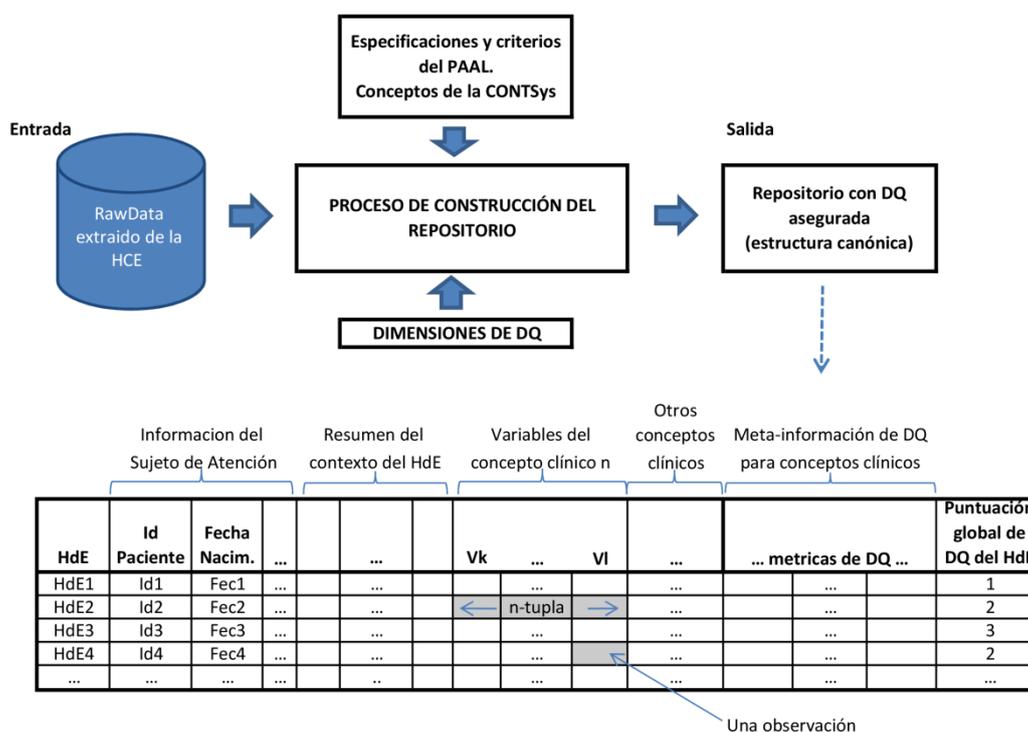
**Tabla 2.3. Ejes medidos en nuestro estudio para las diferentes dimensiones de la calidad de los datos. El conjunto de ejes seleccionado cubre la diferente granularidad de la información incluida en los repositorios.**

Axis	Definition
Hecho de estudio (HdE)	El registro de un caso compuesto por elementos según las necesidades de análisis para su reutilización. Para el PAAL, se puede asignar un HdE al Sujeto de la asistencia, Episodio asistencial o Contacto
N-tupla	Observaciones de N variables de un concepto clínico
Variable	Un atributo de un concepto clínico
Observación	El valor que toma una variable para un sujeto de la asistencia
Formulario	Una plantilla de formulario del sistema de HCE original

Por último, identificamos los ejes de la DQ incluidos en la Tabla 2.3 sobre los que medir las dimensiones de DQ. Así pues, en primer lugar, decidimos medir la DQ a nivel global utilizando el eje “HdE”. A continuación, la DQ se mide también en niveles inferiores: “N-tupla” correspondiente a un concepto clínico, “variable” y “observación”. Además, para algunos pasos hemos incluido el eje “formulario” con el fin de comprobar los aspectos de calidad con respecto a la implementación de la HCE.

### 2.3.3 Definición del proceso de construcción de repositorios del PAAL

El proceso para la construcción de repositorios del PAAL está diseñado como un pipeline secuencial de trece pasos para asegurar la armonización, estandarización, completitud, de-duplicación y consistencia del contenido del conjunto de datos, guiado por la evaluación de DQ descrita anteriormente. La figura 2.2 muestra el esquema general del proceso de construcción de repositorios con calidad asegurada.



**Figura 2.2. Esquema general del proceso de construcción de los repositorios del proceso de atención a la alimentación infantil (PAAL) con calidad de datos (DQ) asegurada a partir de datos de historia clínica electrónica (HCE). En la parte superior el esquema y en la inferior la estructura canónica del repositorio con DQ asegurada con 1 fila por hecho de estudio (HdE)**

Como se puede observar en la figura, la entrada del proceso de construcción del repositorio es una única tabla con los datos brutos sin procesar (denominados comúnmente por su término en inglés como *raw data*), extraídos de las fuentes de datos. La estructura de los datos brutos puede variar en función de la estructura de los registros de HCE originales y del procedimiento de extracción preliminar. El proceso asistencial (PAAL) determina las

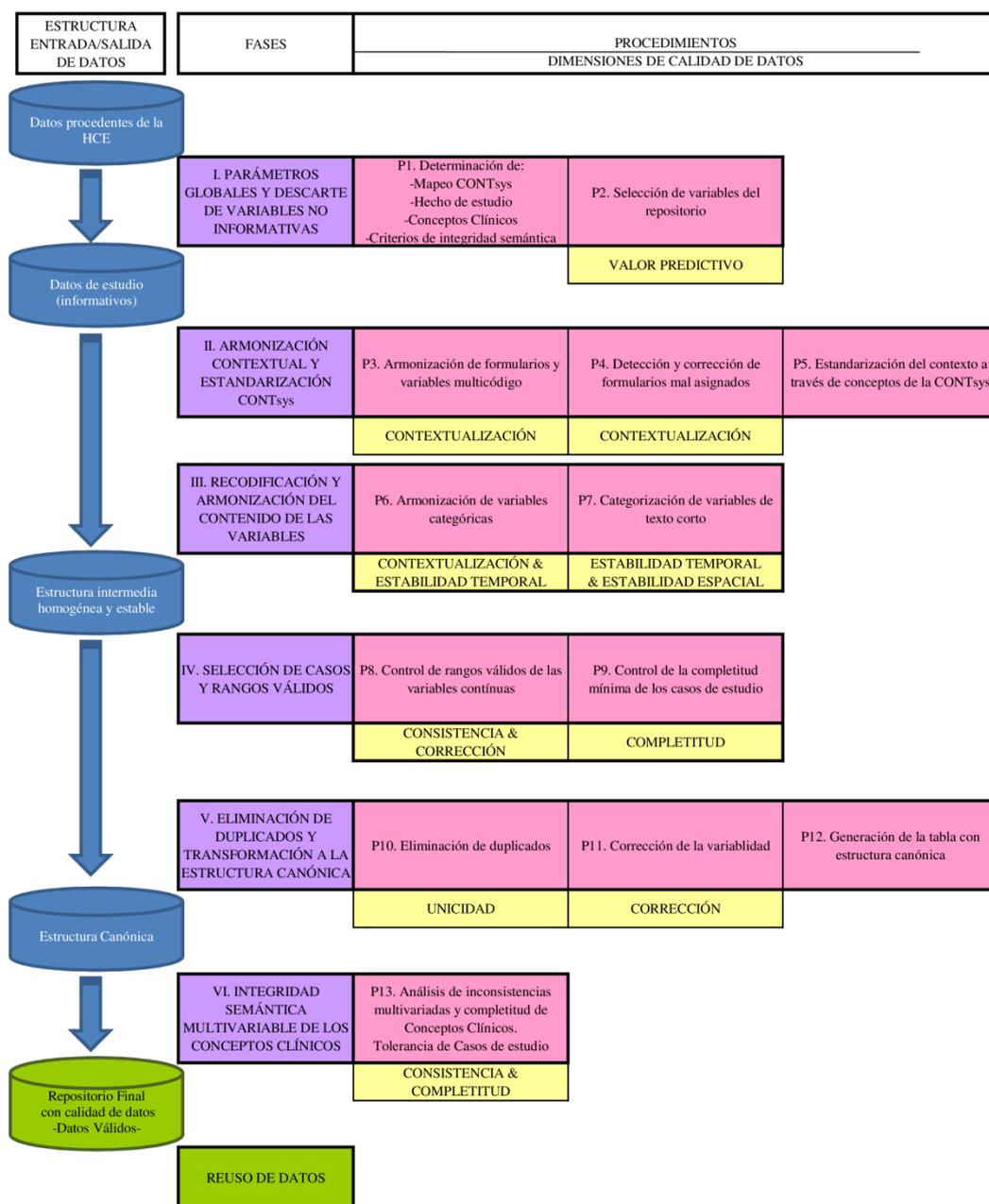
especificaciones y requerimientos del proceso de construcción, estableciendo la base para la definición de los criterios y controles de calidad. Las cuales se realizan conforme a las dimensiones de DQ definidas anteriormente.

Independientemente de la estructura de los datos brutos de entrada, el resultado del proceso es un repositorio con calidad asegurada, presentado en una estructura canónica. La estructura canónica consiste en un conjunto de filas, cada una de las cuales corresponde a un HdE con las variables clínicas desplegadas en columnas y agrupadas por los conceptos clínicos del dominio. Además, cada fila incluye la siguiente información:

- El resumen del contexto para el HdE, que incluye el resumen del sujeto de la asistencia, del episodio asistencial o del contacto y el número de formularios, servicios clínicos y usuarios involucrados en el registro de la información de este HdE en la HCE.
- Meta-información de DQ específica de los conceptos clínicos.
- Una puntuación global de DQ para el HdE con uno de los siguientes tres valores posibles: 1) calidad excelente, 2) con algunos problemas de calidad, pero tolerables, y 3) perdido o más allá del nivel tolerable.

## 2.4 Resultados

Proponemos el pipeline secuencial descrito gráficamente en la Figura 2.3 que está compuesto de trece procedimientos, y el aseguramiento de DQ antes y después de cada uno de ellos. Los procedimientos están diseñados para garantizar la armonización, estandarización, completitud, de-duplicación y consistencia del contenido del conjunto de datos, corrigiendo, descartando o evaluando la tolerancia de los datos afectados, según la funcionalidad de cada procedimiento. La figura 2.3 muestra las fases del proceso, los procedimientos de cada fase y la dimensión de DQ evaluada a lo largo de cada procedimiento. También se identifican las estructuras de datos utilizadas por los procedimientos de cada fase.



**Figura 2.3 Fases y procedimientos del proceso de construcción de los repositorios con calidad asegurada. Las estructuras de datos empleadas se muestran a la izquierda en azul y las dimensiones de calidad afectadas se muestran en amarillo bajo los procedimientos.**

Aunque el proceso se define como un pipeline secuencial, los procedimientos se agruparon en seis fases para presentar con mayor claridad la aplicación de los criterios de integridad y eficiencia de la información, así como las transformaciones de la estructura. Así, tras la parametrización de los criterios de integridad y el descarte de las variables no informativas en la Fase I, se garantiza la integridad estructural del repositorio mediante la armonización, estandarización y recodificación realizadas en las Fases II y III. Seguidamente, la integridad semántica univariante se aborda en las fases IV y V mediante la selección de datos para filtrar los casos y la de-duplicación del conjunto de datos para transformarlos a una estructura de datos canónica. Por último, la integridad semántica

multivariante se aborda en la fase VI, mediante el análisis de la consistencia y la completitud de los conceptos clínicos.

#### **2.4.1 Fase I. Definición de los parámetros del estudio y selección de variables informativas**

En la Fase I se definen los parámetros globales del proceso por el grupo de expertos en función del objetivo del re-uso y se descartan los datos no informativos para satisfacer los criterios de eficiencia. En el procedimiento P1 se determina: i) el mapeo de conceptos organizativos entre el sistema HCE y la norma CONTsys, ii) el tipo de HdE, iii) la clasificación de variables en informativas y no informativas en función de su influencia según el propósito del re-uso, iv) los conceptos clínicos y variables informativas que los componen y v) los criterios de integridad semántica.

A continuación, en el procedimiento P2, la selección de variables informativas del repositorio descarta las variables no predictivas o no informativas a efectos de reutilización. La dimensión de valor predictivo en los ejes de observaciones y variables mide el efecto de P2 en la DQ calculando los porcentajes de observaciones y variables no informativas, respectivamente.

Como resultado, la salida de la Fase I es la selección de los parámetros del proceso y los datos brutos "informativos para el estudio", que contienen el subconjunto de variables informativas de los datos brutos originales.

#### **2.4.2 Fases II y III. Obtención de una estructura homogénea y estable**

La integridad de la estructura de los datos se garantiza mediante la homogeneización, normalización y recodificación del contexto y del contenido de las variables. Esto es necesario porque las organizaciones sanitarias son dinámicas y pueden cambiar sus procesos de adquisición y codificación, afectando a la contextualización y a la estabilidad temporal de los repositorios de datos. Por lo tanto, en la fase II armonizamos el contexto corrigiendo los problemas de contextualización causados por datos no registrados debido a circunstancias indeterminadas del proceso de adquisición.

El primer procedimiento de esta fase, procedimiento P3, identifica y fusiona los formularios y variables multicódigo mediante el análisis de su relación temporal. Los formularios y variables multicódigo son aquellos formularios y variables que, a pesar de tener códigos diferentes en la HCE, son semánticamente equivalentes (por ejemplo, cuando en el sistema de HCE se han generado distintos códigos para las sucesivas versiones de una misma variable). El procedimiento P3 es un procedimiento semiautomático basado en la identificación de los formularios y variables multicódigo realizada por el grupo de expertos. La dimensión de contextualización en los ejes HdE, variable, observación y formulario mide el efecto de P3 en la DQ calculando los porcentajes de instancias afectadas por el problema multicódigo.

A continuación, el procedimiento P4 detecta y corrige los formularios asignados a sujetos incorrectos (por ejemplo, formularios de recién nacidos asignados incorrectamente a la madre). De nuevo, la dimensión de contextualización en los ejes HdE y de observación mide el efecto de P4 en la DQ calculando los porcentajes de instancias con formularios mal asignados.

Después, en el procedimiento P5, la información de contexto se estandariza a la CONTsys según la parametrización realizada en P1. Se trata de un procedimiento automático que implementa una función inyectiva, por lo que no afecta a la DQ.

La fase III aborda principalmente la variabilidad temporal debida a los cambios en proceso de adquisición y codificación de la información en los entornos cambiantes de las organizaciones sanitarias. Esto se hace mediante la armonización del contenido de los datos para obtener una estructura estable y homogénea a lo largo del periodo de estudio.

El procedimiento P6 evalúa y corrige los ejes HdE, variable y observación del efecto sistemático de las agregaciones y desagregaciones aplicadas a lo largo del tiempo a los sistemas de codificación de las variables categóricas de la HCE, debido a los cambios en la aplicación de los protocolos sanitarios. Así, se analiza la temporalidad de los valores categóricos de cada variable y se asigna un código armonizado siguiendo las indicaciones del grupo de expertos. Identificamos dos casos que pueden ser corregidos por este procedimiento: efecto 1) cambios producidos por la recodificación o renombramiento de las categorías en el sistema de HCE sin aumentar o disminuir la precisión de los valores; y efecto 2) cambios producidos por los cambios de protocolo que afectan a la precisión de los valores al disminuir o aumentar el número de categorías (por ejemplo, la inclusión de una nueva ubicación donde realizar el contacto piel con piel). La dimensión de contextualización en los ejes HdE, observación y variable mide el efecto de P6 en las instancias afectadas por el efecto 1 mediante el cálculo de los porcentajes de observaciones y variables afectadas por la recodificación. Además, la dimensión de estabilidad temporal en los ejes HdE, observación y variable mide el efecto de P6 en las instancias afectadas por el efecto 2 calculando los porcentajes de observaciones y variables afectadas por el cambio de protocolo. Además, también se informa de la suma de HdE, observaciones y variables afectadas por el efecto 1 y el efecto 2.

Al igual que las variables categóricas, aquellas cuyos valores son de texto libre pueden verse afectadas por tendencias temporales o por los distintos profesionales sanitarios que han intervenido en la generación de los datos. El procedimiento P7 realiza la categorización de las variables de texto corto utilizando una categorización de patrones simples para obtener una representación homogénea entre los creadores de datos a lo largo del tiempo. El procedimiento está diseñado para que un experto clínico realice actualmente la interpretación del texto libre de forma manual, por lo que no se mide la estabilidad espacial ni la estabilidad temporal. El resultado tras la fase III es un repositorio intermedio con una estructura de datos homogénea y estable.

#### **2.4.3 Fases IV y V. Abordaje de los criterios de integridad semántica de los datos a nivel univariante y preparación de la estructura canónica del repositorio**

La fase IV comprueba la consistencia, corrección y completitud de los datos según los criterios establecidos en P1. Así, el procedimiento P8 realiza dos acciones 1) la discretización de las variables continuas y 2) el control de rangos univariantes válidos. La primera acción se basa en los criterios del grupo de expertos para determinar los rangos de discretización, pero también pueden basarse en algoritmos de agrupación univariante (como k-means). Además, el control de los rangos válidos se realiza en dos niveles: valores fuera de rango y valores improbables. Las dimensiones de consistencia y corrección en los ejes HdE, variable y observación miden el efecto de P8 en la DQ calculando los porcentajes de instancias con valores fuera de rango y con rangos improbables, respectivamente.

A continuación, el procedimiento automático P9 controla la completitud del HdE para obtener la colección de variables clínicas y de contexto que cumplen los criterios de completitud mínima definidos en P1. La dimensión de completitud en los ejes de HdE y observa-

ción mide el efecto de P9 en la DQ calculando los porcentajes de instancias incompletas en el repositorio.

El objetivo de la fase V es obtener la estructura canónica para adaptar la representación del repositorio. Esto requiere una de-duplicación (también conocida como resolución de entidades) para obtener un único valor por variable y HdE (criterio de unicidad). La de-duplicación se realiza mediante dos pasos diferentes sobre los HdE y las observaciones del repositorio. En primer lugar, el procedimiento P10 elimina las observaciones duplicadas. En segundo lugar, es necesario determinar el valor más fiable para las observaciones que han sido adquiridas varias veces con valores diferentes (por ejemplo, el momento del contacto precoz piel con piel se registra durante las dos primeras horas desde el nacimiento y, posteriormente, en el ingreso en planta, pero sólo debe quedar una observación en el repositorio). Después de la de-duplicación, el conjunto de datos puede transformarse en la estructura canónica en la que cada fila representa un HdE con las variables clínicas desplegadas en columnas.

Así, el procedimiento P10 elimina automáticamente las observaciones duplicadas que son instancias con el mismo valor obtenidas en diferentes formularios (duplicados inter-formulario) o en registros sucesivos del mismo formulario (duplicados intra-formulario). La dimensión de duplicación en los ejes de HdE y observación mide el efecto de P10 en la DQ calculando los porcentajes de HdE y observaciones duplicadas, respectivamente. Esta última métrica puede desagregarse en duplicados intra-formulario e inter-formulario.

A continuación, en el procedimiento P11, para resolver la variabilidad (tanto intra-formulario como inter-formulario) se pueden aplicar procedimientos automáticos (último valor, media, etc.) o puede requerir una selección manual por parte del experto. La dimensión de corrección en los ejes de HdE y observación mide el efecto de P11 en la DQ calculando los porcentajes de HdE afectados y los porcentajes de observaciones con adquisiciones múltiples, respectivamente. Esta última métrica puede desagregarse en observaciones intra-formulario e inter-formulario para propósitos de monitorización.

El resultado de la fase V es una estructura canónica realizada en el procedimiento P12 sobre el conjunto de HdE de-duplicados. Además, el resumen de la información de contexto de cada HdE también se calcula mediante funciones de agregación y también se añade a cada fila de la estructura canónica.

#### **2.4.4 Fase VI. Evaluación de la integridad semántica multivariante de los conceptos clínicos**

Por último, en la fase VI se evalúa la integridad semántica multivariante. Esta se realiza mediante un único procedimiento (P13) a nivel de conceptos clínicos, en dos pasos. En primer lugar, el procedimiento genera el espacio formado por cada par de valores que toman las variables de un HdE en cada concepto clínico, y el experto determina la consistencia de cada par diferente según los criterios definidos en P1 (por ejemplo, es inconsistente que la variable de momento de inicio de lactancia materna indique que no se inicia lactancia materna y que la variable de lugar de inicio de lactancia materna indique que se inicia en paritorio). Posteriormente se establece automáticamente que una instancia de un concepto clínico (N-tupla) es inconsistente si contiene al menos un par de valores inconsistente. Además, el procedimiento marca las N-tuplas incompletas. A continuación, para cada N-tupla inconsistente y/o incompleta, un experto determina su conveniencia para la reutilización, dado el contexto clínico de la N-tupla (relacionado con toda la información clínica del HdE y basado en los criterios definidos en P1). Por último, se asigna una puntuación global de DQ a cada HdE basada en la evaluación del procedimiento P13. Aquellos HdE sin ninguna inconsistencia/incompletitud recibieron una puntuación global de DQ de "1) Excelente DQ"; aquellos HdE con alguna inconsistencia/incompletitud tolerable reci-

bieron una puntuación de “2) Con problemas de calidad tolerables”; y el resto de HdE recibieron una puntuación de “3) Con inconsistencia/incompletitud no tolerable o caso perdido”. La dimensión de consistencia en los ejes HdE, concepto clínico y observación mide el efecto de P13 en la DQ calculando los porcentajes de instancias multivariantes inconsistentes. De forma similar para la dimensión de completitud.

El resultado de la Fase VI es el repositorio final de datos con calidad asegurada. Este repositorio en estructura canónica puede explotarse directamente con fines de monitorización e investigación. Además, la meta-información de DQ proporciona una evaluación directa de cada HdE incluido en el repositorio. Por último, las medidas de DQ realizadas durante el proceso antes y después de cada procedimiento proporcionan 1) una evaluación comparativa del efecto del proceso de construcción y 2) una evaluación absoluta del repositorio de datos, desglosada por dimensiones de DQ.

## 2.5 Discusión

En este trabajo hemos propuesto un proceso computacional para extraer repositorios con calidad asegurada representados por una estructura canónica a partir de datos brutos de múltiples sistemas de HCE. El proceso consiste en trece procedimientos para la armonización, estandarización, completitud, de-duplicación y consistencia del repositorio de datos. Un marco de DQ controla el efecto sobre el repositorio de cada procedimiento que afecta a los datos para asegurar una gestión de calidad total del proceso de construcción del repositorio.

### 2.5.1 Significancia

La disponibilidad de repositorios de datos clínicos de calidad para su reutilización es un objetivo sobre el que existe consenso internacional y en el que se están realizando grandes esfuerzos (M. G. Kahn & Weng, 2012; Klazinga et al., 2011; Safran et al., 2007; Weiner & PJ, 2009). El uso de repositorios como fuente de datos para la investigación facilita el reclutamiento y seguimiento de los sujetos, reduciendo los inconvenientes y el tiempo y dinero invertido (Weng et al., 2012). Además, también son una buena fuente de datos para construir modelos predictivos (Klann et al., 2013; Mathias et al., 2013, p. 5; Watson et al., 2011). También cabe destacar su utilidad en el ámbito de la calidad asistencial, ya que los repositorios de datos clínicos son la principal fuente de datos para orientar las decisiones de gestión estratégica (Box et al., 2010; Foglia et al., 2013; Persell et al., 2011), monitorizar los procesos clínicos (Aizpuru et al., 2012; Bundy et al., 2012; Reed et al., 2012) o diseñar y evaluar los procedimientos clínicos para garantizar la seguridad de los pacientes (Coloma et al., 2011; Patterson et al., 2013; Wise & Lovell, 2013).

No hemos encontrado referencias sobre indicadores poblacionales procedentes de la HCE sobre alimentación infantil. Este primer artículo, de la serie de dos, justifica la necesidad de reutilizar repositorios del proceso de atención a la alimentación infantil, y describe el diseño de un proceso para construirlos a partir de la HCE. Se trata de un paso clave para poder realizar estudios poblacionales sobre importantes problemas de salud infantil.

La incorporación de conceptos del estándar CONTsys (UNE-EN ISO 13 940) (European Committee for Standardization Health Informatics, 2007) para definir el contexto organizativo en nuestro modelo de información, así como la agrupación de las variables en torno a conceptos clínicos, nos permitió i) integrar en el repositorio elementos procedentes de diferentes fuentes de datos con el mismo significado; ii) facilitar la definición de métricas de

DQ independientes del nivel global de análisis requerido para la reutilización (HdE); y iii) guiar el proceso de construcción y facilitar la incorporación del investigador clínico en el proceso de construcción.

La adopción de un marco de DQ basado en dimensiones de DQ para medir el efecto de los procedimientos permite una gestión estandarizada de la calidad total del proceso. Esto garantiza la calidad de los conjuntos de datos y permite la comparabilidad y benchmarking de los repositorios.

En consecuencia, la combinación del uso de los conceptos estandarizados de la norma CONTsys y las dimensiones de DQ de (Sáez et al., 2012) permiten la replicación de la metodología y su uso en estudios poblacionales multicéntricos, incluso en contextos organizativos diferentes.

Al mismo tiempo, los procedimientos secuenciales, basados en los criterios de eficacia e integridad de los datos, están diseñados para mejorar gradualmente la DQ y facilitar el uso de este enfoque en los problemas más complejos, como la variabilidad y la consistencia, que requieren la intervención de expertos en las etapas finales. En concreto, el descarte de variables no informativas en la fase I produce una mejora considerable de la eficacia de los procedimientos posteriores sin perder valor predictivo. A continuación, se aborda la integridad de la estructura (fases II y III), corrigiendo los problemas de contextualización y estabilidad temporal. Esto permite obtener una estructura homogénea y estable sin cambiar el sentido clínico original, lo que facilita el análisis semántico realizado por los expertos posteriormente (fases IV a VI).

Nuestro enfoque proporciona una metodología cuantitativa que permite el procesamiento de la información sanitaria recogida en la HCE, para generar información fiable y válida para ser utilizada en estudios poblacionales. Este estudio destaca la relevancia de la transparencia de la DQ en los repositorios de investigación biomédica y explora la aplicabilidad de marcos estandarizados para su evaluación en escenarios reales.

En consecuencia, los procesos de reutilización del repositorio para la extracción de conocimiento y la toma de decisiones dispondrán de meta-información sobre la DQ que se resume en la puntuación global de DQ. Esto permitirá monitorizar el proceso asistencial para realizar mejoras aplicando metodologías de bucle cerrado como los métodos de gestión plan-do-check-act (PDCA) (Röthlin, 2010).

### **2.5.2 Limitaciones**

El proceso de construcción del repositorio propuesto en este trabajo se limita a datos categóricos o numéricos, con un nivel de detalle suficiente para que la información pueda ser agrupada semánticamente, lo que facilita el análisis multifactorial.

Dependiendo del nivel de calidad de los datos de entrada, la cantidad de trabajo manual puede ser elevada en los procedimientos que abordan la variabilidad (P11) y el análisis de consistencia (P13). A partir de este trabajo manual, se pueden implementar reglas definidas por el usuario para resolver estos problemas de forma automática para las nuevas extracciones. El análisis de inconsistencias de los conceptos clínicos es la tarea manual más compleja y requiere una formación y dedicación adicionales, pero la cantidad relativa de trabajo manual disminuyó a medida que se cargaron más datos con la misma estructura.

En el procedimiento P6 hemos utilizado un método basado en tablas para la descripción de los cambios en las categorías de las variables a lo largo del tiempo. A pesar de que en un periodo de tiempo habrá el mismo conjunto de categorías, es posible que su distribución varíe con el tiempo. Del mismo modo, puede ocurrir cuando diferentes usuarios o centros generan datos clínicos (Röthlin, 2010). En ambos casos, podrían surgir problemas de

estabilidad espacial o temporal relacionados con los cambios de concepto. No tener en cuenta estos hechos podría dar lugar a problemas en posteriores estudios de investigación poblacional.

Como posible debilidad de nuestro método, los filtros aplicados a lo largo de las diferentes fases pueden estar distorsionando el espacio muestral real, al descartar los datos que no cumplen ciertos criterios de calidad. Publicaciones recientes estudian el efecto de distorsión de la realidad que puede provocar la limpieza de datos (Sáez, Robles, et al., 2017).

### 2.5.3 Trabajos futuros

El diseño de procedimientos para la extracción de datos relevantes, así como la evaluación de la DQ, nos ha permitido especificar los requisitos para la construcción de repositorios para la reutilización de datos clínicos. En la actualidad, los autores están revisando estos procedimientos para su generalización a otros servicios sanitarios, con el fin de desarrollar un software genérico para la construcción de repositorios con calidad asegurada para la reutilización de la información clínica y el aseguramiento de calidad on-line.

La definición de los conceptos clínicos se realizó agrupando las variables relevantes de la HCE. Actualmente, estamos trabajando en la definición de los conceptos clínicos siguiendo la estructura de arquetipos UNE-EN ISO 13 606 para mejorar su definición mediante una estructura de árbol interoperable estandarizada en la que podamos anexar las reglas de integridad del concepto clínico.

En función de los problemas de estabilidad espacial mencionados anteriormente, exploraremos las distancias entre las distribuciones de probabilidad de los datos relacionados con las diferentes fuentes que los generan (Sáez, Robles, et al., 2017). Esto permitirá la identificación automática de los cambios de categoría, así como la detección de anomalías en las fuentes de datos.

Otra dimensión de la DQ, aunque no sea intrínsecamente medible a partir de los datos, sigue siendo importante: la confiabilidad. En (Sáez et al., 2012) definen la confiabilidad como el grado de reputación de los actores e instituciones que participan en la adquisición de datos. Así, la fiabilidad está relacionada con el grado en que la organización gestiona la DQ y, en este sentido, se pueden medir los siguientes factores: 1) si el proceso asistencial de referencia está estructurado por metodologías de gestión de procesos (en nuestro caso el PAAL en el marco de la *European Foundation for Quality Management*, EFQM); 2) si la organización cuenta con acreditaciones en el ámbito clínico de referencia (en nuestro caso BFHI) y 3) un equipo multidisciplinar con formación y experiencia en investigación clínica en DQ. Estos aspectos se estudiarán en trabajos posteriores.

## 2.6 Conclusiones

En este primer estudio de una serie de dos, hemos propuesto y diseñado el proceso de construcción de repositorios con calidad asegurada de datos del proceso de alimentación infantil desde el nacimiento hasta los dos años de edad. Para ello, se definieron trece pasos para la armonización, estandarización, completitud, de-duplicación y consistencia del contenido del conjunto de datos. Además, la calidad de entrada y salida de datos para cada paso se controla en base a ocho dimensiones de calidad: valor predictivo, corrección, unicidad, consistencia, completitud, contextualización, estabilidad temporal y estabilidad espacial bajo un paradigma de gestión de la calidad de datos total (TDQM). La contribución de este trabajo permite la construcción del primer repositorio con calidad asegurada para la reutilización de la información sobre la alimentación infantil en el periodo perinatal

para el seguimiento de las actividades sanitarias y la investigación que se presenta en el segundo trabajo de esta serie.



## 3 Construcción del repositorio perinatal

En este capítulo, describe la aplicación del proceso computacional de construcción de repositorios con calidad asegurada basados en datos poblaciones de HCE, definido en el capítulo anterior. Se detalla la aplicación del proceso a la construcción del repositorio de alimentación infantil para el ámbito perinatal, a partir de los datos de HCE del HIS del hospital relativas al parto-nacimiento y primeros días de vida. El objetivo de la construcción de este primer repositorio con calidad asegurada era, por un lado, comprobar la usabilidad del proceso de construcción y, por otro, el re-uso de la información clínica para estudiar preguntas relevantes de la BFHI y monitorizar su despliegue en nuestro hospital. Se presenta la aplicación de los 13 procedimientos semi-automáticos definidos en el proceso de construcción para evaluar, recuperar o descartar datos clínicos, midiendo la DQ antes y después de cada procedimiento utilizando métricas relacionadas con ocho dimensiones de calidad: valor predictivo, corrección, unicidad, consistencia, completitud, contextualización, estabilidad temporal y estabilidad espacial. Como resultado se obtuvo el repositorio perinatal con calidad de datos asegurada (C2).

En los siguientes capítulos se abordará en primer lugar la experiencia de integración de repositorios de distintas áreas de salud (Área de Salud V – Altiplano de Murcia y Hospital 12 de Octubre de Madrid), mediante el estándar de información sanitaria UNE-EN ISO 13 606 y uso concurrente de UNE-EN ISO 13 940. Posteriormente, en el capítulo 5 se abordará la ampliación del repositorio perinatal descrito en este capítulo con la información de seguimiento de la alimentación hasta los dos años, así como la paquetización de elementos software del proceso de construcción y la utilidad del nuevo IDR en la implementación Buenas Prácticas.

*El contenido de este capítulo es una adaptación para la tesis del artículo publicado (García-de-León-Chocano et al., 2016) – contribución Pu2. Como resultado, se construyó el repositorio con calidad de datos asegurada de alimentación infantil para el ámbito perinatal en la Gerencia del Área de Salud V – Altiplano, contribución C2. La publicación del artículo Pu2 y la realización de la contribución C2, fueron posibles gracias a los trabajos previos desarrollados en los proyectos de investigación HQTQ-1 y HQTQ-2.*

### 3.1 Introducción y antecedentes

El éxito de la lactancia materna depende de múltiples factores (Meedya et al., 2010), aunque entre todos, los que tienen lugar durante el período perinatal son cruciales. En este sentido, la mejora de las prácticas hospitalarias es un aspecto clave para elevar las tasas de lactancia materna en la población, ya que las actividades profesionales durante el parto (Bai et al., 2013) y los primeros días de vida (WHO, UNICEF, 1989) tienen una fuerte influencia en el inicio y consolidación de la lactancia materna (Häggkvist et al., 2010; Perrine et al., 2012; Venancio et al., 2012).

En España, nuestra estrategia nacional de salud perinatal (Grupo de trabajo de la Guía de Práctica Clínica sobre atención al parto normal, 2010) junto con la BFHI (WHO UNICEF, 2009), acuerda limitar las intervenciones innecesarias que interfieren con los procesos

biológicos materno-infantiles (parto - nacimiento, vinculación y lactancia materna), y también promueve actividades profesionales que refuerzan la confianza y autonomía de las mujeres. Sin embargo, la implementación de los consejos propuestos en ambas estrategias es extremadamente limitada, como revelan los datos europeos sobre salud perinatal. Además, estos datos confirman una gran variabilidad entre países, asociada a diferencias en el tipo de atención sanitaria prestada, así como en el tipo de registro utilizado como fuente de datos ([Euro-Peristat project with SCPE and Eurocat, 2010](#)).

Los datos perinatales actuales no son suficientes para monitorizar las intervenciones de salud que tienen como objetivo mejorar la implementación de las estrategias de salud de las madres y los recién nacidos ([Euro-Peristat project with SCPE and Eurocat, 2010](#)). En la actualidad no existen fuentes de datos poblacionales publicadas sobre el período perinatal que comprendan variables predictivas sobre el abandono prematuro de la lactancia materna. Las variables clínicas relacionadas con la interrupción temprana de la lactancia materna, como el contacto temprano piel con piel ([Moore & Bergman, 2007](#)), el inicio temprano de lactancia materna ([Callendret et al., 2015](#)), la suplementación con la lactancia materna ([Chantry et al., 2014](#); [Häggkvist et al., 2010](#)) o los problemas de lactancia materna ([Häggkvist et al., 2010](#)), no se registran sistemáticamente en la HCE.

La disponibilidad de datos completos, comparables y accesibles sobre la salud de las madres y los bebés es una preocupación para los principales grupos europeos de seguimiento perinatal ([Euro-Peristat project with SCPE and Eurocat, 2010](#)). En este sentido, los registros de HCE son la fuente de datos ideal, cuando se puede asegurar la calidad de los datos (DQ) ([Weiskopf & Weng, 2013](#)), ya que brindan una cobertura poblacional ininterrumpida y permitirían la evaluación comparativa si sus registros estuvieran estandarizados ([Klazinga et al., 2011](#)).

El "Hospital Virgen del Castillo" fue designado como Hospital BFHI en 2004. Posteriormente, dentro de la experiencia HQTQ, el Proceso de Atención a la Alimentación del Lactante (PAAL) permitió ampliar el alcance desde el hospital hasta la atención primaria, con el fin de adecuar las actividades clínicas a las recomendaciones de evidencia desde el nacimiento hasta los dos años de edad. En nuestro caso, la disponibilidad de información de calidad sobre las prácticas y procedimientos realizados, así como sobre el estado de salud de las madres y los bebés, fue un tema clave para la sostenibilidad de la BFHI. Por lo tanto, se planteó la necesidad de construir un repositorio de alimentación infantil a partir de datos poblacionales de HCE.

Con este repositorio podemos monitorizar el PAAL y abordar preguntas sobre la alimentación infantil planteadas por la BFHI, tales como: "¿Cuáles son las tasas de alimentación al alta?", "¿Cuántos niños reciben suplementos y qué razones podrían explicarlo?", "¿Cuáles son las tasas de contacto precoz piel con piel?", "¿Cuáles son las razones de la separación madre-bebé?", "¿Qué problemas tienen las madres al amamantar?" y "¿Son eficaces las intervenciones que abordan cada uno de los problemas identificados?", entre otras.

En este segundo artículo, presentamos el primer repositorio del PAAL con calidad de datos asegurada a partir de datos de HCE, con el objetivo de monitorizar el PAAL en el ámbito perinatal y estudiar preguntas relevantes de la BFHI. Por lo tanto, en esta segunda parte de nuestro estudio, describimos la construcción del repositorio de datos; que es el primer caso de uso del proceso de construcción de repositorios bajo la gestión de la calidad de datos total (TDQM) descrita en el capítulo anterior ([García-de-León-Chocano et al., 2015](#)), y que constituye la pieza central de un repositorio completo sobre el PAAL desde el nacimiento hasta los dos años de edad.

## 3.2 Metodología

Hemos aplicado el proceso de construcción de repositorios con calidad asegurada descrito definido en el capítulo 2 ([García-de-León-Chocano et al., 2015](#)) al caso perinatal. Como punto de partida, se definieron las especificaciones y requisitos del repositorio perinatal dadas las necesidades de información para la monitorización del PAAL y la investigación sobre el período perinatal. De acuerdo con estas necesidades, el episodio asistencial perinatal se definió como el Hecho de Estudio (HdE) del repositorio. Este episodio asistencial incluye información sobre madres y recién nacidos relativa a la gestación, parto-nacimiento y adaptación hasta el alta. El recién nacido es el sujeto de atención del PAAL, y los datos relacionados con la madre nos permiten completar la información necesaria para la monitorización del recién nacido. El período de estudio fue de enero de 2009 a diciembre de 2011.

### 3.2.1 Materiales

En el "Hospital Virgen del Castillo", la información del episodio asistencial perinatal se registra en el Sistema de Información Hospitalario (HIS) desde 2008 para todos los partos. El HIS es el software Siemens AG Selene, que contiene formularios clínicos, peticiones e informes, entre otros elementos. Los datos estructurados de los formularios clínicos del HIS incluyen toda la información necesaria para su reutilización. Cada registro de formulario contiene las observaciones y la información contextual de las variables clínicas registradas. La información contextual incluye la identidad del usuario, la marca de tiempo de la observación y el identificador del proceso asistencial del paciente, entre otra información.

La información del episodio asistencial perinatal se registra en 11 formularios diferentes a lo largo de los dos contactos el episodio asistencial: paritorio/quirófano y planta (maternidad y/o neonatología). Nueve de los formularios pertenecen al recién nacido y los otros dos formularios pertenecen a la madre. Cada formulario está diseñado para recoger la información proporcionada por cada uno de los grupos profesionales involucrados en el proceso: matronas, obstetras, enfermeras de reanimación post-anestésica, enfermeras de planta y pediatras. En total, la HCE perinatal comprende 223 variables (101 de ellas se registran en formularios asignados a la madre y 122 en formularios asignados al recién nacido). Varias variables se definen en más de un formulario, ya que pueden ser observadas por diferentes grupos profesionales en diferentes etapas del proceso asistencial.

### Conjunto de datos inicial

La entrada del proceso de construcción del repositorio es una tabla única que denominamos "*raw data*" o datos brutos sin procesar, que corresponde al conjunto de datos extraído de la base de datos Oracle del HIS Selene y cargado en una tabla de SQL Server utilizando la herramienta de integración de datos Pentaho. Cada fila representa una observación de una variable clínica para un episodio perinatal junto con su información contextual. Los datos brutos corresponden a 2048 episodios perinatales (identificados por el `Id_Nacimiento`) para el período 2009 - 2011, con un total de 433 308 observaciones de 223 variables clínicas. Estas observaciones se recopilaron en 14 663 registros de formulario de los 11 formularios diferentes identificados. Las variables son categóricas, numéricas y de texto corto, aptas para categorización. En resumen, el conjunto de datos inicial son los datos brutos formados por 433 308 filas donde cada fila incluye la siguiente información:

- El código de la variable.
- El valor de la observación
- El Id\_Sujeto correspondiente al recién nacido y otra información del sujeto de la asistencia
- El Id\_Nacimiento y otra información del episodio asistencial perinatal
- El Id\_Contacto y otra información del contacto
- La identificación (Id\_Usuario) del usuario (profesional sanitario) que registró la información
- Rol del usuario que registró la información (matrona, obstetra, enfermera de reanimación post-anestésica, enfermera de planta o pediatra)
- La fecha y hora a la que se registró la información
- El código del formulario en el cual se registró la información

### 3.2.2 Métodos

La descripción de los procedimientos para la construcción del repositorio y las métricas realizadas para asegurar su DQ general fue definida en el capítulo anterior ([García-de-León-Chocano et al., 2015](#)). En concreto, en el apartado Metodología se presentó el marco de DQ que guió el diseño del proceso de construcción, en base a las ocho dimensiones de DQ ([Sáez et al., 2012](#)): Completitud, Consistencia, Duplicación, Corrección, Estabilidad temporal, Estabilidad espacial, Contextualización y Valor predictivo. Además, en nuestro repositorio, medimos las dimensiones de DQ en los cinco ejes definidos en el capítulo anterior: HdE, concepto clínico (N-tupla), variable, observación y formulario. Finalmente, el proceso de construcción del repositorio perinatal siguió el proceso secuencial descrito en la sección de resultados del capítulo anterior (figura 2.3).

Para el repositorio perinatal, los parámetros de estudio de la fase I fueron especificados por el grupo de expertos, guiado por el propósito de la reutilización: la monitorización del PAAL en el campo perinatal y el estudio de preguntas relevantes de la BFHI, específicamente:

- El mapeo entre los conceptos organizacionales de CONTsys y Selene HIS se describe en la Tabla 3.1:

**Tabla 3.1 Mapeo entre los conceptos organizativos de la norma CONTsys y el sistema de información hospitalario (HIS) Selene.**

Conceptos de la Contsys	Conceptos del HIS Selene	Equivalencia en nuestro caso de estudio
Sujeto de la asistencia	Paciente	Recién nacido
Episodio asistencial	Proceso	Episodio asistencial perinatal
Contacto	Acto clínico	Paritorio / Quirófano Planta (Maternidad o Neonatología)

- El HdE se configuró para el episodio asistencial correspondiente al episodio perinatal y se identificó por Id\_Nacimiento.

- Las variables informativas se identificaron a partir de revisiones de evidencia sobre los factores de riesgo de abandono prematuro de lactancia materna, en las que se determinaron las variables independientes, dependientes y de confusión.
- Los conceptos clínicos y las variables que los conforman se definieron a partir de los indicadores clínicos para el seguimiento de la BFHI (WHO, UNICEF, 2007). Además, los indicadores previamente establecidos en el PAAL se definieron a partir de una revisión exhaustiva de la evidencia y las principales bases de los indicadores internacionales de salud perinatal (Euro-Peristat project with SCPE and Eurocat, 2010). La tabla 3.2 muestra los conceptos clínicos del PAAL para el período perinatal.

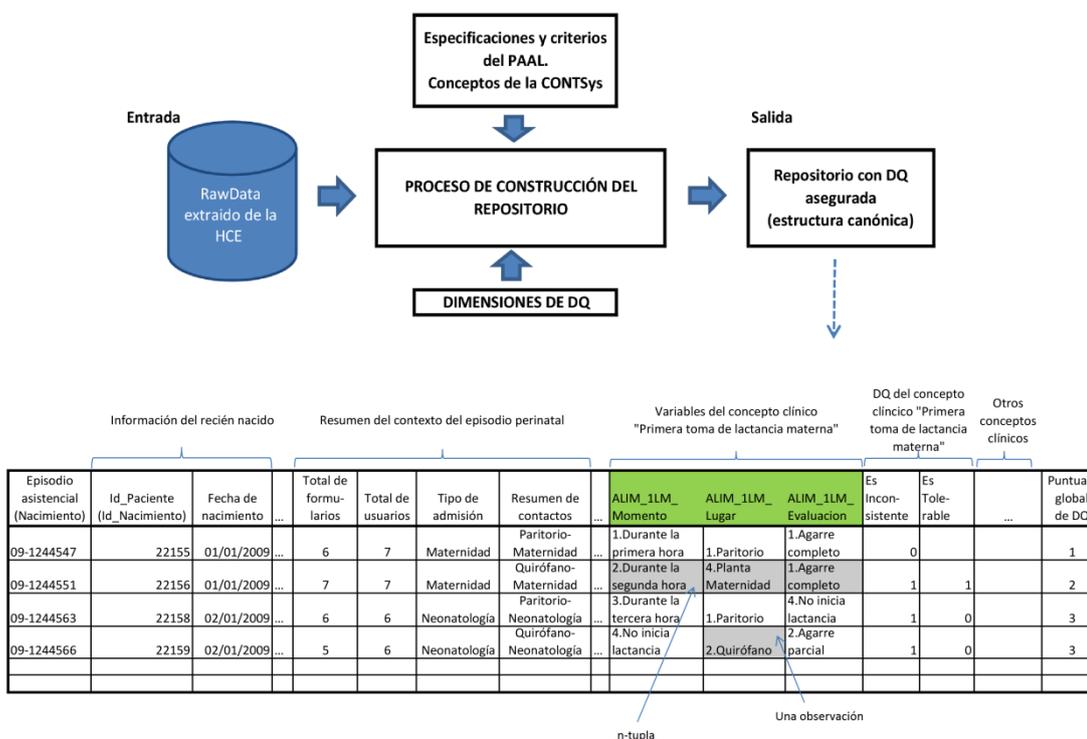
**Tabla 3.2 Conceptos clínicos del proceso de atención a la alimentación infantil (PAAL) durante el periodo perinatal**

Conceptos clínicos del PAAL durante el periodo perinatal
Hábitos tóxicos durante la gestión
Intervenciones obstétricas
Reanimación neonatal
Contacto precoz piel con piel
Inicio temprano de lactancia materna
Evaluación del pecho
Diagnósticos enfermeros de lactancia
Alimentación al alta (Recuerdo 24 horas)

- La tabla A.1 del apéndice A enumera el subconjunto de variables identificadas como variables informativas para el repositorio perinatal. La tabla incluye para cada variable: su descripción, el tipo de variable (dependiente, independiente y de confusión) y el concepto clínico al que pertenece la variable.
- Se establecieron los criterios mínimos de completitud del HdE en el repositorio perinatal se establecieron a: los nacimientos de bebés que completaron el proceso de atención de salud en el hospital.
- Se establecieron los rangos de discretización y valores de aceptabilidad para cada variable numérica.
- Para resolver la variabilidad (P11), se seleccionó la última observación para la variabilidad intra-formulario ya que refiere a variabilidad generada por el mismo grupo profesional. Además, para resolver la variabilidad de la inter-formulario, se decidió realizar una selección manual del valor más fiable ya que refiere a variabilidad generada por diferentes grupos profesionales. Los criterios para esta selección manual se definieron de acuerdo a todo el contexto clínico del nacimiento.
- Se establecieron los criterios para la consistencia multivariante. La completitud multivariante del procedimiento P13 se deshabilitó porque todas las variables relevantes para el repositorio eran obligatorias en el HIS.
- Se establecieron los criterios para establecer la tolerancia de los nacimientos de acuerdo con las inconsistencias en los conceptos clínicos en todo el contexto clínico.

### 3.3 Resultados

El resultado del proceso es el repositorio perinatal con calidad asegurada, presentado en una estructura canónica) que consta de una fila por Episodio asistencial perinatal (identificado por el Id\_Nacimiento), con las variables clínicas informativas desplegadas en columnas y agrupadas por los conceptos clínicos, relativos a los estados de salud del recién nacido y las actividades de atención de salud perinatal del PAAL (ver Tabla 3.2). Además, cada fila incluye la información sobre el recién nacido, el resumen del contexto del episodio asistencial perinatal y la puntuación global de DQ para el episodio perinatal. La Figura 3.1 muestra los detalles de la estructura canónica del repositorio perinatal con calidad asegurada. Por limitaciones de espacio solo hemos incluido el concepto clínico "Primera toma de lactancia" en la figura.



**Figura 3.1 Construcción del repositorio de datos perinatales del proceso de atención a la alimentación infantil (PANA), con calidad de datos (DQ) asegurada. En la parte superior el proceso de construcción y en la inferior la estructura canónica del repositorio final (ejemplo del concepto clínico "Primera toma de lactancia"). La variable ALIM\_1LM\_Momento es "Momento de la primera toma de lactancia", ALIM\_1LM\_Lugar es "Lugar de la primera toma de lactancia" y ALIM\_1LM\_Evaluación es "Evaluación de la primera toma de lactancia".**

En el siguiente apartado detallamos los resultados de la evaluación de calidad obtenidos durante el proceso de construcción. Las métricas correspondientes al episodio asistencial perinatal se presentan en relación a los nacimientos para facilitar la comprensión. Las tablas 3.3 - 3.6 detallan los resultados de la evaluación de DQ de los procedimientos agrupados por fases. Para evaluar el efecto de cada procedimiento, se cuantificó cada métrica antes y después de la aplicación de cada procedimiento a los datos. Para simplificar la comparación de resultados de diferentes procedimientos, los resultados del procedimiento P6 incluidos en la Tabla 3.5 se calcularon como un porcentaje sobre el número total de observaciones y no solo sobre las observaciones de variables categóricas. Además, en el

procedimiento P13, los nacimientos y n-tuplas que permanecen afectados son aquellos que muestran alguna inconsistencia que el clínico señaló como no tolerada.

### Fase I. Descarte de variables no informativas

**Tabla 3.3 Evaluación comparativa de calidad de datos (DQ) antes y después del procedimiento P2**

ID	Métrica	Evaluación de DQ antes del procedimiento	Evaluación de DQ después del procedimiento
P2	% variables no informativas	143/223 (64.13%)	0/80 (0%)
	% observaciones no informativas	229 229/433 308 (52.90%)	0/204 079 (0%)

### Fases II y III. Obtención de una estructura homogénea y estable

**Tabla 3.4 Evaluación comparativa de calidad de datos (DQ) antes y después de cada procedimiento (P3-P7)**

ID	Métrica	Evaluación de DQ antes del procedimiento	Evaluación de DQ después del procedimiento
P3	% formularios multicódigo	2/11 (18.8%)	0/9 (0%)
	% variables multicódigo	7/80 (8.75%)	0/73 (0%)
	% observaciones afectadas por contextualización en formularios y/o variables	65 967/204 079 (32.32%)	0/204 079 (0%)
	% nacimientos afectados por contextualización en formularios y/o variables	1598/2048 (78.03%)	0/2048 (0%)
P4	% observaciones de formularios mal asignados	1794/204 079 (0.88%)	0/204 079 (0%)
	% nacimientos con formularios mal asignados	164/2048 (8.01%)	0/2048 (0%)
P6	% variables categóricas afectadas por contextualización en sus valores	32/46 (69.57%)	0/46 (0%)
	% observaciones afectadas por contextualización en sus valores	13 392/204 079 (6.56%)	0/204 079 (0%)
	% nacimientos afectados por contextualización en sus valores	1487/2048 (72.61%)	0/2048 (0%)
	% variables categóricas afectadas por estabilidad temporal	9/46 (19.57%)	0/46 (0%)
	% observaciones con dominio de valores variable en el tiempo	1850/204 079 (0.91%)	0/204 079 (0%)
	% nacimientos con dominio de valores variable en el tiempo	649/2048 (31.69%)	0/2048 (0%)
	% observaciones afectadas por contextualización o estabilidad temporal	15 242/204 079 (7.47%)	0/204 079 (0%)
	% nacimientos afectados por contextualización o estabilidad temporal	2048/2048 (100%)	0/2048 (0%)

### Fases IV y V. Abordaje de los criterios de integridad semántica de los datos a nivel univariante y elaboración de la estructura canónica del repositorio

Tabla 3.5 Evaluación comparativa de DQ antes y después de cada procedimiento (P8-P12)

ID	Métrica	Evaluación de DQ antes del procedimiento	Evaluación de DQ después del procedimiento
P8	% variables numéricas con valores fuera de rango	3/12 (25%)	0/12 (0%)
	% observaciones con valores fuera de rango	6/204 079 (0.003%)	0/204 073 (0%)
	% nacimientos con valores fuera de rango	5/2048 (0.24%)	0/2048 (0%)
	% variables numéricas con valores en rangos poco probables	1/12 (8.33%)	0/12 (0%)
	% observaciones con valores en rangos poco probables	2/204 079 (0.001%)	0/204 073 (0%)
	% nacimientos con valores en rangos poco probables	1/2048 (0.04%)	0/2048 (0%)
P9	% nacimientos incompletos	123/2048 (6.01%)	0/1925 (0%)
	% observaciones de nacimientos incompletos	7384/204 073 (3.62%)	0/196 686 (0%)
P10	% observaciones duplicadas	88 415/196 686 (44.95%)	0/108 271 (0%)
	% duplicados inter-formulario (sistemáticos) con respecto al total de observaciones duplicadas	82 071/88 415 (92.82%)	
	% duplicados intra-formulario (no sistemáticos) con respecto al total de observaciones duplicadas	6344/88 415 (7.18%)	
	% nacimientos duplicados	0/1925 (0%)	0/1925 (0%)
P11	% observaciones únicas con variabilidad	742/108 271 (0.69%)	0/107 529 (0%)
	% variabilidad intra-formulario con respecto al total de observaciones con variabilidad	235/742 (31.67%)	
	% variabilidad intra-formulario con respecto al total de observaciones con variabilidad	507/742 (68.33%)	
	% nacimientos afectados por variabilidad	315/1925 (16.36%)	0/1925 (0%)

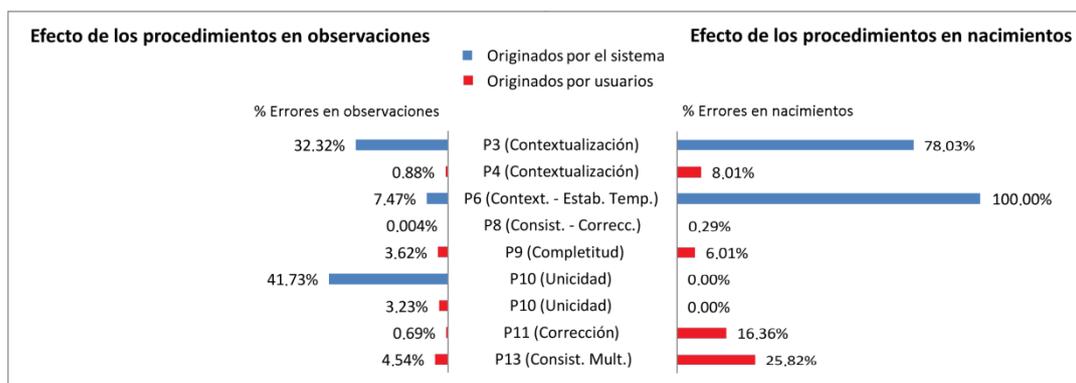
### Fase VI. Evaluación de la integridad semántica multivariante de los nacimientos

Tabla 3.6 Evaluación comparativa de DQ antes y después del procedimiento P13

ID	Métrica	Evaluación de DQ antes del procedimiento	Evaluación de DQ después del procedimiento
P13	% n-tuplas inconsistentes	831/15 400 (5.40%)	421/15 400 (2.73%)
	% nacimientos con alguna inconsistencia en sus conceptos clínicos	497/1925 (25.82%)	207/1925 (10.75%)
	% observaciones afectadas por inconsistencia multivariante (perteneciente a n-tuplas inconsistentes)	4877/107 529 (4.54%)	2414/107 529 (2.30%)

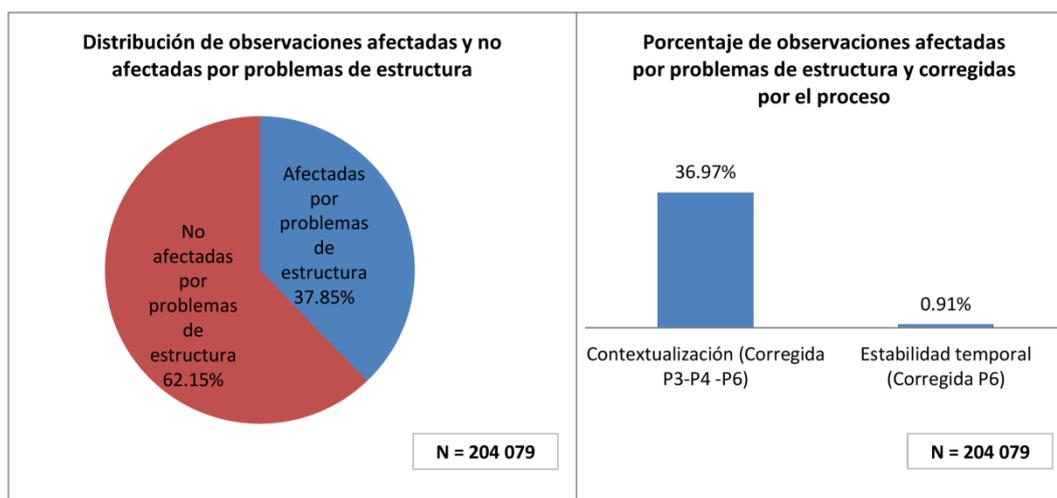
## Resultados globales

Una vez asegurada la DQ, el repositorio final de nuestro caso de estudio en el período perinatal 2009 - 2011 consta de 1925 nacimientos, de los cuales el 74.18% obtuvo la puntuación global de calidad 1 (DQ excelente), el 15.06% obtuvo la puntuación 2 (con problemas de calidad tolerables) y el 10.75% obtuvo la puntuación 3 (con inconsistencias 'no toleradas').



**Figura 3.2 Efecto de cada procedimiento en observaciones y nacimientos. Para cada procedimiento, las barras azules representan el porcentaje de errores originados por el sistema y las barras rojas representan los errores originados por usuarios.**

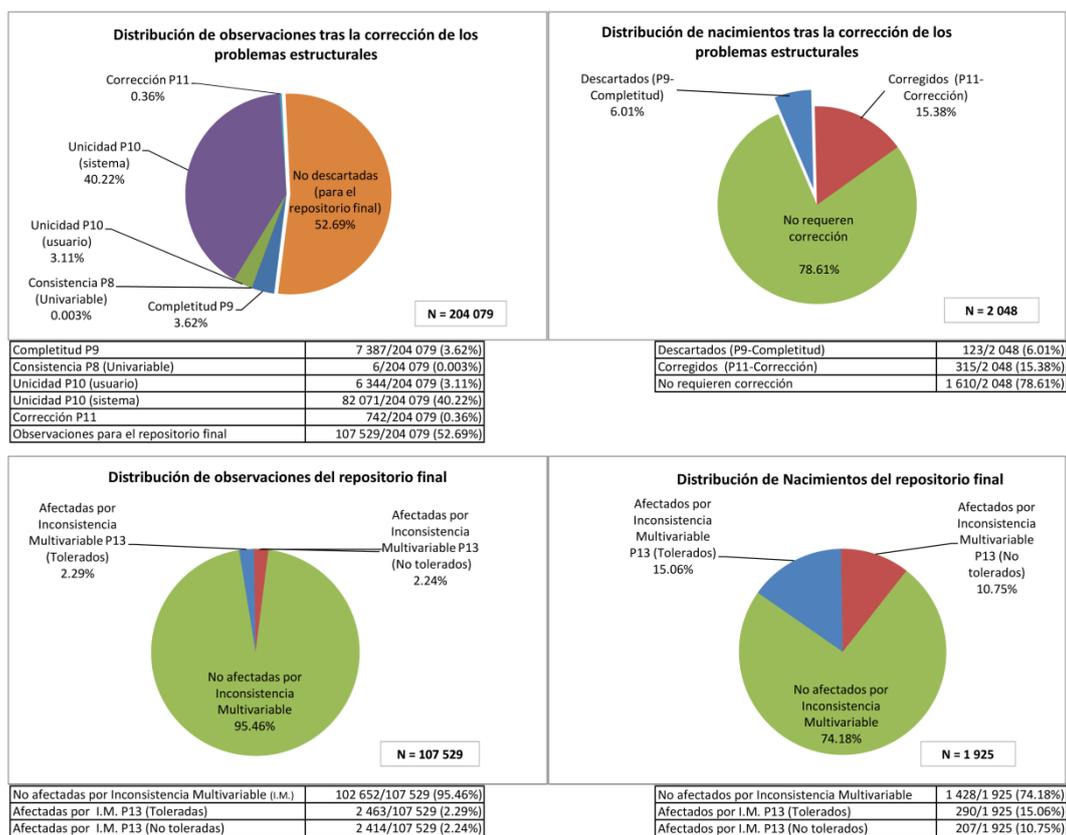
Las Figuras 3.2 - 3.4 resumen los resultados de las fases II-VI, una vez descartadas las observaciones de las variables no informativas (52.90% en la fase I). En concreto, la Figura 3.2 representa el efecto de cada procedimiento en la detección de errores en observaciones y nacimientos, diferenciando los procedimientos que tratan los problemas originados por el propio sistema y los problemas originados por los usuarios. Los procedimientos identificados que abordan problemas atribuibles al sistema de HCE son P3 y P6, así como los duplicados inter-formulario del P10, debido a la implementación del flujo de trabajo de los formularios de la HCE perinatal. El resto de los procedimientos abordan problemas asociados a la introducción de datos por parte de los usuarios. La mayoría de los valores de la Figura 3.2 se obtienen de las Tablas 3.3 - 3.6. Además, los valores que se muestran en P8 son la suma de los dos tipos de rangos de control (fuera de rango y rangos poco probables) incluidos en la Tabla 3.4. Además, el efecto de P10 se dibuja dos veces para diferenciar el efecto de duplicados sistemáticos, 41.73% (82 071 duplicados inter-formulario respecto a las 196 686 observaciones totales antes del procedimiento), de los duplicados por causa de usuarios 3.23% (6344 duplicados inter-formulario respecto a las 196 686 observaciones totales antes del procedimiento).



**Figura 3.3. A la izquierda, observaciones afectadas y no afectadas por problemas estructurales. A la derecha, el porcentaje de observaciones afectadas por problemas estructurales corregidos durante el proceso.**

La Figura 3.3 muestra la distribución de observaciones afectadas y no afectadas por problemas estructurales abordados en las fases II y III. Además, muestra los porcentajes de observaciones corregidas para las dimensiones de DQ de contextualización y estabilidad temporal. Los resultados presentados en la Figura 3.3 se calculan a partir de las observaciones con problemas estructurales (65 967 de P3, 1794 de P4 y 15 242 de P6) con respecto al total de observaciones a lo largo de estos procedimientos (204 079). Vale la pena tener en cuenta que algunas de estas observaciones son tratadas por más de un procedimiento, por lo que el total de observaciones diferentes afectadas por problemas estructurales es de 77 249 (37.85%). Esta superposición también afecta los resultados del panel derecho de la figura, por lo que las observaciones corregidas son 75 454 (36.97%) para la dimensión de contextualización.

En la parte superior, la Figura 3.4 presenta las observaciones y nacimientos afectados a nivel univariante diferenciados por dimensión de DQ (fases IV y V), una vez resueltos los problemas estructurales. A la izquierda, presentamos la distribución de las observaciones descartadas en comparación con las no descartadas (incluidas en el repositorio final). Las observaciones descartadas se diferencian por dimensión DQ, lo que también permite identificar el efecto de de-duplicación (Duplicación P10 y Corrección P11). A la derecha, presentamos la misma información, pero en términos de nacimientos. En la parte inferior, la Figura 4 muestra las distribuciones de observaciones (a la izquierda) y nacimientos (a la derecha) del repositorio final una vez que se corrige la estructura y se abordan los criterios de integridad univariante. La distribución representa las observaciones y los nacimientos afectados por la inconsistencia multivariante, tolerada y no tolerada, que se abordan en el último procedimiento.



**Figura 3.4. Distribución de observaciones y nacimientos descartados y corregidos por DQ. Tolerancia de la inconsistencia multivariante del repositorio final.**

La Figura 3.5 amplía los resultados de inconsistencia al nivel de conceptos clínicos. A la izquierda, la línea azul representa el porcentaje de nacimientos afectados por inconsistencia multivariante (tanto tolerada como no tolerada) en 1, 2, 3, 4 o 5 conceptos clínicos, mientras que la línea roja muestra el porcentaje de nacimientos afectados por inconsistencia multivariante no tolerada. A la derecha, presentamos el análisis individualizado de inconsistencia multivariante por conceptos clínicos, de tal modo que para cada concepto clínico, mostramos el porcentaje de nacimientos con inconsistencias en ese concepto, tanto toleradas como no toleradas.

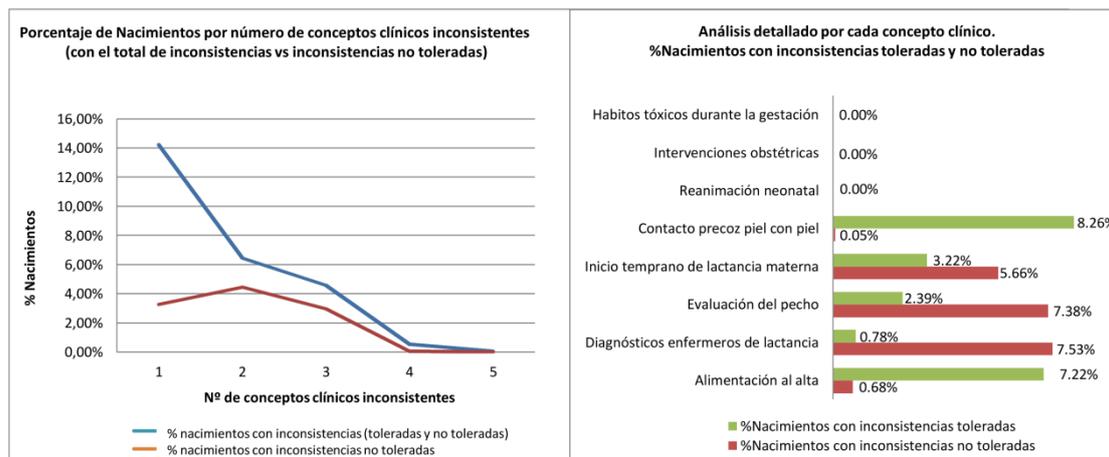


Figura 3.5. Porcentaje de nacimientos por número de conceptos clínicos inconsistentes. Análisis detallado de conceptos clínicos.

### 3.4 Discusión

#### Significancia

En nuestra búsqueda bibliográfica, no hemos encontrado referencias previas a indicadores poblacionales sobre alimentación infantil basados en HCE. Nuestro repositorio poblacional es el primer repositorio que incorpora los aspectos más relevantes del nacimiento del recién nacido y el proceso de alimentación. Por tanto, puede constituir un conocimiento relevante para la epidemiología, el control de calidad y la toma de decisiones en este área clínica específica. En particular, puede constituir conocimiento relevante para estudios poblacionales sobre importantes problemas de salud como las enfermedades inmunológicas en niños o la obesidad infantil (MelniK, 2014), que son problemas relacionados con el abandono de lactancia materna.

Este trabajo ha puesto de manifiesto el valor de los grupos multidisciplinares integrados por expertos clínicos y en tecnologías de la información para desarrollar proyectos a escala regional para la construcción de repositorios con calidad asegurada. En nuestra opinión, este diseño es la forma más realista de incorporar la HCE como fuente de datos de repositorios para su reutilización.

El repositorio presentado en este trabajo ha sido crucial para el seguimiento del PAAL en el área perinatal y detectar retrocesos en la implementación de la BFHI en nuestro hospital. Además, este repositorio con calidad asegurada se ha utilizado como un "sistema de aprendizaje de la salud", tal como lo propone el Instituto de Medicina de EE. UU. (Institute of Medicine (US), 2007) y otros, que utilizan datos recopilados de forma rutinaria para orientar la atención. En este sentido, la información obtenida del repositorio ha sido fundamental para guiar el proceso de aprendizaje en la mejora de la calidad de la organización.

Basado en los procedimientos de DQ implementados en el capítulo anterior (García-de-León-Chocano et al., 2015), para la creación de nuestro repositorio se requirieron los siguientes recursos: 1) 70 horas-persona para la definición y parametrización de las reglas de DQ junto con varias reuniones de consenso; 2) la homogeneización y estabilización de los datos brutos requirió 10 horas-persona. Esto fue optimizado usando los procedimientos

P3 y P6; y 3) el análisis semántico requirió 30 horas-persona. Vale la pena señalar que este esfuerzo depende del volumen de datos y la DQ inicial. Para futuras actualizaciones del repositorio, se requerirían los siguientes procedimientos y esfuerzo: 1) para corregir posibles cambios en la nueva estructura (dada la naturaleza cambiante de los sistemas de HCE) estimamos 5 horas-persona; y 2) para abordar nuevos casos de variabilidad e inconsistencia (que disminuye considerablemente con cada nueva extracción) estimamos no más de 10 horas-persona.

El enfoque de gestión de la calidad total aplicado para controlar el proceso de construcción ha permitido evaluar su configuración para la funcionalidad esperada y, por lo tanto, mejorar su diseño durante la construcción del repositorio. La parametrización del procedimiento P1 permitió configurar el diseño genérico del proceso de construcción a las necesidades específicas del escenario perinatal. Además, los puntos de corte establecidos durante la construcción del repositorio fueron útiles para incorporar las actividades de selección, análisis y corrección realizadas por el experto. Ambos aumentaron la calidad final del repositorio al tiempo que redujeron la cantidad de nacimientos descartados.

La medición de las dimensiones de DQ a través del proceso de carga del repositorio nos permitió cuantificar la calidad de su información. Tras el proceso de carga, en nuestro repositorio de alimentación infantil perinatal: 1) el 100% de las variables finalmente incorporadas tuvieron valor predictivo para el estudio, 2) el 100% de los nacimientos fueron contextualizados adecuadamente y temporalmente estables, 3) el 100% de las observaciones estuvieron en los rangos establecidos, 4) después del procedimiento P11 se corrigió el 100% de la variabilidad y 5) el 89.25% de los partos-nacimientos fueron consistentes con las condiciones de los expertos.

El descarte de variables no informativas (64.13%) en la fase I es alto debido a las peculiaridades del conjunto de datos perinatales, que incluye datos de la madre y el recién nacido. La mayoría de las variables descartadas pertenecen a la madre y no tienen un valor predictivo conocido para su reutilización. También se descartaron las variables del recién nacido sin valor predictivo. Este descarte representa una mejora considerable de la eficiencia de los procedimientos posteriores. A continuación, se abordó la integridad de la estructura (fases II y III) corrigiendo problemas de contextualización y estabilidad temporal que afectaban al 37.85% de las observaciones y al 100% de los nacimientos. Esto nos permitió obtener una estructura homogénea y estable sin cambiar el sentido clínico original que facilitó el análisis semántico realizado posteriormente por expertos clínicos (fases IV a VI).

En cuanto al análisis semántico de observaciones individuales, el porcentaje de errores de consistencia y corrección fue insignificante debido al rango de las variables cuantitativas. Además, dada la obligatoriedad de las variables que se identificaron como predictivas, solo el 6.01% de los nacimientos fueron incompletos y su análisis determinó que corresponden bien a mortinatos o recién nacidos trasladados a otro hospital (para los cuales no se completa todo el flujo de trabajo de formularios) y que no se corresponden con casos de estudio.

El porcentaje total de observaciones descartadas en la de-duplicación es alto (44.69%) debido al flujo de adquisición distribuida en la HCE, y es inherente a la implementación del proceso de atención perinatal en la HCE dado que muchas de las variables deben ser recogidas de diferentes puntos. Solo el 0.36% de las observaciones se vieron afectadas por la variabilidad y fueron resueltas, ya sea automáticamente (para la intra-formulario) o seleccionando el valor más fiable por el experto (para la inter-formulario). El resto de las observaciones descartadas en la de-duplicación corresponden a duplicados. De esta forma, la de-duplicación nos permite generar seguidamente la estructura canónica a través de una transformación simple pivotando sobre el `Id_Nacimiento` sin perder información relevante. Esta estructura cumple con los requisitos de presentación del repositorio final establecido en la Fase I, y facilita el abordaje del problema de inconsistencia multivariante desarrollado en la última fase.

Posteriormente, el análisis en dos pasos de la integridad semántica multivariante requiere entrenamiento y control de concordancia. En nuestro caso, fue realizado por un único ob-

servador entrenado (uno de los autores de este trabajo) detectando que el 25.82% de los nacimientos tenía alguna inconsistencia en sus conceptos clínicos. El 58.35% se recuperó utilizando la tolerancia establecida por el grupo de expertos, marcando los restantes como "no tolerados".

En general, se desprende de la Figura 3.2 que los procedimientos con mayor efecto sobre las observaciones y los nacimientos son los de corrección de problemas causados por el sistema, específicamente los relacionados con formularios y variables multicódigo (P3), recodificaciones (P6) y duplicaciones producidas por el flujo de trabajo de formularios (P10). Vale la pena señalar que P3 y P6 afectan directamente a casi todos los nacimientos debido a los cambios estructurales durante el período de estudio. Sin embargo, el efecto de los duplicados sistémicos corregidos por P10 que afectan al 41.73% de las observaciones no afecta a los nacimientos porque corresponden a nacimientos duplicados. Finalmente, en la Figura 5, el área entre la línea azul y la línea roja (nacimientos con inconsistencias y nacimientos con inconsistencias "no tolerados") representa la efectividad de la tolerancia en P13. Además, a la vista de esta figura, pudimos determinar el umbral para el número de conceptos clínicos inconsistentes pero tolerables para aceptar un nacimiento para el re-uso de datos en estudios clínicos.

El repositorio presentado en este artículo demuestra la validez y utilidad del proceso de construcción de repositorios de datos del PAAL de calidad asegurada, que se presentó en el capítulo anterior ([García-de-León-Chocano et al., 2015](#)).

## Limitaciones

El proceso de construcción del repositorio requiere que un grupo de expertos clínicos lleve a cabo un trabajo preliminar para establecer un conjunto completo de reglas, especialmente en el procedimiento P13. Además, el trabajo manual necesario para garantizar la integridad de la estructura (P3 y P6) puede ser realizado por un médico experto, así como por un profesional informático familiarizado con la implementación de la HCE. En cuanto a la cantidad de datos descartados: 1) el descarte de variables no informativas no tiene efecto sobre los sesgos de selección ya que estas variables no tienen valor predictivo conocido sobre el abandono prematuro de la lactancia materna según la evidencia actual, 2) es obvio que descartar observaciones debido a la de-duplicación no influye en el sesgo de selección, 3) ~ 6% de los nacimientos descartados en P9 no son casos de estudio y 4) el 10.75% de los nacimientos no tolerados es bastante bajo considerando la naturaleza poblacional del repositorio. Estos nacimientos no tolerados podrían causar un sesgo de selección, pero no se descartan del repositorio final; solo quedan marcados, lo que permite un análisis más preciso de los posibles sesgos de selección. Así, podemos concluir que la muestra final del estudio es representativa de la población objetivo.

## Trabajos futuros

Estamos ampliando el repositorio para incluir el período completo de PAAL con información sobre el episodio asistencial correspondiente al programa de niño sano hasta la edad de los dos años. Estamos definiendo los procesos de extracción de las fuentes de datos de atención primaria (el sistema OMI-AP en nuestra región) y especificando los requisitos de DQ del repositorio. El repositorio completo se construirá bajo el sistema de conceptos CONTSys, y los conceptos clínicos se definirán bajo el estándar UNE-EN 13 606, asegurando la interoperabilidad semántica y organizativa del repositorio. La información generada en este repositorio permitirá diseñar un sistema completo de monitorización del PAAL basado en datos poblacionales.

Además, dada la importancia de contar con repositorios de calidad asegurada sobre alimentación infantil, actualmente estamos trabajando en la ampliación del repositorio con nuevas fuentes de datos, por ejemplo, completándolo con datos sobre episodios mórbidos y consumo farmacéutico. Se está preparando una publicación con las conclusiones clínicas extraídas del repositorio de datos para complementar la metodología descrita en este trabajo. Además, los autores están liderando un proyecto nacional para desarrollar el modelo estandarizado de HCE para niños sanos desde el nacimiento hasta los dos años de edad, permitiendo la monitorización del PAAL a escala nacional utilizando métodos novedosos de monitorización de DQ (Sáez, Robles, et al., 2017), (Sáez et al., 2015).

Estos procedimientos se pueden aplicar a repositorios multicéntricos para la reutilización de información clínica. En este sentido, el modelo perinatal de HCE analizado en este estudio se está implantando en todos los hospitales de la Región de Murcia. Esto ampliará el alcance de este estudio a nivel regional.

### **3.5 Conclusiones**

En este estudio, hemos podido construir el primer repositorio con calidad asegurada para la reutilización de información de HCE perinatal sobre el nacimiento y la alimentación infantil que permitirá su explotación en estudios poblacionales y el desarrollo de estrategias clínicas. El proceso de construcción obtuvo un repositorio de datos con calidad de datos asegurada: con un 100% de variables informativas, donde la homogeneización y estabilización corrigió el 100% de la variabilidad, se de-duplicaron todas las observaciones necesarias y el 89.25% de los partos-nacimientos fueron consistentes con las condiciones de los expertos.



## 4 Integración de repositorios materno-infantiles basada en estándares de información sanitaria

En este capítulo se presenta la experiencia de integración, basada en estándares de información sanitaria, de repositorios materno-infantiles de dos hospitales: Hospital Virgen del Castillo del Area de Salud V – Altiplano y Hospital 12 de Octubre de Madrid. Partimos, de repositorios obtenidos a partir de datos de la HCE de ambos hospitales, sin evaluación previa de DQ. Se presenta una propuesta de integración mediante la construcción de un nuevo IDR utilizando estándares de información de salud, en concreto con el uso concurrente de las normas UNE-EN ISO 13 606 y UNE-EN ISO 13 940 – Consys, que permiten dar soporte a la continuidad asistencial. Se definieron y diseñaron modelos de información estandarizados de la atención perinatal y de alimentación hasta los dos años, basado en arquetipos UNE-EN ISO 13 606 para asegurar su interoperabilidad semántica y utilizando los conceptos de la norma UNE-EN ISO 1390 – Consys. A partir del modelo de información estandarizado evaluamos la calidad de datos y generamos el repositorio de datos integrados (IDR) con datos de gestación, nacimiento y alimentación de ambos hospitales (C6).

Como resultado adicional, los dos arquetipos UNE-EN ISO 13 606 que componen el modelo de información estandarizado se presentaron en el Ministerio de Sanidad, Servicios Sociales e Igualdad como propuesta para una integración a nivel nacional (C4).

Sobre los modelos de información estandarizados se definió un conjunto de indicadores de Buenas Prácticas para evaluar el grado de implementación de estrategias de salud materno-infantiles como la BFHI y monitorizar sus resultados.

El desarrollo de los modelos de información estandarizados e integración de los repositorios con calidad de datos asegurada estuvieron enmarcados dentro un proyecto piloto realizado para el Ministerio de Sanidad, Servicios Sociales e Igualdad (MSSSI): “Proyecto piloto para mejorar la calidad de la información en la atención perinatal, basada en buenas prácticas del Sistema Nacional de Salud (SNS), en el marco de la estrategia de atención al parto y salud reproductiva. Expediente nº 2015/07PN0010” (*P. Piloto MSSSI – 2015*), fruto de la colaboración de los hospitales Virgen del Castillo de Yecla, 12 de Octubre de Madrid y la empresa Veratech for Health, spin-off de la UPV.

*El contenido de este artículo es una adaptación para la tesis del artículo (Sáez, Moner, et al., 2017) – publicación Pu4. Como resultado, se construyó el IDR materno-infantil (gestación, nacimiento y alimentación hasta los dos años) de dos hospitales (Murcia y Madrid), basado en estándares de información sanitaria, contribución C3 y se presentó la “Propuesta de arquetipos UNE-EN ISO 13 606 al MSSSI para integración a nivel nacional”, contribución C4.*

## 4.1 Introducción

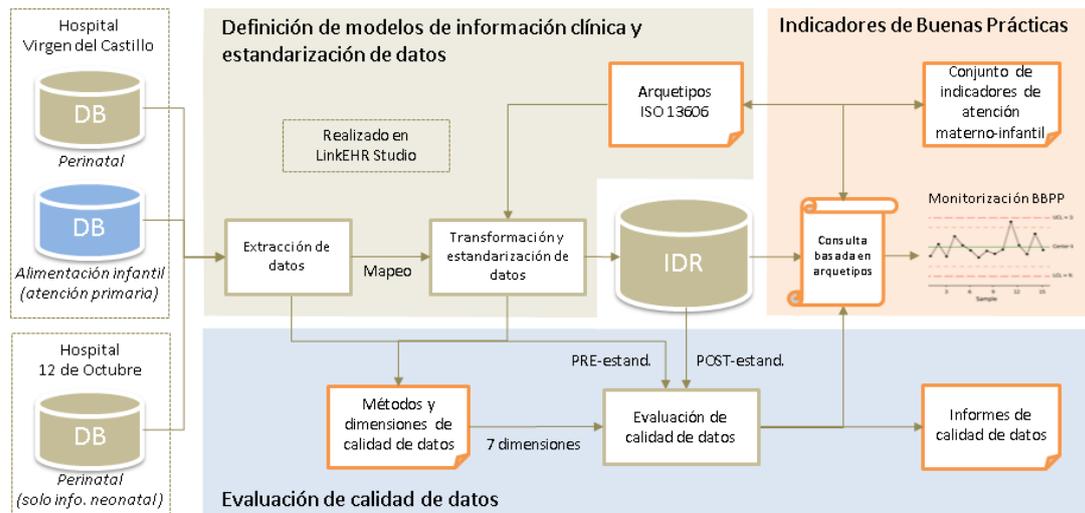
Los repositorios de datos integrados (IDR) se están convirtiendo en un recurso esencial que permite la reutilización de datos biomédicos a partir de grandes cantidades de datos de diversas fuentes (Toubiana & Cuggia, 2014). Se han llevado a cabo varias iniciativas en IDRs para proporcionar acceso a datos de investigación biomédica, ya sea como herramientas de consulta federadas (McMurry et al., 2013; Weber et al., 2011) o como repositorios centralizados (Murphy et al., 2010). En la mayoría de las soluciones, la adopción de un formato de datos común fue clave. Sin embargo, hasta donde sabemos, el uso de estándares específicos de información sanitaria fue limitado (Walker et al., 2014). Además, se acuerda que la fiabilidad de la reutilización de datos depende en gran medida de su DQ (Weiskopf & Weng, 2013). Ciertamente, la evaluación DQ se considera un componente clave para cualquier IDR (MacKenzie et al., 2012), donde se pueden encontrar algunos ejemplos exitosos en la literatura reciente (G. M. Kahn et al., 2012; Murphy et al., 2010).

Este capítulo se enmarca en un proyecto piloto del Ministerio de Sanidad, Servicios Sociales e Igualdad de España (Exp. nº 2015/07PN0010), destinado al desarrollo de un IDR nacional de información materno-infantil. El proyecto tuvo dos motivaciones principales. En primer lugar, la evaluación de las estrategias de salud materno-infantil (Observatorio de Salud de la Mujer y del Sistema Nacional de Salud. MSSSI, 2008; WHO, 1998) con el objetivo último de difundir buenas prácticas (BBPP). En segundo lugar, proporcionar un repositorio para la investigación poblacional, con un enfoque especial en la lactancia materna, como uno de los principales determinantes de la salud materno-infantil (WHO, 1998). Los objetivos específicos fueron: 1) Definir un modelo estandarizado de HCE de la infancia temprana (incluidas la gestación, nacimiento y seguimiento de la alimentación hasta los dos años), que integrase tanto los aspectos clínicos como los datos de contexto organizativo, 2) generar un repositorio de datos integrados con calidad asegurada para su utilización en la monitorización de buenas prácticas y 3) definir una propuesta de indicadores para la monitorización de la atención materno-infantil, los cuales sean medibles sobre el repositorio estandarizado. Se desarrolló un IDR como solución, que, basado en estándares de información de salud, aseguró una interfaz común para monitorizar BBPP de diferentes hospitales y regiones, y tener su DQ evaluada asegura una reutilización fiable de los datos.

## 4.2 Material y métodos

Para la realización del proyecto piloto se formó un grupo de trabajo multidisciplinar y participativo formado por expertos del dominio clínico, gestores y responsables de informática de los hospitales Virgen del Castillo y 12 de Octubre, así como expertos en tecnologías de la información del Grupo de IBIME, actual BDSLab, del Instituto ITACA de la UPV y de la empresa VeraTech for Health (spin-off de la UPV).

El proyecto piloto se dividió en tres tareas principales: (1) definición de modelos de información clínica y estandarización de datos, (2) definición de criterios y evaluación de DQ (PRE estandarización y POST estandarización) y (3) definición de una propuesta de indicadores de BBPP. El soporte clínico y de datos fue proporcionado por los dos hospitales participantes: Hospital Virgen del Castillo, Yecla y Hospital 12 de Octubre, Madrid. La figura 4.1 muestra la arquitectura de la solución propuesta.



**Figura 4.1** Arquitectura de la solución del repositorio de datos integrados (IDR) para una reutilización estandarizada y fiable de las fuentes de datos (DB) de atención materno-infantil

### Definición de modelos clínicos y estandarización de datos

Se decidió usar modelos de información estandarizados basados arquetipos UNE-EN ISO 13 606 (ISO, 2008, p. 13606) para proporcionar al IDR un modelo de información sobre la estructura de datos (cómo se organizan los datos) y las restricciones de datos que deben cumplirse (qué valores son válidos). Un modelo de información adecuado es fundamental para garantizar que la información clínica relevante esté disponible para los fines particulares de la reutilización de datos. Por lo tanto, la metodología de creación de arquetipos incluyó la identificación de las principales estructuras de datos clínicos, la selección y agrupación de elementos de datos relevantes para los dominios clínicos requeridos, la búsqueda de arquetipos reutilizables, la creación de nuevos arquetipos o la adaptación de los existentes y su validación por expertos clínicos.

En concreto, los modelos de información estandarizados se definieron de arriba abajo y de abajo a arriba de forma confluyente:

- En primer lugar, se analizaron los modelos de datos de las HCE actuales de ambos hospitales.
- A continuación se siguió un método sistemático, con búsqueda de especificaciones de datos estándares para la atención al recién nacido y seguimiento de la alimentación en documentos (de registros epidemiológicos, estrategias de atención neonatal, ...) o formalismos informáticos ya empleados en otros proyectos de desarrollo de HCE (por ejemplo, arquetipos ya existentes en la biblioteca pública openEHR Clinical Knowledge Manager, ontologías, ...).

Los modelos se representaron en primer lugar en una hoja de cálculo, lo cual facilitó el proceso inicial de discusión y la creación de la estructura básica de información. A continuación se desarrollaron arquetipos ISO 13 606 según la sintaxis ADL 1.4. Se utilizó la herramienta LinkEHR Studio (Maldonado et al., 2009) (www.linkehr.com) para el modelado de los arquetipos.

Por último, se utilizó la misma herramienta para mapear los arquetipos definidos con los datos de atención materno-infantil extraídos de los sistemas originales de HCE, para obtener documentos conformes a los arquetipos ISO 13 606 a partir de programas basados en XQuery. De este modo, el programa XQuery transforma los datos a partir de la HCE no estandarizada para generar documentos XML estandarizados según la norma UNE-EN

ISO 13 606 y conformes a la estructura de información definida en los arquetipos desarrollados (Figura 4.2).

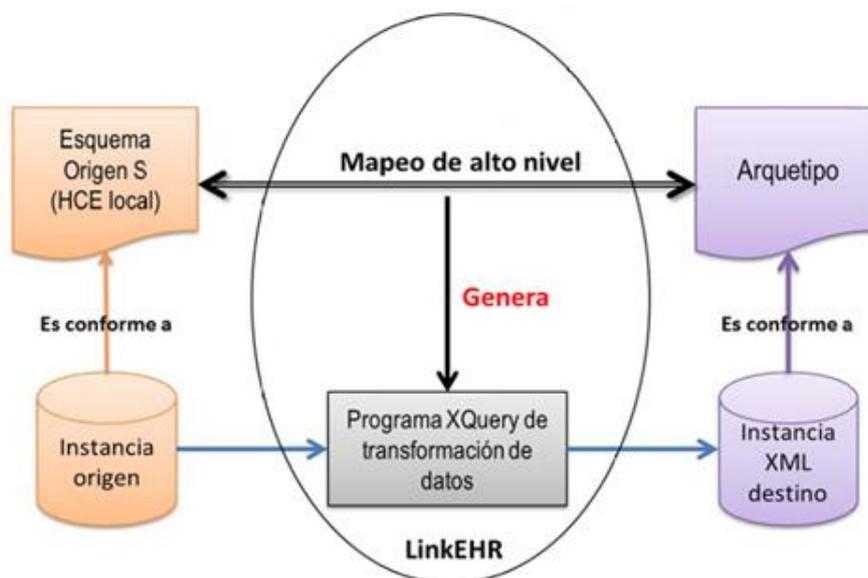


Figura 4.2 Metodología de generación de mapeos XQuery

De esta forma, fue posible la construcción de un repositorio de datos estandarizados conforme a la norma UNE-EN ISO13 606 y a los arquetipos definidos. La arquitectura se diseñó siguiendo las especificaciones de seguridad de la norma UNE-EN ISO 13 606, con disociación de datos demográficos y clínicos. El IDR se implementó en ExistDB y las consultas se definieron utilizando el lenguaje XQuery nativo.

### Evaluación de calidad de datos

Los criterios para medir la DQ de los repositorios y su posterior evaluación se definieron y llevaron a cabo con procedimientos específicos en base a 8 dimensiones de DQ (Sáez et al., 2012). La completitud (datos no faltantes), consistencia (conformidad con las restricciones y reglas del esquema) y unicidad (identificadores de instancias no duplicados) se calcularon de acuerdo con los requisitos de conformidad del arquetipo y en base a los estudios previos en el ámbito perinatal (García-de-León-Chocano et al., 2016) donde quedaron establecidos los criterios y estándares de calidad de datos para este ámbito, sin embargo el modelo de arquetipos con LinkEHR no permitió establecer criterios de integridad semántica de consistencia y completitud multivariantes. Por ello, la completitud se evaluó en base al número máximo de datos atómicos que puede contener una instancia mínima del arquetipo, ponderando elementos obligatorios y optativos y la consistencia como la proporción de restricciones univariantes sobre el número total de restricciones posibles. Por otro lado, la dimensión de contextualización se evaluó en función del porcentaje de información de contexto anotada en cada instancia sobre el máximo de datos de datos de contexto del arquetipo. La estabilidad temporal y espacial de datos (grado de concordancia de datos entre diferentes fuentes y en el tiempo) se evaluó en función de los métodos probabilísticos de DQ propuestos por (Sáez et al., 2015, 2016; Sáez, Robles, et al., 2017). Para la corrección de los datos se midió el número de registros posiblemente anómalos

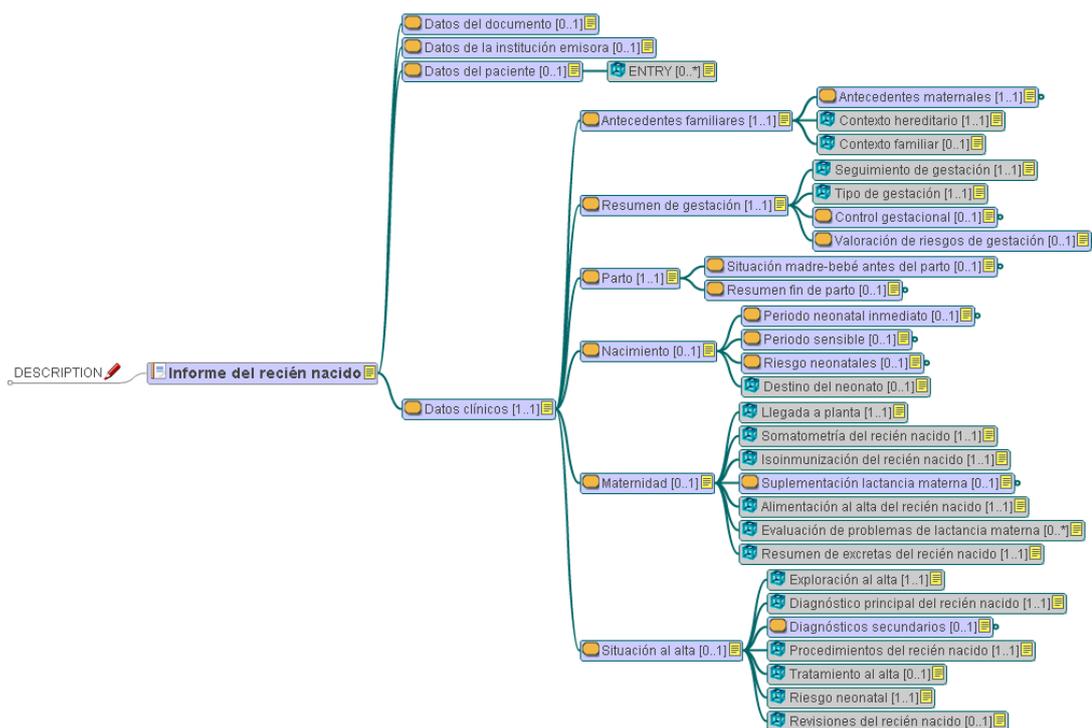
(valores atípicos multivariados), empleando una distribución de la población como referencia estándar proveniente de múltiples fuentes y comparando frente a ella los valores de tendencia central y variabilidad del conjunto de datos analizado, descubriendo los datos anómalos con diagramas de Box-Whisker. Finalmente, la dimensión de valor predictivo midió el potencial base del conjunto de datos para predecir la continuidad de la lactancia materna al mes usando el área bajo la curva ROC (AUC) de un modelo predictivo *Naïve Bayes* con una estimación de validación cruzada en 10 bloques, como una medida de la utilidad de los datos para esta tarea.

### **Definición de una propuesta de indicadores de BBPP**

Por último, la definición de una propuesta de indicadores de BBPP incluía dos sub-tareas. Primero, la definición formal de un conjunto de indicadores de BBPP basada en una extensa revisión de la literatura. Y segundo, la definición operativa de los indicadores y mapeo de sus variables a la información del arquetipo para su cálculo automático.

## **4.3 Resultados**

Se crearon dos arquetipos principales, uno para la gestación y nacimiento y otro para la alimentación del lactante, basados en la norma UNE-EN ISO 13 606 en sintaxis ADL 1.4, como base para 1) un informe de información de salud perinatal “Informe del recién nacido” (Figura 4.3), incluyendo información sobre antecedentes familiares, gestación, parto, nacimiento y maternidad y 2) un informe de la alimentación del lactante hasta los dos años, incluyendo información sobre lactancia materna e introducción de diferentes tipos de alimentos. Los dos arquetipos principales estaban compuestos subarquetipos re-utilizables, en total se definieron un total de 41 arquetipos. Los arquetipos pueden ser accedidos desde <http://mm.linkehr.com/>. Para su diseño, se reutilizaron 11 arquetipos pertenecientes al “Conjunto Mínimo de Datos de Informes Clínicos” (CMDIC) desarrollados por el MSSSI dentro del marco del proyecto de Historia Clínica Digital del Servicio Nacional de Salud. En la figura 4.3 se muestra a modo de ejemplo el arquetipo principal “Informe del recién nacido”. En la tabla B.1, tabla B.2 y figuras B.1 a B.8 del apéndice B se muestran el detalle de los arquetipos utilizados, arquetipos re-utilizados del CMDIC y la representación gráfica de los contenidos de los principales arquetipos desarrollados, respectivamente.



**Figura 4.3 Contenido del arquetipo del informe del recién nacido. Los términos de la figura se incluyen en inglés por ser una captura del programa original.**

Para poblar el IDR, los arquetipos se asignaron a las fuentes de datos de los hospitales y los datos se transformaron a instancias XML UNE-EN ISO 13 606 a través de programas XQuery definidos en LinKEHR. El Hospital Virgen del Castillo proporcionó datos de 3781 registros para el informe de recién nacidos y 2133 para el informe de alimentación infantil. El Hospital 12 de Octubre proporcionó registros de 1949 individuos para el informe del recién nacido, aunque solo de su base de datos neonatal. En ambos casos del periodo 2009 – 2014.

La DQ se evaluó en los datos del Hospital Virgen del Castillo y Hospital 12 de Octubre por separado para los conjuntos de datos perinatal y de seguimiento de la alimentación, con evaluaciones de los datos de HCE originales (PRE) y después de la estandarización en el IDR (POST). Se generó un informe de DQ, que incluye los siguientes hallazgos principales. Unicidad: 100%, no se encontraron identificadores de pacientes duplicados. Completitud: en POST, la completitud general fue baja debido a requisitos de información más estrictos definidos en los arquetipos UNE-EN ISO 13 606 con criterios de máximos. Particularmente, a pesar de la mayor completitud del conjunto de datos neonatales PRE del Hospital 12 de Octubre (77%), este llenó una pequeña parte del arquetipo perinatal más general en POST (8%), mientras que la completitud promedio del Hospital Virgen del Castillo permaneció más estable (56% a 52%), completando en mayor grado la información requerida por el arquetipo. Consistencia: en este proyecto piloto, debido a la gran cantidad de variables, no se incluyeron las comprobaciones de tipos de datos ni de rango, por lo que los resultados de consistencia se restringieron a la falta de conformidad con la cardinalidad de los datos (número de veces que debe aparecer un dato), obteniéndose mediciones altas en general, superiores al 95%. El valor de contextualización poblacional obtenido fue del 100%. Estabilidad temporal: el método advirtió sobre un pequeño número de registros mal fechados en el Hospital 12 de Octubre, apareciendo como un subgrupo temporal anómalo; los datos perinatales del Hospital Virgen del Castillo mostraron tres sub-

grupos temporales (que muestran datos no concordantes entre sus períodos) relacionados con dos cambios en el sistema de HCE original; los datos de alimentación infantil del Hospital Virgen del Castillo se mantuvieron estables a lo largo del tiempo. Estabilidad de múltiples fuentes: se encontró una baja estabilidad entre los datos perinatales del Hospital Virgen del Castillo y Hospital 12 de Octubre (0.08 sobre 1). Corrección: se encontró un promedio del 1% de registros atípicos en todos los conjuntos de datos. Valor predictivo: se obtuvo un AUC de 0,60 utilizando todas las variables del conjunto de datos perinatales del Hospital Virgen del Castillo. Seis registros no pasaron el proceso de estandarización debido a importantes problemas de calidad.

Finalmente, se definió un conjunto de 127 indicadores de BBPP, agrupados en seis categorías: indicadores centrales (contacto precoz, primera toma de lactancia materna, suplementación, problemas de lactancia, alimentación al alta y duración de la lactancia materna y alimentación complementaria), antecedentes maternos, condiciones obstétricas, entorno obstétrico, intervenciones obstétricas y estado de salud del bebé, que involucran diferentes procesos clínicos (desde la gestación hasta el seguimiento en Atención Primaria), según recomendaciones nacionales e internacionales (Euro-Peristat, OMS, UNICEF). Todas las variables en las definiciones operativas de indicadores se mapearon a los arquetipos. Se desarrolló un prototipo de sistema de monitorización de BBPP utilizando programas XQuery sobre los datos estandarizados.

## 4.4 Discusión

Los arquetipos definidos deben considerarse como una versión inicial, candidata a ser revisada por más profesionales, hacia un modelo clínico detallado armonizado. Las versiones posteriores también deben incorporar enlaces de terminología (por ejemplo, a SNO-MED-CT) para proporcionar la semántica de las estructuras de información y los valores de los datos, que no se tratan en este piloto. Mapear los vocabularios locales de los dos hospitales con el vocabulario controlado utilizado en los arquetipos fue una tarea intensa. El uso de terminologías estándar tanto en los arquetipos como en las fuentes de datos originales puede resolver este problema en el futuro.

La evaluación de DQ proporcionó conocimientos novedosos sobre el efecto en DQ de la estandarización de datos utilizando estándares de información de salud. Los requisitos de información más estrictos pueden mejorar la DQ en términos de integridad y consistencia, y mejorar la usabilidad de los datos dada su contextualización. Sin embargo, el cambio del espacio de variables dado por la estandarización debe ser considerado al comparar las medidas PRE y POST: los arquetipos definen un conjunto mayor de información en comparación con los datos existentes. En concreto, los resultados de completitud POST fueron bajos en ambos repositorios (alrededor del 50% en el Hospital Virgen del Castillo y 8% en el 12 de Octubre). El valor de completitud tan bajo para el 12 Octubre se debió a la fuente origen de datos, proveniente de un repositorio del Servicio Neonatal (recién nacidos ingresados), el cual tiene un enfoque orientado a la morbilidad, claramente distinto al modelado en los arquetipos. Este hecho, también explica la notable diferencia entre la completitud PRE y POST de dicho hospital.

Los resultados de consistencia son altos, debido principalmente a las pocas características evaluadas. La evaluación de la consistencia se puede mejorar, incluyendo las verificaciones de tipo y rango de datos, lo que puede respaldarse con el uso de terminologías. Así mismo, la limitación de la herramienta para implementar restricciones simultáneas sobre varios elementos del arquetipo, impidió definir criterios para evaluar la integridad semántica multivariante, tanto a nivel de consistencia como de completitud.

Los resultados de contextualización indicaron que toda la información de contexto fue proporcionada para todos los registros.

Los resultados de la estabilidad temporal pueden ser de suma utilidad para respaldar el mapeo de datos con representaciones variables a lo largo del tiempo. Los tres subgrupos temporales identificados en el Hospital Virgen del Castillo, motivados por dos cambios en el sistema de HCE original, coincidieron con los periodos de cambios de formularios y variables identificados por los métodos semi-automáticos empleados en el capítulo 3. Por tanto, se ha podido constatar que los métodos probabilísticos propuestos por (Sáez et al., 2015, 2016; Sáez, Robles, et al., 2017) obtienen resultados que permiten identificar la variabilidad debida a variables y formularios multicódigo, al igual que la producida en el dominio de las variables. Variabilidad que debe ser abordada para asegurar un conjunto de datos estable a lo largo del tiempo.

Por otro lado, la baja estabilidad encontrada entre los datos del Hospital Virgen del Castillo y los del 12 Octubre se debió también a sus diferentes poblaciones (unidades de maternidad vs neonatales) y completitudes derivadas de esta razón.

Respecto a la corrección, este primer repositorio obtenido podría servir como primera referencia para futuros cálculos de sesgos con nuevos repositorios de información perinatal.

La monitorización de BBPP basada en datos estandarizados permitiría a futuros hospitales enrolados en el proyecto obtener una monitorización instantánea de la BBPP, una comparación y una evaluación de DQ, simplemente proporcionando datos estandarizados equivalentes.

Se encontraron limitaciones con la herramienta ExistDB empleada para la implementación del IDR. Por tanto, se explorarán otras tecnologías de repositorio para respaldar las necesidades de consultas avanzadas y mejorar la eficiencia del IDR.

## 4.5 Conclusiones

La estandarización se ha mostrado útil para obtener un repositorio de datos integrado e interoperable de ambos hospitales. Los modelos de información estandarizados propuestos permiten la compartición de información materno-infantil de distintas organizaciones y ha sido posible definir un conjunto de indicadores a partir de los mismos, habilitando una monitorización común y, por tanto, comparable. De este modo, el proyecto piloto puede sentar las bases de un IDR nacional para una reutilización fiable de datos materno-infantiles y un seguimiento estandarizado de BBPP. Las lecciones aprendidas discutidas pueden facilitar la ampliación del proyecto en acciones nacionales e internacionales. El enfoque desarrollado puede reproducirse en otros ámbitos de la atención sanitaria.

Sin embargo, a nivel local de una misma organización sanitaria, las ventajas de la estandarización se contrarrestan con la sencillez de formatos planos más eficientes y versátiles para el análisis de datos. Por otro lado, la limitación actual de la herramienta de modelado de arquetipos para establecer restricciones simultaneas entre distintos elementos del arquetipo, supone un serio obstáculo para establecer criterios de integridad multivariante de alto nivel como los definidos en el capítulo 3 (García-de-León-Chocano et al., 2016) para el ámbito perinatal; criterios que será necesario ampliar para evaluar la integridad semántica de un repositorio completo de alimentación infantil hasta los dos años. Por estos motivos, se decidió prescindir de la estandarización UNE-EN ISO 13 606 para la construcción del IDR completo (desde el nacimiento hasta los dos años) del Área de Salud V – Altiplano que se abordará en el capítulo siguiente.

# 5 Ampliación del repositorio con el seguimiento de la alimentación hasta los dos años

En este capítulo se describe la aplicación del proceso de construcción de repositorios para ampliar la estructura y contenido del repositorio perinatal descrito en el capítulo 3, con la información del seguimiento de la alimentación del niño hasta los 2 años. Los datos se obtienen de las revisiones de salud infantil realizadas en atención primaria según el Programa de Atención al Niño y Adolescente de la Región de Murcia (PANA). De este modo, se obtuvo un repositorio de datos integrados (IDR, del inglés *Integrated Data Repository*) de alimentación infantil hasta los dos años (C5), que integra los datos de alimentación de los dos niveles asistenciales y, cubre el ámbito completo del proceso asistencial de atención a la alimentación infantil (PAAL) definido en el Área de Salud V – Altiplano.

También se describe la paquetización de los elementos software de dicho proceso (C6), para la obtención de una aplicación web que cubre todas las fases del proceso.

En el capítulo siguiente se abordará la estimación robusta de indicadores de alimentación infantil a partir del IDR del Área V, ampliando la evaluación y aseguramiento de DQ desde un enfoque longitudinal.

*El contenido de este capítulo está compuesto principalmente por un texto original para la tesis describiendo la ampliación del repositorio. Los apartados relativos a la paquetización son una adaptación y ampliación para la tesis del Trabajo Fin de Grado en Ingeniería Informática de la Universidad de Castilla-La Mancha “Homogeneización y estabilización de datos clínicos” co-dirigido por el autor de esta tesis. Como resultado, se construyó el IDR de alimentación infantil hasta los dos años de la Gerencia del Área de Salud V – Altiplano, contribución C5 y se paquetizó el software para la construcción de repositorios, contribución C6. Las contribuciones C5 y C6 fueron posibles gracias a los trabajos previos desarrollados en los proyectos de investigación HQTQ-1 y HQTQ-2.*

## 5.1 Introducción

En la actualidad, los datos de procedentes de HCE son una excelente fuente de datos candidata para la monitorización, investigación y vigilancia de la alimentación infantil debido a su cobertura poblacional y su continuidad en el tiempo (Ajetunmobi et al., 2014; American Academy of Pediatrics et al., 2008; Euro-Peristat project with SCPE and Eurocat, 2010; Halvorsen et al., 2015; Nickel et al., 2017; Paul et al., 2019; Weng et al., 2012). Para la evaluación de la atención a la alimentación infantil desde el nacimiento hasta los dos años, es preciso integrar las fuentes de datos de ambos niveles asistenciales (atención especializada y atención primaria).

En este escenario, los repositorios de datos integrados (IDR, del inglés *Integrated Data Repository*) permiten la reutilización de datos biomédicos a partir de grandes cantidades de datos de diversas fuentes de HCE (Toubiana & Cuggia, 2014) y, de este modo, pueden servir de base para la evaluación y seguimiento de Procesos Asistenciales Integrados (PAI) que requieran de la integración de distintas fuentes. Para este propósito, se debe realizar una evaluación de la DQ para garantizar la validez y fiabilidad de la reutilización de

los datos (Khare et al., 2017; Reimer et al., 2016; Verheij et al., 2018; Weiskopf & Weng, 2013; Weng et al., 2012).

En el Área de Salud V – Altiplano, tras la acreditación del Hospital Virgen del Castillo como hospital BFHI (García-de-León-González et al., 2011), se definió la experiencia HQTQ como una intervención de mejora de la calidad asistencial en la atención a la alimentación infantil para implementar las recomendaciones de la BFHI en toda el área de salud, con objeto de mejorar la calidad de la atención materno-infantil durante los primeros 1000 días de vida. Para su implementación, se definió el proceso de atención a la alimentación al lactante (PAAL) (García-De-León-González et al., 2008), garantizando la continuidad asistencial entre el hospital y atención primaria.

Con objeto de poder evaluar el PAAL, tanto a nivel de proceso como de resultados, era necesario ampliar el repositorio perinatal inicial creado en el hospital (García-de-León-Chocano et al., 2016), integrando en dicho repositorio la información de alimentación infantil hasta los dos años proveniente de las revisiones de salud infantil de atención primaria realizadas según el Programa de Atención al Niño y Adolescente de la Región de Murcia (PANA).

El objetivo entonces era aplicar el proceso de construcción de repositorios con calidad asegurada (García-de-León-Chocano et al., 2015), utilizado previamente para construir el repositorio perinatal (García-de-León-Chocano et al., 2016), a la información del seguimiento de la alimentación hasta los 2 años, con motivo de obtener un IDR que recoja e integre la información necesaria de los dos niveles asistenciales implicados en el PAAL; con el fin último de constituir el IDR como fuente para la monitorización del PAAL. El uso de los conceptos de la norma UNE-EN ISO 13 940 (Consys) permite dar soporte a la continuidad asistencial a nivel organizativo.

Paralelamente a la aplicación del proceso de construcción, se desarrolló la paquetización de los elementos software del proceso con el objetivo de facilitar su aplicabilidad y uso por personal sanitario no técnico.

El objetivo de la paquetización era el desarrollo de una aplicación web para cubrir todas las fases del proceso de construcción de repositorios con calidad asegurada (García-de-León-Chocano et al., 2015), simplificando su aplicación y uso por personal clínico, no técnico. La aplicación, denominada *ProDQA*, debía aportar herramientas e interfaces para la aplicación del proceso de construcción a los distintos ámbitos clínicos. En concreto: la recolección rutinaria de información procedente de las distintas fuentes de HCE, la anonimización de datos, los interfaces para la parametrización del proceso y la implementación de los procedimientos y métricas de DQ, tanto los relativos al aseguramiento de la integridad de la estructura de los datos (homogeneización y estabilización para obtener la estructura canónica), como los relativos a la integridad semántica de los datos. La aplicación debía generar el repositorio final con calidad de datos asegurada y los repositorios intermedios necesarios.

## 5.2 Metodología

A partir del repositorio perinatal descrito en el capítulo 3, que contiene los datos del episodio perinatal (atención especializada), hemos ampliado su estructura y contenido aplicando el proceso de construcción de repositorios con calidad asegurada descrito en el capítulo 2 (García-de-León-Chocano et al., 2015) sobre los datos de seguimiento de las revisiones de salud infantil del PANA hasta los dos años (atención primaria). El nuevo IDR debe integrar la información de ambos niveles asistenciales, cubriendo de este modo el ámbito completo del PAAL, según su modelo de información definido en la figura 2.1. y

utilizando los conceptos del estándar de información sanitaria UNE-EN ISO 13 940 (CONTsys), según las correspondencias con el PAAL definidas en la tabla 2.1.

Como punto de partida, se ampliaron las especificaciones y requisitos del repositorio dadas las necesidades de información para la monitorización completa del PAAL y la investigación sobre la alimentación infantil hasta los dos años. De acuerdo con estas necesidades, el episodio asistencial del seguimiento del PANA hasta los dos años se definió como un nuevo Hecho de Estudio (HdE) del repositorio. El nuevo período de estudio fue del 1 de enero de 2009 al 31 de diciembre de 2018.

### 5.2.1 Materiales

Partimos del repositorio perinatal existente, con calidad de datos asegurada. Consta con un total de 5858 sujetos con toda su información perinatal tras la carga rutinaria de los episodios perinatales del HIS del hospital (Selene) relativas a la ampliación del periodo de estudio.

Respecto a la información del episodio del seguimiento del niño hasta los dos años, en los 3 centros de salud del Área V de Salud – Altiplano se registra en el Sistema de Información de Atención Primaria, OMI-AP. El software OMI-AP contiene formularios clínicos, peticiones e informes, entre otros elementos. Los datos estructurados de los formularios clínicos incluyen toda la información necesaria para su reutilización. Cada registro de formulario contiene las observaciones y la información contextual de las variables clínicas registradas. La información contextual incluye la identidad del usuario, la marca de tiempo de la observación y el identificador del proceso asistencial del paciente, entre otra información.

#### Conjunto de datos inicial

La nueva entrada para el proceso de construcción del repositorio es una tabla única de “raw data” o datos brutos sin procesar, que corresponde al conjunto de datos extraído de la base de datos Oracle del sistema de información de atención primaria OMI-AP y cargado en una tabla de SQL Server utilizando la herramienta de integración de datos Pentaho. Cada fila representa una observación de una variable clínica del episodio de seguimiento de la alimentación junto con su información contextual. Los datos brutos corresponden a 6427 niños (identificados por el Id\_Sujeto) con un total de 58 368 contactos de seguimiento (revisiones de salud infantil) para el período 2009 - 2018, con un total de 1 881 732 observaciones de 109 variables clínicas, recopiladas de los 17 formularios diferentes utilizados para las revisiones hasta los 2 años del PANA durante el periodo de estudio. Las variables son categóricas, numéricas y de texto corto. En resumen, el conjunto de datos inicial son los datos brutos formados por 1 881 732 filas donde cada fila incluye la siguiente información:

- El código de la variable.
- El valor de la observación.
- El Id\_Sujeto y otra información del sujeto de la asistencia.
- El Id\_Episodio y otra información del episodio asistencial del PANA.
- El Id\_Contacto y otra información del contacto (revisión de salud infantil del PANA).
- La identificación (Id\_Usuario) del usuario (profesional sanitario) que registró la información

- Rol del usuario que registró la información (enfermera de atención primaria o médico de atención primaria)
- La fecha y hora a la que se registró la información.
- El código del formulario en el cual se registró la información.

El identificador del sujeto (Id\_Sujeto) en el sistema de información de atención primaria OMI-AP es común al identificador del recién nacido ya existente en repositorio perinatal (proveniente del HIS Selene).

### 5.2.2 Métodos

Siguiendo la misma metodología utilizada para la construcción del repositorio perinatal, se aplicaron los procedimientos para la construcción de repositorios descritos el capítulo 2 (García-de-León-Chocano et al., 2015), utilizando métricas similares adaptadas al nuevo conjunto de datos. En concreto, se aplicaron las siguientes fases de la figura 2.3: Fase I de determinación de parámetros globales y descarte de variables no informativas, fases II y III para el aseguramiento de la integridad estructural de los datos, así como la fase V de resolución de la de-duplicación de datos y obtención de la estructura canónica, la cual se automatizó completamente y se adelantó a la fase IV. Tras la aplicación de este bloque de fases, quedaron evaluadas y aseguradas las dimensiones de DQ de valor predictivo, contextualización, estabilidad temporal, estabilidad espacial, unicidad y corrección.

Las fases y procedimientos para el aseguramiento de la integridad semántica, tanto univariante (fase IV) como multivariante (fase VI) se describen posteriormente en el capítulo 6 desde un enfoque centrado en la estimación robusta de indicadores de alimentación infantil y aplicadas sobre el IDR completo.

Para el episodio de seguimiento, los parámetros de estudio de la fase I fueron especificados por el grupo de expertos, guiado por el propósito de la reutilización: la monitorización completa del PAAL y la investigación de cuestiones relacionadas con la alimentación infantil hasta los dos años, específicamente:

- El mapeo entre los conceptos organizacionales de CONTsys y OMI-AP se describe en la Tabla 5.1:

**Tabla 5.1 Mapeo entre los conceptos organizativos de la norma CONTsys y el sistema de información de atención primaria OMI-AP**

Conceptos de la Contsys	Conceptos de OMI-AP	Equivalencia en el caso de estudio
Sujeto de la asistencia	Paciente	Niño
Episodio asistencial	Episodio	Episodio asistencial de seguimiento del niño hasta los dos años
Contacto	Visita	visita a los 15 días, visita a los 2 meses, 4 meses, 6 meses, entre los 9-12 meses, 15 meses, 18 meses y 2 años

- El HdE se configuró inicialmente para el episodio de asistencial correspondiente al episodio de seguimiento, identificado por el Id\_Episodio. Sin embargo, al ser un episodio único por Sujeto de la Asistencial, al igual que ocurría con el HdE perinatal, el HdE del IDR completo se configuró al Sujeto de la Asistencia (identificado de forma homogénea en todo el IDR por Id\_Sujeto)

- No se incluyeron los formularios de la revisión de los 2 años al contrastar que se utilizaba para revisiones de niños a partir de 2 años de edad, fuera del ámbito de estudio.
- Las variables informativas se identificaron a partir de revisiones de evidencia sobre las variables necesarias para calcular tasas y duración de lactancia materna, además de los factores que determinan dicha duración. La tabla A.2 del apéndice A enumera el subconjunto de variables identificadas como variables informativas del episodio asistencial de seguimiento. La tabla incluye para cada variable: su descripción y el concepto clínico al que pertenece la variable.
- La selección de variables informativas no incluyó variables de texto sin categorizar por lo que no fue necesario implementar el proceso P7 de categorización.

Adicionalmente, respecto a la implementación de los procedimientos para el caso perinatal, en la ampliación para el episodio de seguimiento del niño hasta los dos años: se automatizaron los procesos P10 y P11 de la fase V, seleccionando la última toma tanto para resolver la unicidad como para la variabilidad. Por otro lado, al no existir dos tipos de sujetos diferenciados no fue necesario aplicar el procedimiento P4 – Detección y corrección de formularios mal asignados

#### 5.2.2.1 Paquetización del proceso de construcción

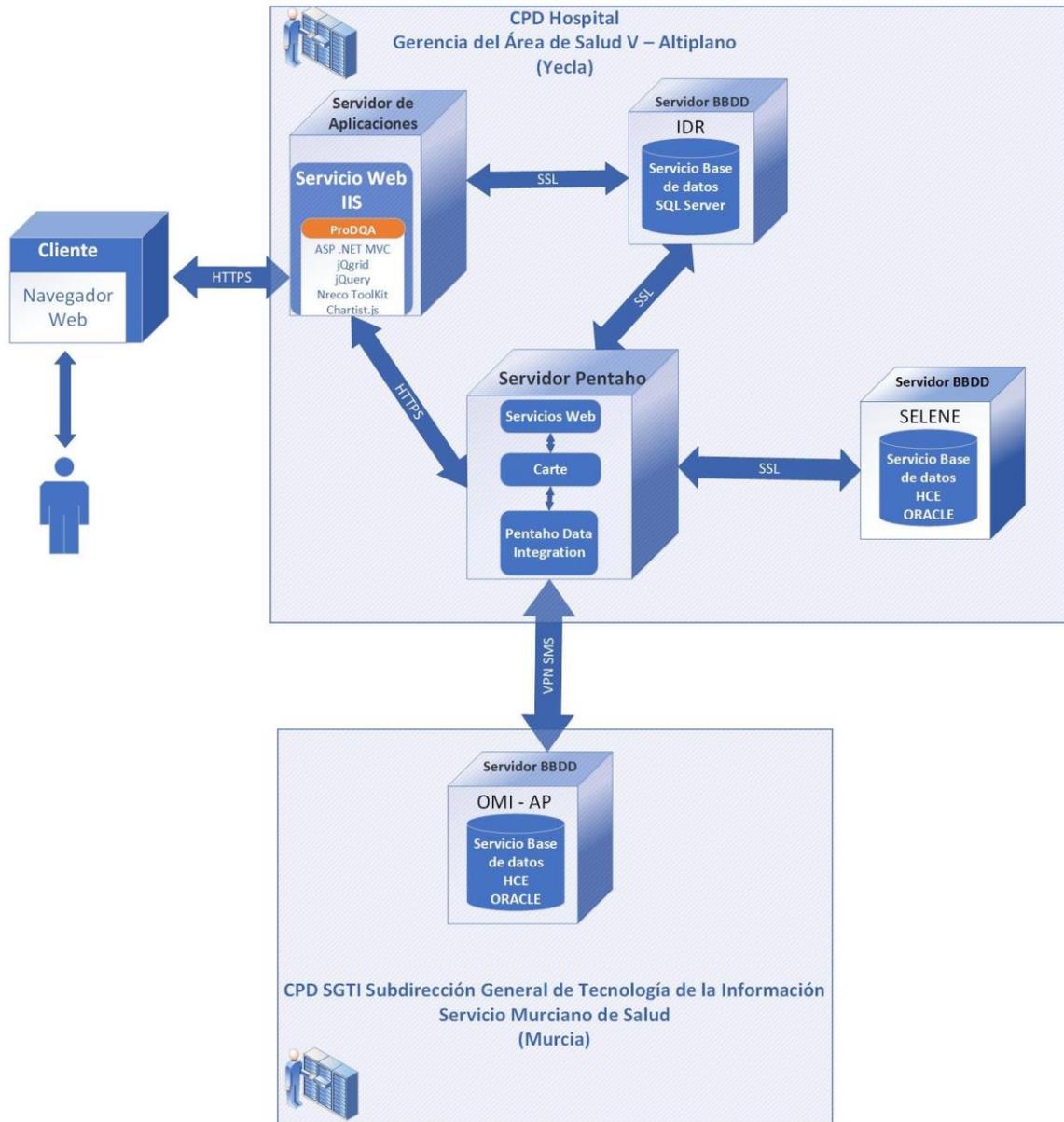
La planificación, desarrollo y mantenimiento de la aplicación web desarrollada para paquetizar el proceso de construcción se realizó en base a la metodología Métrica V3.

Los requerimientos funcionales de la aplicación web eran los establecidos en el propio proceso de construcción de repositorios con calidad asegurada (García-de-León-Chocano et al., 2015).

Respecto a requerimientos no funcionales, la usabilidad debía estar adaptada a uno de los roles principales de uso de la aplicación (rol clínico), tanto en facilidad de uso como en aportación de controles de protección a errores de usuario. Por otro lado, los requerimientos de eficiencia no fueron estrictos puesto que los principales procesos son *batch*. Por último, la seguridad debía incorporar la gestión de usuarios necesaria para un control de accesos por roles.

Respecto a la arquitectura, la aplicación web se basó en tecnología .Net de Microsoft. Específicamente se empleó Asp.net MVC como tecnología principal por su escalabilidad, flexibilidad y facilidad de mantenimiento, así como su facilidad para integrar múltiples componentes. En concreto, se añadieron los siguientes componentes a la aplicación principal: 1) Pentaho Data Integration para las tareas de Extracción, Transformación y Carga (ETL), integrado con la aplicación principal mediante servicios web, 2) JqGrid para la representación de datos tabulares, 3) Nreco olap library como componente para la gestión de datos multidimensionales en memoria, 4) Nreco bi toolkit para representar tablas dinámicas en HTML con funciones avanzadas de analítica y reporting en tiempo real, y 5) Chartist.js para visualización de gráficas interactivas.

A continuación, se presenta en la figura 5.1 el diagrama de despliegue para ilustrar la representación de la arquitectura. Los distintos componentes del sistema están representados por nodos, entendiendo por nodo cada partición física o parte significativa del sistema con características propias de ejecución o función.



**Figura 5.1** Arquitectura de la aplicación web. Se presentan los distintos nodos relacionados con la aplicación: Servidor de aplicaciones web implementado en Internet Information Server (IIS), servidor de extracción, transformación y carga implementado en Pentaho y sus conexiones con los servidores de bases de datos (BBDD) origen y el repositorio de datos integrado (IDR). Los nodos se presentan agrupados por su ubicación en los distintos Centros de Proceso de Datos (CPD).

Tal y como se puede observar en la figura, la aplicación web ProDQA se ejecuta en un nodo de servidor de aplicaciones implementado en Microsoft Internet Information Server (IIS). Este nodo aloja las capas de interfaz, lógica y persistencia de la aplicación que rige el sistema. Utiliza servicios web para conectarse con el nodo de servidor de Pentaho Data Integration donde se ejecutan las extracciones, transformaciones y cargas de datos. Existen 3 nodos de servidor de base de datos (BBDD), uno para el IDR, implementado en Microsoft SQL Server y dos nodos Oracle con las fuentes de HCE de atención especializada (Selene) y atención primaria (OMI-AP), respectivamente. Los dos primeros nodos de BBDD se encuentran en el propio Centro de Proceso de Datos (CPD) del Hospital Virgen del Castillo. El servidor de BBDD de OMI-AP se encuentra alojado en el CPD de la Subdi-

rección General de Tecnologías de la Información ubicado en Murcia. Por este motivo, las conexiones del servidor Pentaho con el servidor de BBDD de OMI-AP se realiza a través de una red privada virtual (VPN, del inglés, *Virtual Private Network*).

### 5.3 Resultados

El resultado de la aplicación de las fases descritas (I, II, III y V) del proceso de construcción de repositorios, para ampliar la estructura y contenido del repositorio perinatal con la información de seguimiento del niño en atención primaria, es un IDR con la calidad de la estructura asegurada, presentado en una estructura canónica que consta de una fila por Sujeto de la asistencia (identificado por el Id\_Sujeto). Cada fila incluye las variables contextuales del sujeto, tanto del episodio asistencial perinatal como del episodio asistencial de seguimiento, así como las variables clínicas informativas de ambos episodios desplegadas en columnas y agrupadas por los conceptos clínicos, relativas a los estados de salud del recién nacido, las actividades de atención de salud perinatal del PAAL y la información de seguimiento de alimentación en atención primaria. Además, cada fila incluye la toda la meta-información de DQ del episodio perinatal descrita en el capítulo 3, más la meta-información de DQ relativa a las fases de I, II, III y V del episodio de seguimiento.

A partir del repositorio perinatal existente con toda la información perinatal de 5858 recién nacidos, la aplicación del proceso de construcción de repositorios permitió generar el nuevo IDR donde se integró la información de seguimiento de 6427 niños seguidos en uno de los centros de salud del Área de Salud V, consistente en 58 368 contactos de revisiones de salud infantil del PANA realizadas en atención primaria.

A continuación presentamos los resultados de la evaluación de DQ obtenidos durante el proceso de construcción, relativos al nuevo conjunto de datos procedentes del episodio asistencial de seguimiento en atención primaria, agrupados por fases (tablas 5.2, 5.3 y 5.4). Para simplificar la comparación de resultados de diferentes procedimientos, los resultados del procedimiento P6 incluidos en la tabla 5.3 se calcularon como un porcentaje sobre el número total de observaciones y no solo sobre las observaciones de variables categóricas.

#### Fase I. Descarte de variables no informativas

**Tabla 5.2 Evaluación comparativa de calidad de datos (DQ) antes y después del procedimiento P2**

ID	Métrica	Evaluación de DQ antes del procedimiento	Evaluación de DQ después del procedimiento
P2	% variables no informativas	47/109 (43.12%)	0/62 (0%)
	% observaciones no informativas	827 330/1 881 732 (43.97%)	0/1 054 402 (0%)

## Fases II y III. Obtención de una estructura homogénea y estable

**Tabla 5.3 Evaluación comparativa de calidad de datos (DQ) antes y después de cada procedimiento (P3-P7)**

ID	Métrica	Evaluación de DQ antes del procedimiento	Evaluación de DQ después del procedimiento
P3	% formularios multicódigo	11/17 (64.71%)	0/6 (0%)
	% variables multicódigo	0/62 (0%)	0/62 (0%)
	% observaciones afectadas por contextualización en formularios y/o variables	516 624/1 054 402 (49.00%)	0/1 054 402 (0%)
P6	% variables categóricas afectadas por contextualización en sus valores	2/37 (5.41%)	0/37 (0%)
	% observaciones afectadas por contextualización en sus valores	12 829/1 054 402 (1.22%)	0/1 054 402 (0%)
	% sujetos afectados por contextualización en sus valores	4681/6427 (72.83%)	0/6427 (0%)
	% variables categóricas afectadas por estabilidad temporal	0/37 (0%)	0/37 (0%)
	% observaciones con dominio de valores variable en el tiempo	0/1 054 402 (0%)	0/1 054 402 (0%)
	% sujetos con dominio de valores variable en el tiempo	0/6427 (0%)	0/6427 (0%)

## Fase V. De-duplicación y elaboración de la estructura canónica del repositorio

**Tabla 5.4 Evaluación comparativa de calidad de datos (DQ) antes y después de cada procedimiento (P10-P12)**

ID	Métrica	Evaluación de DQ antes del procedimiento	Evaluación de DQ después del procedimiento
P10-P11	% observaciones de-duplicadas	11 138/1 054 402 (1.06%)	0/1 043 219 (0%)

Respecto al IDR completo, a través del identificador común para el sujeto de la asistencia en el HIS hospitalario y el sistema de información de atención primaria se pudo determinar que el 84.59% (5436) de los niños con seguimiento en atención primaria tenían episodio perinatal en el IDR, coincidiendo con niños nacidos en el Hospital Virgen del Castillo durante el periodo de estudio. Por otro lado, los 422 recién nacidos restantes existentes en el repositorio perinatal corresponden a recién nacidos en el Hospital Virgen del Castillo que no han tenido seguimiento en atención primaria en ninguno de los 3 centros de salud del Área de Salud V - Altiplano.

La información necesaria para el seguimiento de la alimentación es homogénea para los dos episodios asistenciales que recoge el IDR. En concreto, para el episodio de seguimiento, se identificaron las variables de los conceptos clínicos "Recuerdo 24h", "Hitos de alimentación" y "Somatometría". Para el episodio perinatal encontramos variables equivalentes adaptadas a la edad del recién nacido, salvo las relativas a los hitos de alimentación por tratarse de la cronología de introducción de alimentos. Se puede consultar el detalle de estas variables y sus correspondientes conceptos clínicos en las tablas A.1 y A.2 del apéndice A. De este modo, la información de alimentación del episodio perinatal de los 5436 niños que nacieron en el Hospital Virgen del Castillo (recogida fundamentalmente al

alta hospitalaria) se añade como contacto inicial (al alta hospitalaria) del seguimiento de alimentación. Así, el IDR resultante contiene un total de 63 804 contactos para el seguimiento de la alimentación.

### 5.3.1 Resultados de la paquetización

El resultado de la paquetización fue la aplicación web *ProDQA*, compuesta por los siguientes módulos funcionales:

- Parametrización de repositorios: Corresponde al proceso P1 de la fase I del proceso de construcción de repositorios ([García-de-León-Chocano et al., 2015](#)), donde además de los interfaces para especificar los parámetros definidos en el proceso P1 (mapeo de conceptos Contsys, determinación de HdE, clasificación de variables en informativas y no informativas, definición de conceptos clínicos y variables que los componen y definición de criterios de integridad semántica), se especifican también los parámetros necesarios para la creación de la base de datos SQL Server que compondrá el repositorio y que incluye todas las estructuras necesarias para el proceso (tabla de *raw data* inicial, estructuras intermedias y estructura final), definidas en la figura 2.3
- Recolección rutinaria de HCE y carga de datos brutos: Incluye interfaces para especificar los parámetros de conexión a las distintas fuentes de información de HCE, consultas de extracción de datos y procesos de carga en a la tabla de *raw data* para la generación de datos brutos. También para definir la periodicidad de la extracción y carga de datos. A partir de estas definiciones, un proceso *batch* implementado mediante trabajos y transformaciones de Pentaho Data Integration ejecuta la extracción periódica y carga de los datos brutos en la tabla de *raw data*.
- Anonimización: Se invoca automáticamente tras cada carga de datos. Utiliza algoritmos de pseudo-anonimización tanto para los datos personales (Id\_Sujeto e Id\_Usuario), como para determinados datos organizativos como el centro de salud.
- Transformaciones de aseguramiento de DQ: Implementación de los procedimientos del proceso de construcción de repositorios ([García-de-León-Chocano et al., 2015](#)) a partir de P2. La implementación de los mismos se realizó utilizando trabajos y transformaciones diseñadas en Pentaho Data Integration. Los procedimientos se van invocando en orden tras cada carga y anonimización. El proceso de construcción se detiene cuando es necesario un punto de corte en algún procedimiento, motivado por nuevas casuísticas en los datos cargados que requieren la introducción de nueva información para el aseguramiento. Cuando se produce un punto de corte se genera un log con el motivo, especificando el modulo y procedimiento implicado para que un usuario introduzca la información necesaria.
- Gestión de la homogeneización: Presenta los interfaces necesarios para la realización de los procedimientos P3 (Armonización de variables y formularios multicódigo), P6 (Armonización de variables categóricas) y P7 (Categorización de variables de texto corto).
- Gestión de la de-duplicación: Presenta los interfaces necesarios para la realización de los procedimientos P10 y P11.
- Gestión de la integridad semántica: Presenta los interfaces necesarios para la realización de los procedimientos P8, P9 y P13
- Gestión de usuarios: Presenta los interfaces para definir los usuarios, roles y sus permisos.

La figura 5.2 muestra, a modo de ejemplo, el interfaz de Gestión de la homogeneización para la armonización de formularios multicódigo (P3).

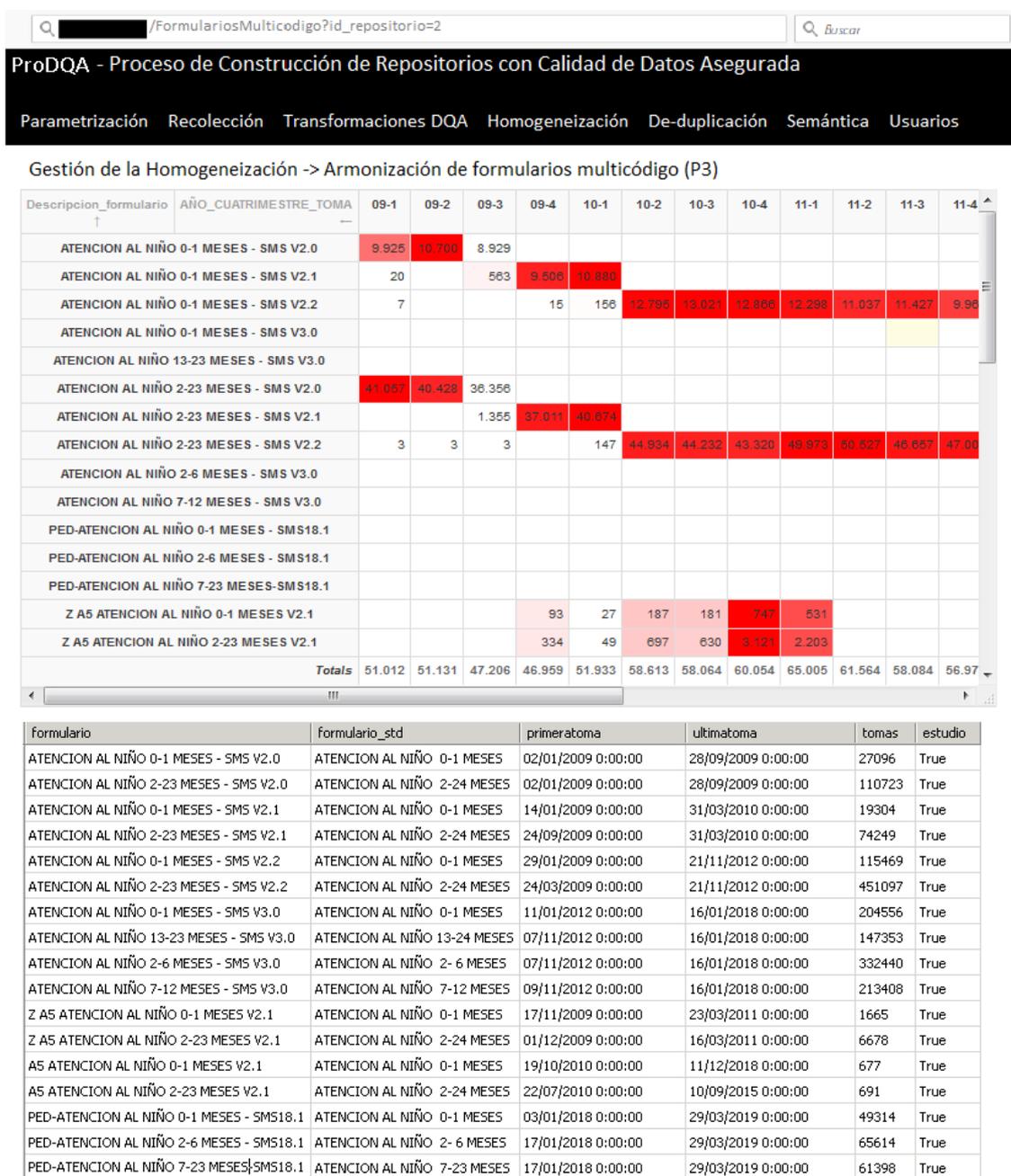


Figura 5.2 Interfaz de Gestión de la homogeneización para la armonización de los formularios multicódigo (correspondiente al procedimiento P3). En la parte superior, gráfico de ayuda a la homogeneización. En la parte inferior, interfaz para resolver formularios multicódigo mediante la asignación de códigos únicos (formulario\_std)

En la figura 5.2, correspondiente al interfaz de Gestión de la homogeneización para los formularios multicódigo (P3), se observa en la parte superior la gráfica de ayuda a la homogeneización con el total de tomas por cuatrimestre de cada formulario para detectar problemas de contextualización. En la parte inferior se encuentra una tabla con una línea por formulario origen donde en la columna formulario\_std se debe especificar un código único para los formularios identificados como multicódigo. Se puede observar, por ejemplo, que los formularios “ATENCIÓN AL NIÑO 0-1 MESES – SMS V2.0”, “ATENCIÓN AL NIÑO 0-1 MESES – SMS V2.1”, “ATENCIÓN AL NIÑO 0-1 MESES – SMS V2.2”, “ATENCIÓN AL NIÑO 0-1 MESES – SMS V3.0”, “Z A5 ATENCIÓN AL NIÑO 0-1 MESES – SMS V2.1”, “A5 ATENCIÓN AL NIÑO 0-1 MESES – SMS V2.1”, “PED ATENCIÓN AL NIÑO 0-1 MESES – SMS V18.1” se han identificado como multicódigo y se les ha puesto en consecuencia un único código en la columna formulario\_std “ATENCIÓN AL NIÑO 0-1 MESES”.

## 5.4 Discusión

El proceso de construcción de repositorios definido en el capítulo 2 ([García-de-León-Chocano et al., 2015](#)), utilizado previamente para construir el repositorio perinatal definido en el capítulo 3 ([García-de-León-Chocano et al., 2016](#)), se ha mostrado útil para ampliar nuestro repositorio de investigación con la información de seguimiento en atención primaria, obteniendo un nuevo IDR, que integra la información de los dos niveles asistenciales (hospital y atención primaria).

El descarte de variables no informativas (42.12%) en la fase I es alto, como ya ocurrió con el repositorio perinatal, debido a que se extrajeron todas las variables de los formularios de seguimiento para posteriormente descartar las variables sin valor predictivo para el seguimiento de la alimentación. Este descarte representa una mejora considerable de la eficiencia de los procedimientos posteriores. A continuación, se abordó la integridad de la estructura (fases II y III) corrigiendo problemas de contextualización y estabilidad temporal que afectaban a casi la mitad de las observaciones. Cabe destacar que, a diferencia del repositorio perinatal, la estructura de las variables se ha mantenido bastante estable a lo largo del periodo de estudio: No se encontraron variables multicódigo a pesar de tener un alto porcentaje (64.71%) de formularios multicódigo, y respecto al contenido de las mismas, solo el 5.41% de las variables estuvieron afectadas por problemas contextualización en sus valores debido a cambios en la codificación de las categorías. No se encontraron cambios que afectasen al número de categorías por lo que no se reportaron problemas de estabilidad temporal. En cualquier caso, se obtuvo una estructura homogénea y estable sin cambiar el sentido clínico original, lo que facilitará el análisis semántico posterior, descrito en el capítulo siguiente.

La automatización de los procedimientos de la fase V facilitó que se pudiese adelantar esta fase respecto a la fase IV para obtener la estructura canónica antes de abordar el aseguramiento de la integridad semántica, lo que supone una mejora de diseño respecto a la aplicación del proceso para el caso perinatal. A diferencia de lo ocurrido con el repositorio perinatal, el porcentaje total de observaciones descartadas en la de-duplicación para la obtención de la estructura canónica es bajo (1.06%) debido a que según el flujo de formularios definido en la HCE para el seguimiento del niño en atención primaria, es poco común que se registre más de un formulario para cada contacto (visita de revisión).

El Id\_Sujeto común en los dos sistemas de información de origen (Selene y OMI-AP) es un elemento fundamental para disponer de información integrada de los ámbitos perinatal y de seguimiento en atención primaria para un mismo sujeto de la asistencia. Adicionalmente, el uso de un sistema de conceptos común para ambos sistemas, basado en el es-

tándar UNE-EN ISO 13 940, facilita el mapeo de conceptos y variables equivalentes entre los dos sistemas, dando soporte a la continuidad asistencial a nivel organizativo (interoperabilidad organizativa).

Gracias a la estructura homogénea de los datos de alimentación en todo el IDR, la información de alimentación al alta hospitalaria recogida durante el contacto en planta de los 5436 niños con seguimiento de alimentación en primaria y nacidos en el Hospital Virgen del Castillo, se pudo añadir como contacto inicial del seguimiento de alimentación, ampliando el número de contactos para el seguimiento a 63 804 contactos.

Al descartar el formulario de la revisión de los dos años, la última revisión incluida en el IDR para la monitorización del PAAL es la de los 18 meses.

La aplicación web fue usada para ampliar el repositorio perinatal con la información relativa al seguimiento de alimentación descrita en el apartado 5.2 de este capítulo.

### **Trabajos futuros**

Los procedimientos empleados en el repositorio perinatal para el aseguramiento de la integridad semántica, descritos en el capítulo 3, utilizaban una perspectiva transversal. Aunque se tuvo en cuenta el componente temporal en las evaluaciones de DQ, éste estuvo circunscrito a una misma etapa (etapa perinatal). Sin embargo, para el seguimiento de la alimentación no es suficiente con replicar el enfoque transversal aplicando el foco a asegurar la DQ de cada revisión de salud de forma independiente. La propia naturaleza de la información de seguimiento requiere ampliar el abordaje del aseguramiento de la DQ al plano longitudinal, para lo que será necesario ampliar el proceso de construcción de repositorios con nuevos procedimientos para tal fin. Estos procedimientos estarán orientados a la obtención de una estimación robusta de los indicadores de alimentación infantil.

## 6 Estimación robusta de indicadores de alimentación infantil

En este capítulo, se presenta la estimación robusta de indicadores de alimentación infantil a partir de la evaluación de calidad de datos del registro longitudinal de alimentación. El registro longitudinal está conformado por una serie de los contactos que incluyen el contacto durante el periodo perinatal (que finaliza con el alta hospitalaria tras el nacimiento) y la serie de revisiones de salud infantil de atención primaria hasta los 18 meses.

En este capítulo, a partir de los datos procedentes de la recolección rutinaria de HCE que alimentan el IDR de alimentación infantil del Área de Salud V – Altiplano descrito en el capítulo anterior, la evaluación de DQ se realiza para cada uno de los indicadores de alimentación infantil, con el objetivo de obtener una estimación robusta, al tiempo que se describe la DQ de cada indicador y se determina el porcentaje de sujetos con información válida para cada uno de ellos. La propia naturaleza de los datos e indicadores a estimar requirió implementar un sistema de evaluación longitudinal de DQ de la serie de contactos. La meta-información de DQ obtenida guió el cálculo de la duración de los diferentes tipos de lactancia materna: lactancia materna exclusiva (EBF, del inglés *Exclusive Breastfeeding*), lactancia materna completa (FBF, del inglés *Full Breastfeeding*) y cualquier tipo de lactancia materna (ABF, del inglés *Any Breastfeeding*). La duración de lactancia materna y meta-información de DQ fueron los elementos principales para obtener la estimación robusta de indicadores de alimentación infantil.

Se trata de la última pieza desarrollada en la tesis que, junto con el resto de contribuciones, nos permitió habilitar finalmente en el Área de Salud V – Altiplano un sistema de monitorización poblacional, continuo y fiable de indicadores de alimentación infantil definidos en estrategias de salud internacionales como la BFHI, a partir de datos procedentes de la recolección rutinaria de HCE (C7).

*El contenido de este artículo es una adaptación para la tesis del artículo publicado (García-de-León-Chocano et al., 2021) – publicación Pu6 y de la publicación “An information-based best practice for mother and child care outcomes improvement. A successful bottom-up strategy” para el congreso i-HD Hospital Network Workshop, 2018 – publicación Pu5. Como resultado, quedó definido e implantado en la Gerencia del Área de Salud V – Altiplano el “Sistema de monitorización de indicadores de alimentación infantil con calidad de datos asegurada”, contribución C7.*

### 6.1 Introducción y antecedentes

Los primeros 1000 días de vida, desde la concepción hasta los dos años de vida, son un período único de oportunidad donde se establecen las bases de una salud, crecimiento y desarrollo neurológico óptimos a lo largo de la vida (United Nations Children’s Fund, World Health Organization, World Bank, Early Childhood Development Action Network, The Partnership for Maternal, Newborn and Child Health, 2018). Uno de los aspectos clave de este período es una nutrición adecuada. La lactancia materna es la norma biológica para la alimentación infantil durante la vida temprana y, por lo tanto, proporciona numerosos beneficios para la salud de la madre y el bebé (Cattaneo, 2012; Victora et al., 2016).

La Organización Mundial de la Salud (OMS) recomienda lactancia materna exclusiva durante los primeros 6 meses de vida y con alimentos complementarios adecuados hasta al menos los 2 años de edad (World Health Organization, 2003). Sin embargo, menos de la mitad de los bebés son amamantados de acuerdo con estas recomendaciones. Mejorar esta realidad es fundamental para lograr los objetivos mundiales de nutrición, salud y supervivencia, crecimiento económico y sostenibilidad ambiental (Rollins et al., 2016; Victora et al., 2016; WHO, UNICEF, 2018).

La Iniciativa BFHI es una estrategia internacional de la OMS / UNICEF destinada a mejorar las prácticas de lactancia materna en los servicios de salud (Rollins et al., 2016) y es una de las intervenciones más eficaces (Pérez-Escamilla et al., 2016) y rentables (Horton et al., 1996) para la mejora general de las tasas de lactancia materna. Sin embargo, en la actualidad, solo el 10% de los nacimientos en el mundo tienen lugar en un hospital acreditado como BFHI (WHO, 2017, p.).

Un factor importante para la implementación y el mantenimiento exitosos de la BFHI es la monitorización regular de indicadores de proceso y resultado (Pérez-Escamilla, 2007; Pérez-Escamilla & Chapman, 2012; Zakarija-Grković et al., 2018). Los indicadores de resultado recomendados para el seguimiento de la BFHI son los “indicadores para evaluar las prácticas de alimentación del lactante y del niño pequeño (IYCF, del inglés *Indicators for assessing infant and young child feeding practices*)” definidos por la OMS / UNICEF (WHO, UNICEF, 2007). En particular, el inicio, la duración y la exclusividad de la lactancia materna son resultados importantes de los estudios sobre lactancia materna (Wood & Woods, 2018). Además, para lograr la acreditación BFHI, los establecimientos sanitarios deben demostrar una tasa de al menos 75% de lactancia materna exclusiva al alta.

A pesar de ello, en la actualidad, una minoría de países lleva a cabo una monitorización regular de la alimentación infantil (Pérez-Escamilla & Moran, 2016; UNICEF, 2020). Una de las principales barreras para establecer un sistema de monitorización regular es la obtención de datos poblacionales fiables, lo que plantea el problema de obtener un conjunto relevante de variables para medir y evaluar la calidad de grandes conjuntos de datos (Chapman & Pérez-Escamilla, 2009; European Centre for Disease Control Prevention, 2014; Nickel et al., 2017; Paul et al., 2019; Pérez-Escamilla & Chapman, 2012).

Actualmente, los datos de procedentes de recolección rutinaria de HCE a partir de revisiones de salud infantil de atención primaria son una excelente fuente de datos candidata para la monitorización, investigación y vigilancia de la alimentación infantil debido a su cobertura poblacional y su continuidad en el tiempo (Ajetunmobi et al., 2014; American Academy of Pediatrics et al., 2008; Euro-Peristat project with SCPE and Eurocat, 2010; Halvorsen et al., 2015; Nickel et al., 2017; Paul et al., 2019; Weng et al., 2012). Con respecto a esto, existen estudios de investigación previos en alimentación infantil basados en datos procedentes de recolección rutinaria de HCE (Ajetunmobi et al., 2014; Bartsch et al., 2018; Busck-Rasmussen et al., 2014; Halvorsen et al., 2015).

No obstante, dado que los datos de HCE se recopilan de forma rutinaria con fines clínicos en lugar de para la investigación y monitorización, se debe realizar una evaluación de la DQ para garantizar la validez y fiabilidad de la reutilización de los datos en la investigación, la calidad de la atención sanitaria o la monitorización de (Khare et al., 2017; Reimer et al., 2016; Verheij et al., 2018; Weiskopf & Weng, 2013; Weng et al., 2012). En particular, para el re-uso en monitorización, la DQ tiene un impacto real y directo en la validez y fiabilidad de los indicadores (Dentler et al., 2014). A pesar de ello, estudios recientes sobre alimentación infantil basados en HCE no especifican análisis de DQ (Ajetunmobi et al., 2014; Bartsch et al., 2018; Busck-Rasmussen et al., 2014; Halvorsen et al., 2015) y algunos de ellos no utilizaron criterios estándar para asignar el estado de lactancia materna (Ajetunmobi et al., 2014; Busck-Rasmussen et al., 2014).

El Área de Salud V del Servicio Murciano de Salud (España) cuenta con una política establecida de protección, promoción y apoyo a la lactancia materna desde que su hospital de referencia fue acreditado como hospital BFHI en 2004. La política, definida en el marco de la experiencia “Hasta Que Tú Quieras” (HQTQ), incluyó una monitorización poblacional a

partir de datos HCE para evaluar no sólo el mantenimiento de criterios de calidad basados en la BFHI, sino también la promoción de la lactancia materna en la escuela y en el entorno laboral y social.

En trabajos anteriores, los autores propusieron un proceso de evaluación y aseguramiento de DQ para datos de HCE (García-de-León-Chocano et al., 2015; Sáez et al., 2012), incluida una aplicación a datos perinatales (García-de-León-Chocano et al., 2016) basada en 8 dimensiones de DQ: valor predictivo, corrección, duplicación, consistencia, completitud, contextualización, estabilidad temporal y estabilidad espacial. Posteriormente, diseñaron un proyecto piloto en España (Sáez, Moner, et al., 2017) que sentó las bases para una monitorización fiable de indicadores de alimentación infantil, centrándose en la estandarización y el análisis de DQ, principalmente de las variables perinatales, sin abordar el análisis del registro longitudinal de alimentación.

En este estudio, prestamos atención a la DQ del registro del seguimiento de la alimentación infantil. Utilizamos el registro longitudinal formado por los datos al alta hospitalaria y los datos de atención primaria de las revisiones de salud infantil. Este registro longitudinal contiene las variables necesarias para estimar todos los indicadores IYCF.

La naturaleza longitudinal de estos registros introduce varios problemas de calidad adicionales que deben abordarse. Primero, para generar indicadores derivados de variables clínicas que no fueron desarrolladas para ese re-uso es necesaria una construcción meditada de variables válidas (Khare et al., 2017). Además, debido a que la DQ debe analizarse a lo largo del tiempo (Guidry et al., 2015; M. G. Kahn et al., 2016; Li et al., 2014; Reimer et al., 2016; Whitney et al., 1998), es necesario evaluar problemas longitudinales de completitud y consistencia. Por lo tanto, un simple mapeo no es suficiente para monitorizar intervenciones a partir de registros de seguimiento (Khare et al., 2017).

Dada esta complejidad para generar información válida y fiable para la monitorización a partir de registros longitudinales, la pregunta es: ¿Es posible asegurar una estimación robusta de los indicadores de alimentación infantil a partir de datos procedentes de recolección rutinaria de HCE?

El objetivo de esta investigación es realizar una evaluación de DQ y un proceso de aseguramiento de los datos procedentes de recolección rutinaria de EHR para permitir estimaciones robustas de los indicadores de alimentación infantil. En este trabajo, informamos sobre el proceso de construcción de variables derivadas y generación de meta-información DQ necesaria para una estimación válida y fiable de los indicadores de alimentación infantil.

## 6.2 Materiales

Utilizamos los datos poblacionales de alimentación infantil cargados en el IDR definido en el capítulo anterior a partir de los datos de HCE del Área de Salud V del Servicio Murciano de Salud. El Área de Salud V – Altiplano está compuesta por tres centros de salud públicos y un hospital público (Hospital Virgen del Castillo) para atender a una población de 60 032 personas, con un 6.74% de origen extranjero. El porcentaje de mujeres en edad fértil sobre la población total del Área V es del 20.12%. Los datos del estudio comprenden 6427 niños nacidos entre el 1 de enero de 2009 y el 31 de diciembre de 2018 y seguidos en uno de los 3 centros de salud, lo que representa el 98.68% de cobertura del total de niños del área de salud. El 84.59% (5436) de estos niños nacieron en el Hospital Virgen del Castillo. Tal y como se definió en el modelo de información del Proceso de Atención a la Alimentación del Lactante (PAAL), la serie de datos de alimentación infantil proviene de los formularios de HCE cumplimentados desde el nacimiento hasta los 18 meses de edad a lo largo de una serie de contactos consecutivos de salud. Todos los niños nacidos en el

hospital tuvieron un primer contacto en el hospital cumplimentado al alta por una enfermera. El resto de contactos correspondieron a siete revisiones de salud de atención primaria del Programa de Atención al Niño y Adolescente de la Región de Murcia (PANA), cumplimentados por un médico de atención primaria o una enfermera, en 8 momentos específicos: antes del primer mes, a los 2 meses, a los 4 meses, a los 6 meses, entre los 9 y los 12 meses, entre los 15 y los 18 meses. Todas las revisiones, excepto las revisiones “antes del primer mes” y “entre los 9 y 12 meses”, coinciden con los puntos de vacunación. Todos los profesionales colaboradores recibieron formación específica y participaron en el diseño y estudio piloto para el desarrollo de los formularios de alimentación infantil.

Los formularios de alimentación de la HCE cargados en el IDR tenían dos tipos de información: 1) Datos transversales referidos al tipo de lactancia materna en cada contacto obtenidos mediante un recuerdo de 24 horas (“datos transversales” es el término habitual en los artículos de investigación sobre alimentación infantil para referirse al estado actual de lactancia materna en una encuesta puntual), y 2) datos retrospectivos informados por las madres referidos a la edad del niño cuando ocurren los hitos relevantes de alimentación infantil.

El conjunto de datos inicial estaba formado por 6427 sujetos con un total de 63 804 contactos de salud procedentes de la recolección rutinaria de HCE, que componen el IDR de alimentación infantil descrito en el capítulo anterior. También se obtuvieron datos relevantes como la edad del niño en cada contacto. La Tabla 6.1 presenta la estructura de los datos brutos de los contactos.

**Tabla 6.1 Estructura de los datos brutos de recolección rutinaria de historia clínica electrónica (HCE) presentes en el repositorio de datos integrados (IDR) de alimentación infantil, compuestos por datos contextuales del sujeto de la asistencia y el contacto, datos de somatometría y datos estructurados de alimentación infantil. Los datos estructurados de alimentación infantil incluyeron datos transversales y datos retrospectivos.**

<b>Datos contextuales</b>
ID Sujeto de la Asistencia (identificador común en los sistemas de información de hospital y atención primaria)
Fecha de nacimiento
Centro de salud al que pertenece el niño
Unidad básica de salud a la que pertenece el niño
Fecha de la revisión de salud
Edad del niño durante el contacto
Usuario que registra los datos
Fecha y hora en la que se registró la observación
<b>Somatometría</b>
Peso el día de la revisión de salud
Altura el día de la revisión de salud
Perímetro cefálico el día de la revisión de salud
<b>Datos estructurados de alimentación infantil</b>
<b>Recuerdo de 24 horas de alimentación infantil [5 elementos de datos transversales]</b>
i) ¿Tipo de lactancia materna? (Lactancia materna Exclusiva, Lactancia materna Predominante, Lactancia materna Parcial, No lactancia materna) *
ii) ¿Ha tomado lactancia materna en las últimas 24 horas? (Sí, No) **
iii) ¿Ha tomado alimentos complementarios en forma de sólidos o semi-sólidos en las últimas 24 h? (Sí, No) **
iv) ¿Ha tomado alimentos complementarios en forma de leche de fórmula en las últimas 24 horas? (Sí, No) **
v) ¿Ha tomado agua, infusiones, soluciones de rehidratación oral y/o zumos en las últimas 24 horas? (Sí, No) **
<b>Hitos de alimentación infantil [14 elementos de datos retrospectivos sobre la edad del niño en cada hito]</b>
Líquidos no nutritivos, Cereales sin gluten, Cereales con gluten, Frutas, Verduras, Leche de fórmula, Pollo, Pescado, Legumbres, Yema de huevo, Huevo entero, Yogur, Leche de vaca, Abandono de lactancia materna

\*: Elemento disponible en formularios de HCE para todas las revisiones;

\*\* : Elemento disponible solo en formularios de HCE para revisiones hasta el sexto mes

Esta estructura de datos se utilizó para construir variables y estimar indicadores de alimentación infantil. Para este estudio, definimos los tipos de lactancia materna de acuerdo con las definiciones más aceptadas propuestas por diferentes autores (M. H. Lobbok & Starling, 2012; M. Lobbok & Krasovec, 1990; WHO, UNICEF, 2007). En cada contacto, la alimentación infantil se clasificó de acuerdo con estas definiciones: La lactancia materna exclusiva (EBF) cuando los bebés solo recibieron leche materna; Lactancia materna predominante, cuando se recibe leche materna con otros líquidos no nutritivos; y Lactancia materna parcial cuando las madres ofrecieron a sus hijos leche de fórmula o alimentos sólidos además de la leche materna. Además, creamos dos categorías principales: Lactancia materna completa (FBF), que incluye a los niños que reciben lactancia materna exclusiva o predominante; y Cualquier Lactancia Materna (ABF) cuando los bebés reciben cualquier cantidad de leche materna con o sin otros líquidos o alimentos. La Tabla 6.2 presenta los indicadores propuestos para el seguimiento de lactancia materna, basados en los indicadores IYCF definidos por la OMS / UNICEF (WHO, UNICEF, 2007) y los datos disponibles en nuestro conjunto de datos. Adicionalmente, considerando los datos disponibles, propusimos indicadores de EBF al alta hospitalaria, EBF menor de 15 días e indicadores de FBF.

**Tabla 6.2. Indicadores para la monitorización de la alimentación infantil para los distintos tipos de lactancia materna: lactancia materna exclusiva (EBF), lactancia materna completa (FBF) y cualquier tipo de lactancia (ABF).**

Indicador	Descripción
EBF al alta hospitalaria	Proporción de lactantes al alta hospitalaria que fueron alimentados exclusivamente con leche materna (EBF) durante el día anterior
EBF antes de los 15 días	Proporción de lactantes de 0 a 15 días de edad alimentados exclusivamente con leche materna (EBF) durante el día anterior
EBF antes de los 6 meses	Proporción de lactantes de 0 a 5 meses de edad alimentados exclusivamente con lactancia materna (EBF) durante el día anterior
FBF antes de los 6 meses	Proporción de lactantes de 0 a 5 meses de edad que recibieron lactancia materna exclusiva o lactancia materna predominante durante el día anterior
Lactancia materna continua al año de vida	Proporción de niños de 12 a 15 meses de edad que son amantados (ABF)
Mediana de duración de lactancia materna ABF	Mediana de duración de lactancia (ABF) entre los niños menores de 24 meses de edad. Referido a cualquier tipo de lactancia materna (ABF)
Mediana de duración de FBF	Mediana de duración de FBF entre los niños menores de 24 meses de edad
Mediana de duración de EBF	Mediana de duración de EBF entre los niños menores de 24 meses de edad

## 6.3 Métodos

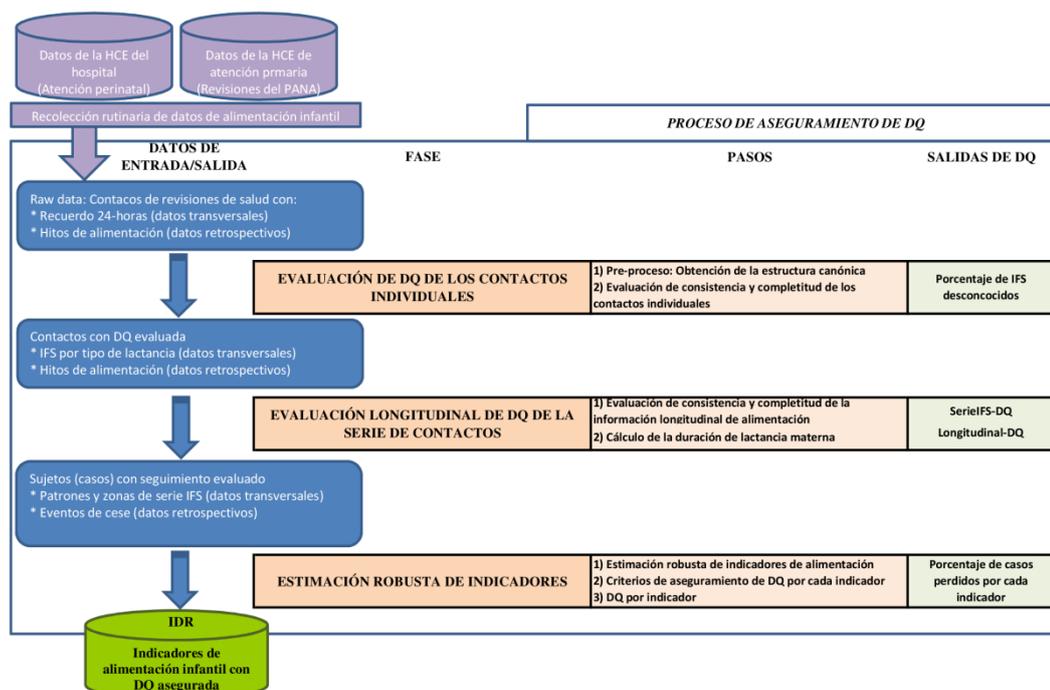
Un grupo de expertos formado por 4 clínicos líderes del área de salud en este dominio y 3 expertos en tecnología de la información, autores del artículo (García-de-León-Chocano et al., 2021), definieron el proceso de aseguramiento de DQ para lograr una estimación robusta de indicadores de la alimentación infantil.

Los criterios de inclusión fueron todos los niños con seguimiento de alimentación infantil en cualquiera de los tres centros públicos de salud del Área de Salud V. Definimos un criterio de exclusión general para los sujetos que no cumplen con un seguimiento mínimo: menos de 3 revisiones del Programa del Niño Sano, primera revisión después de los 120 días de

edad o última revisión antes de los 180 días. Todos los datos se anonimizaron antes de los análisis.

Cada sujeto incluido (caso) en el estudio tenía una serie de contactos de atención sanitaria. Así, para la estimación robusta de los indicadores de alimentación infantil, realizamos un proceso de aseguramiento de DQ sobre datos transversales y retrospectivos, tanto en cada contacto individual como en toda la serie de contactos que conforman cada caso. El proceso de aseguramiento de DQ consistió en tres fases: 1) Evaluación de la información transversal de alimentación de cada contacto individual y definición del estado de alimentación del lactante (IFS, del inglés *Infant Feeding Status*) en cada contacto individual; 2) Evaluación longitudinal de DQ de la serie de contactos mediante el análisis de la completitud y consistencia de la información de alimentación longitudinal en la serie de contactos (combinando información transversal y retrospectiva) para calcular la duración de los diferentes tipos de lactancia materna para cada caso; y 3) Estimación robusta de indicadores para la monitorización de la alimentación infantil y la descripción de la DQ de cada indicador.

La Figura 6.1 resume este proceso de DQ. El resultado final es un IDR de atención materno-infantil hasta los 18 meses de edad que incorpora una estimación DQ-asegurada de los indicadores de alimentación infantil.



**Figura 6.1** Proceso de aseguramiento de calidad de datos (DQ) para una estimación robusta de los indicadores de alimentación infantil. En la parte superior, en color magenta: fuentes de datos de la recolección rutinaria de historia clínica electrónica (HCE). Para el proceso de aseguramiento de DQ se muestran: los datos de entrada / salida involucrados en cada fase, los pasos y salidas de DQ. En verde el repositorio de datos integrados (IDR) final donde se calculan los indicadores.

### 6.3.1 Fase 1: Evaluación de DQ de contactos individuales

La evaluación de DQ de contactos individuales se realizó en dos pasos. Primero realizamos un pre-proceso utilizando métodos y métricas de DQ de trabajos previos (García-

de-León-Chocano et al., 2015, 2016) para asegurar la integridad estructural de los datos y resolver la de-duplicación de datos, obteniendo una estructura canónica donde las dimensiones de DQ de contextualización, estabilidad temporal, estabilidad espacial y unicidad fueron evaluadas. La aplicación del preproceso se puede consultar en los capítulos 3 y 5.

En segundo lugar, evaluamos la completitud y la consistencia de cada contacto individual de cada caso para obtener las variables derivadas de alimentación. Para los datos transversales, evaluamos la DQ de los elementos del recuerdo de 24 horas (ver elementos de datos transversales en la Tabla 6.1) para obtener el IFS en 3 variables binarias derivadas que indicaban si los lactantes recibieron o no ABF, FBF y EBF, respectivamente. El IFS representa si el lactante se alimenta de acuerdo al tipo específico de lactancia materna en cada contacto individual (ver tabla C.1 del apéndice C para el detalle del tipo de lactancia materna según las 3 variables binarias). El elemento i) del recuerdo 24 horas ("Tipo de lactancia") determinó el IFS, el resto de elementos participaron en la evaluación de DQ, comprobando la consistencia del "Tipo de lactancia" de forma secuencial y acumulativa, específicamente el elemento ii) para ABF, los elementos iii) y iv) para FBF y el elemento v) para EBF. Si la DQ de los elementos del recuerdo de 24 horas impedía determinar el IFS para un tipo de lactancia, etiquetábamos el contacto como "IFS desconocido" para ese tipo de lactancia. Para los datos retrospectivos (hitos de alimentación infantil de la Tabla 6.1), se determinaron tres variables derivadas expresadas en días: edad de introducción de líquidos, edad de introducción de fórmula artificial o alimentos y edad de cese de lactancia materna, utilizando el valor mínimo de los hitos de alimentación correspondientes para cada caso.

FBF y EBF se evaluaron sólo en los contactos hasta la revisión de los 6 meses de acuerdo con las recomendaciones de la OMS (World Health Organization, 2003).

### **6.3.2 Fase 2: Evaluación longitudinal de DQ de la serie de contactos**

Esta fase se realizó en dos pasos: 1) Evaluación de la consistencia y completitud de la información longitudinal de alimentación a lo largo la serie de contactos (combinando los datos transversales y datos retrospectivos) utilizando una base de reglas que recoge el conocimiento experto para evaluar esta información; y 2) Cálculo de la duración de los diferentes tipos de lactancia materna para cada caso, guiado por la meta-información de DQ obtenida en el paso anterior.

#### **6.3.2.1 Evaluación de la consistencia y completitud de la información longitudinal de alimentación en la serie de contactos.**

Analizamos morfológicamente la serie de contactos para cada caso, tanto datos transversales (serie de valores de las variables IFS de la fase anterior), datos retrospectivos (serie de valores de hitos de alimentación infantil) como información combinada de ambos. Para el análisis morfológico de estas series, definimos elementos clave que permitieron establecer un conjunto de atributos DQ para la evaluación DQ, detallados en la Tabla 6.3.

**Tabla 6.3 Elementos definidos para el análisis y evaluación de calidad de datos (DQ) de la serie de contactos, tanto para los datos transversales del estado de alimentación del lactante (IFS) como para los datos retrospectivos, y la combinación de ambos. Los elementos, atributos de DQ y puntuaciones de la tabla se definieron y calcularon para cada tipo de lactancia materna.**

<b>Elementos definidos para la serie de datos transversales</b>	
<b>Momentos clave de la serie de IFS</b>	<b>Descripción</b>
T0	Edad (en días) al alta hospitalaria
T1	Edad (en días) en el primer contacto de atención primaria donde el IFS fue positivo (se identificó que el bebé estaba siendo alimentado según el tipo de lactancia)
T2	Edad (en días) en el último contacto donde el IFS fue positivo (siempre antes de que aparezca un IFS negativo)
T3	Edad (en días) en el primer contacto donde el IFS fue negativo (se identificó que el bebé no estaba siendo alimentado según el tipo de lactancia)
<b>Patrones de la serie de IFS</b>	<b>Descripción</b>
Duración no censurada	La lactancia comienza y termina dentro de la cohorte. La serie tiene los momentos T1, T2 y T3. T2 y T3 delimitan el intervalo de cese de lactancia materna
Duración censurada	Continúa lactando antes del final de la cohorte (18 meses). La serie solo tiene T1 y T2. El intervalo de cese de lactancia materna está delimitado por T2 y el final de la cohorte
No lactancia materna	No se inicia lactancia materna (no aparece T1 ni T2, solo existe T3 en la serie) El intervalo de cese de lactancia materna está delimitado por el nacimiento y T3
Patrón nulo	No hay información válida sobre la lactancia materna
<b>Zonas de la serie de IFS</b>	<b>Descripción</b>
Inicio	Entre el nacimiento y T1
Núcleo	Entre T1 y T2
Intervalo de cese de lactancia materna	Ver definición en cada patrón
Final	Desde T3 hasta el final de la serie
<b>DQ de la serie de datos transversales</b>	
<b>Atributos y puntuación de DQ de la serie de IFS</b>	<b>Descripción</b>
Contactos nulos del inicio	Nº de contactos nulos de la zona inicio
Contactos válidos del núcleo	Nº de contactos establecidos a 1-Sí en la zona núcleo
Contactos nulos del núcleo	Nº de contactos nulos de la zona núcleo
Ratio de completitud de los contactos del núcleo	Proporción de contactos válidos frente a contactos nulos dentro de la zona núcleo
Contactos nulos del intervalo de cese	Nº de contactos nulos en el intervalo de cese de lactancia materna
Contactos nulos del final	Nº de contactos nulos de la zona final
Contactos inconsistentes del final	Nº de contactos establecidos a 1-Sí desde T3 hasta el final de la serie
Contactos válidos del final	Nº de contactos establecidos a 2-No desde T3 hasta el final de la serie
SerieIFS-DQ	Puntuación resumen de la DQ de la serie de IFS (de 0 a 100)
<b>Elementos definidos para la serie de datos retrospectivos</b>	
<b>Eventos de cese</b>	<b>Descripción</b>
Edad de introducción de líquidos	El valor mínimo de los hitos relacionados con líquidos no nutritivos
Edad de introducción de alimentos	El valor mínimo de los hitos relacionados con los alimentos (sólidos, semi-sólidos o leche de fórmula)
Abandono de lactancia materna	El valor mínimo de cese de la lactancia materna
<b>Puntuación global de DQ (combinación de datos transversales y retrospectivos)</b>	
Longitudinal-DQ	Valor en escala de Likert según la posición de los eventos de cese en relación al intervalo de cese de lactancia materna

En concreto, para los datos transversales, la serie para cada tipo de lactancia materna se compuso con la secuencia de valores del IFS correspondiente. A partir de 4 momentos clave (T0, T1, T2, T3) descritos en la Tabla 6.3, pudimos identificar un conjunto de patrones de la serie IFS: Duración no censurada, Duración censurada, No lactancia materna y Patrón nulo (información no válida debido a problemas severos de completitud en la serie debido a contactos desconocidos). Además, según los momentos clave, la serie de contactos se dividió en cuatro zonas temporales (Inicio, Núcleo, Intervalo de cese de lactancia materna y Final) que delimitan las transiciones de estado dentro del tipo de lactancia materna. A partir de estos elementos, se estableció un conjunto de atributos de DQ para evaluar la completitud y consistencia de la serie de IFS. Esta evaluación de DQ se resumió en una puntuación DQ que denominamos "SerieIFS-DQ" ponderando los atributos y normalizando el resultado en una escala de 0 (para series de IFS con problemas DQ que impiden determinar el patrón) a 100 (para series de IFS sin problemas de DQ).

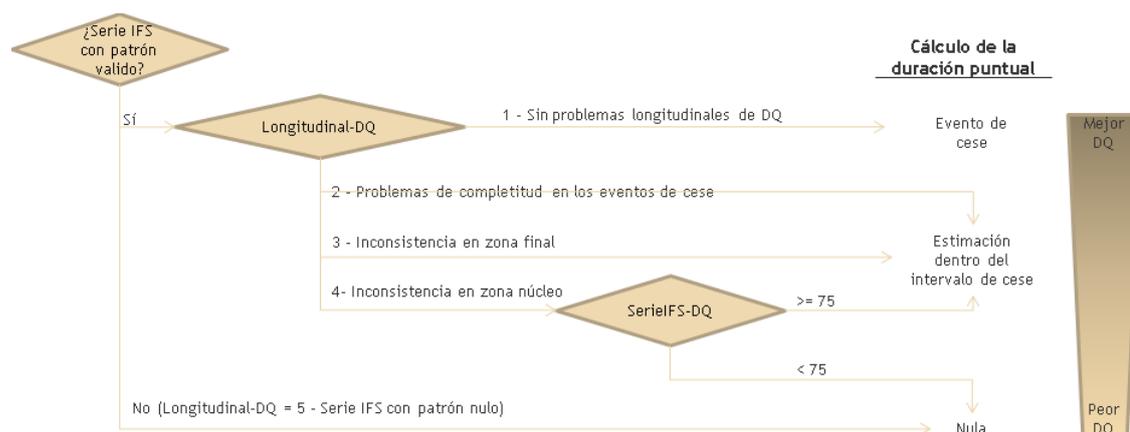
Por otro lado, para los datos retrospectivos (hitos de alimentación infantil), se determinaron tres eventos de cese expresados en días utilizando las siguientes variables derivadas: edad de introducción de líquidos, edad de introducción de alimentos y edad de cese de la lactancia materna, utilizando el valor mínimo de los hitos correspondientes en la secuencia de contactos para cada sujeto.

Finalmente, contrastamos la información extraída de los datos transversales y datos retrospectivos comparando los elementos de la serie de IFS con los eventos de cese. Utilizamos una base de reglas en forma de tabla de valores que incluye 109 entradas según los valores del tipo de lactancia, el patrón y zonas de la serie IFS, así como la posición relativa de los eventos de cese respecto al intervalo de cese de la lactancia materna. El resultado de la evaluación fue la puntuación global de DQ, que denominamos "Longitudinal-DQ", expresada en la siguiente escala Likert: 1- Sin problemas longitudinales de DQ (todos los eventos de cese relacionados con el tipo específico de lactancia materna estaban dentro del intervalo de cese de la lactancia materna de un patrón de serie IFS válido); 2- Problemas de completitud en los eventos de cese (no se registraron eventos de cese relacionados con el tipo específico de lactancia materna); 3- Inconsistencia en zona final (los eventos más relevantes para el tipo de lactancia materna eran posteriores al intervalo de cese); 4- Inconsistencia en zona núcleo (los eventos más relevantes para el tipo de lactancia eran anteriores al intervalo de cese); y 5- Serie IFS con patrón nulo.

#### 6.3.2.2 Cálculo de la duración de lactancia materna para cada caso.

Calculamos la duración de la lactancia materna para cada caso a partir de los datos transversales y retrospectivos, y guiados por la meta-información de DQ de la evaluación longitudinal de DQ. Como se muestra en la Figura 6.2, establecimos una duración puntual de la lactancia materna en dos casos: a) cuando los datos retrospectivos son completos y consistentes con los datos transversales, se estableció la duración puntual de cada tipo de lactancia al evento de cese particular, es decir, a partir de datos retrospectivos; o b) cuando no teníamos eventos de cese, o eran inconsistentes con los datos transversales, el grupo de expertos decidió establecer la duración puntual al punto medio del intervalo de cese de la lactancia materna para el patrón de Duración no censurada y el mínimo del intervalo para el patrón de Duración censurada.

Por otro lado, definimos los casos con información perdida para un tipo de lactancia cuando a) encontramos inconsistencias en la zona núcleo y la DQ de los datos transversales era pobre (SerieIFS-DQ <75) para los que el grupo de expertos decidió no establecer un valor de duración válido; o b) cuando no pudimos obtener el intervalo de cese de lactancia materna debido a un patrón nulo de la serie de IFS por lo que no se pudo determinar un valor de duración.



**Figura 6.2 Cálculo de duración de lactancia materna guiado por calidad de datos (DQ). Algoritmo para calcular la duración según la DQ de la serie de contactos, a partir del patrón de la serie de datos del estado de alimentación del lactante (IFS), las puntuaciones “Longitudinal-DQ” y “SerieIFS-DQ”. Los tipos de cálculo están ordenados de arriba a abajo de mejor a peor DQ. \*: Punto medio del intervalo de cese de lactancia materna para patrones de duración no censurada y el mínimo del intervalo de cese para patrones de duración censurada**

### 6.3.3 Fase 3: Estimación robusta de indicadores

A partir de la duración calculada para cada caso y tipo de lactancia materna en la fase anterior, estimamos el valor de los indicadores para el seguimiento de la alimentación infantil utilizando el método actuarial, y guiados por la meta-información DQ (Figura 6.2). Para el método actuarial, agrupamos los eventos de cese de lactancia materna, desde el nacimiento hasta los 720 días, en intervalos de treinta días, asignando a cada caso, en cada intervalo, un valor de 1 si continuaba lactando, un valor de 2 si había cesado, y un valor “missing” si los datos no estaban disponibles.

El grupo de expertos definió para cada indicador de alimentación infantil los criterios de aseguramiento de DQ para obtener tres niveles de DQ: DQ estándar, problemas de DQ tolerables y problemas de DQ no tolerables, estos últimos como casos perdidos.

Los criterios de aseguramiento de DQ se establecieron en función del Longitudinal-DQ y los casos “missing” resultantes de la estimación de cada indicador específico según el método actuarial. Estableciéndose de la siguiente forma: Problemas de DQ no tolerables (casos perdidos): aquellos con duración de la lactancia materna nula para el tipo de lactancia del indicador y los casos “missing” para el indicador específico; para todos los demás casos (casos válidos), los casos con DQ estándar fueron aquellos en los que la duración fue determinada por el evento de cese, el resto de casos válidos (donde la duración se estableció en el punto medio o al mínimo del intervalo de cese según el patrón) fueron casos con problemas de DQ tolerables. Excepcionalmente, para los criterios de aseguramiento del indicador “EBF al alta hospitalaria” no se tuvo en cuenta la duración de lactancia materna al no ser de aplicación, en su lugar, se empleó el dato transversal de alimentación al alta hospitalaria.

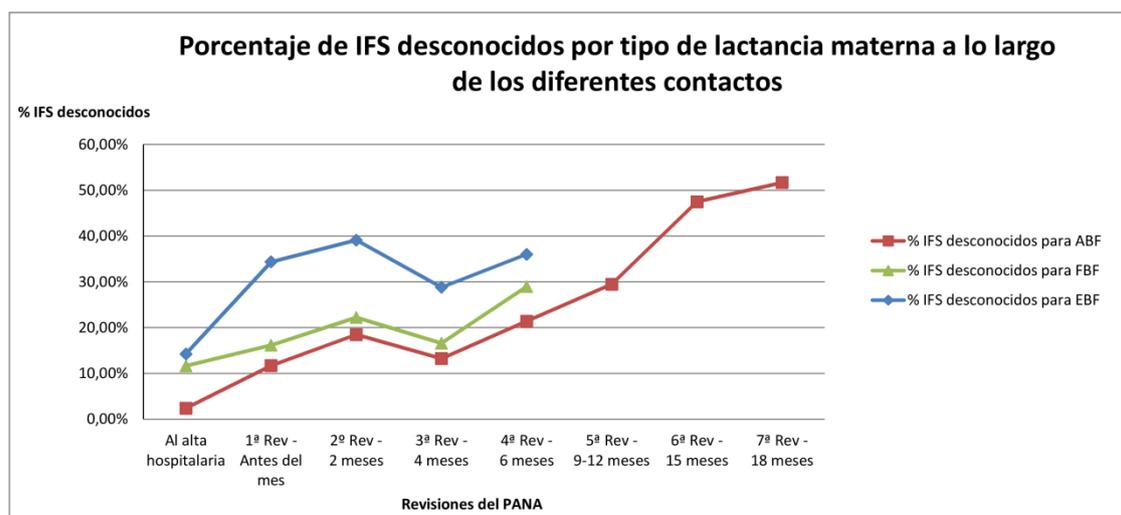
## 6.4 Resultados

De los 6427 niños, el porcentaje de sujetos excluidos por no cumplir con el seguimiento mínimo fue del 10.33% (664 sujetos). La población final incluida fue de 5763 sujetos con un total de 60 885 contactos sanitarios. Estos contactos se distribuyen homogéneamente entre la secuencia de contactos: el 87.51% de los sujetos tuvo un contacto al alta hospitalaria, el 74.51% tuvo la revisión antes del primer mes, el 97.78% a los 2 meses, el 97.88% a los 4 meses, el 97.83% a los 6 meses, 96.41% entre 9 y 12 meses, 83.71% a los 15 meses y 69.18% a los 18 meses.

### 6.4.1 Resultados de la evaluación de DQ de los contactos individuales

Todas las variables y valores del conjunto de datos inicial recolectadas de la HCE fueron armonizados y de-duplicados obteniendo una estructura homogénea y estable de contactos a lo largo del período de estudio.

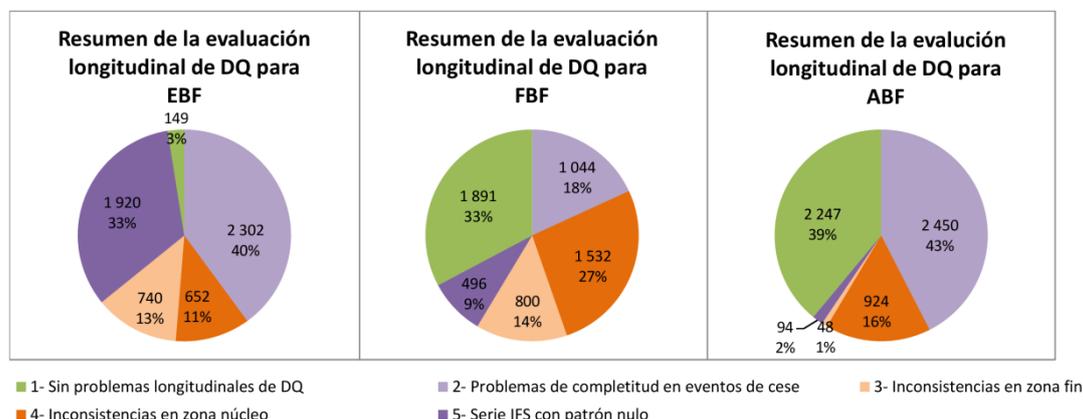
A partir del análisis de DQ de completitud y consistencia de los datos transversales de los contactos individuales: el número de contactos con IFS desconocido, donde no fue posible determinar el IFS (IFS-DQ = 0), según el tipo de lactancia fue: considerando contactos hasta la revisión de 6 meses, 11 730 (30.42%) contactos para EBF y 7332 (19.02%) para FBF; y, considerando todas las revisiones, 13 698 para ABF (22.50%). En particular, la Figura 6.3 muestra la distribución de IFS desconocidos a lo largo de la secuencia de contactos (al alta hospitalaria y revisiones de salud del Programa Niño Sano).



**Figura 6.3** Porcentaje de estados de alimentación del lactante (IFS) desconocidos por tipo de lactancia materna a lo largo de los diferentes contactos: al alta hospitalaria y la serie de revisiones del Programa de Atención al Niño y Adolescente (PANA). En el eje horizontal, mostramos la secuencia de contactos. Con línea azul, presentamos el porcentaje de IFS desconocidos para lactancia materna exclusiva (EBF), la línea verde para lactancia materna completa (FBF) y la línea roja para cualquier tipo de lactancia (ABF)

### 6.4.2 Resultados de la evaluación longitudinal de DQ

Presentamos los resultados de la evaluación longitudinal de DQ de los datos transversales combinados con los datos retrospectivos a lo largo de la serie de contactos. La Figura 6.4 muestra los resultados finales de esta evaluación resumidos por la puntuación global Longitudinal-DQ para cada tipo de lactancia (paso 1 de la fase 2).



**Figura 6.4** Distribución de los problemas longitudinales de calidad de datos (DQ) según la escala de severidad de la puntuación global Longitudinal-DQ (Likert de 1 - Sin problemas longitudinales de DQ a 5 - Serie IFS con patrón nulo). Se muestra un gráfico circular por tipo de lactancia materna: lactancia materna exclusiva (EBF), lactancia materna completa (FBF) y cualquier tipo de lactancia (ABF), para los 5763 sujetos incluidos con los problemas de DQ categorizados

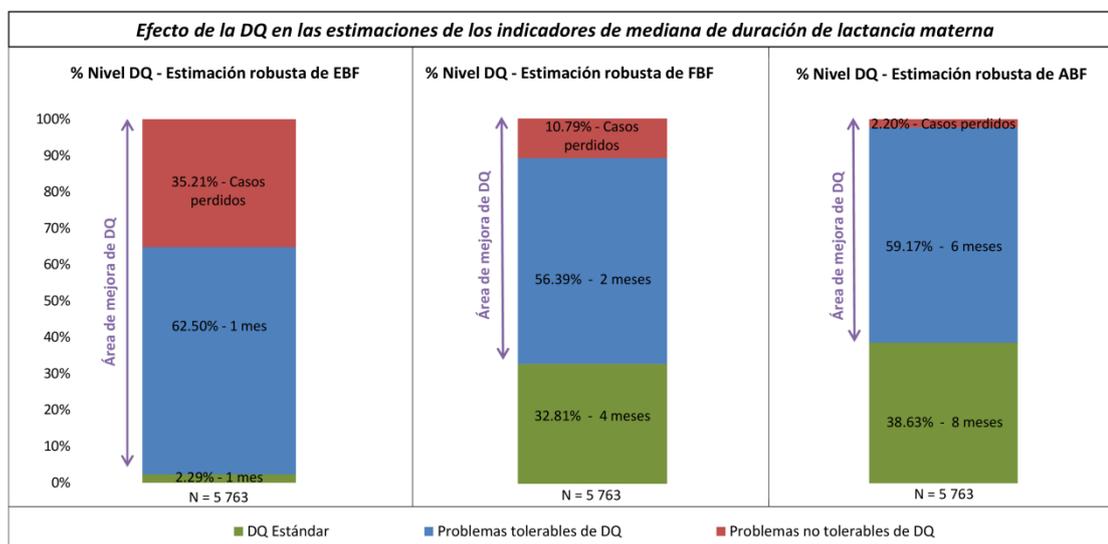
### 6.4.3 Resultados de aseguramiento de DQ para la estimación robusta de indicadores

La Tabla 6.4 presenta los resultados de la estimación robusta de indicadores con el porcentaje correspondiente de casos perdidos (sujetos con problemas de DQ no tolerables respecto a los criterios de aseguramiento establecidos para cada indicador por el grupo de expertos). Se obtuvieron bajos porcentajes de casos perdidos para todos los indicadores excepto para el seguimiento de la EBF.

**Tabla 6.4** Estimación robusta de indicadores para distintos tipos de lactancia materna: lactancia materna exclusiva (EBF), lactancia materna completa (FBF) y cualquier tipo de lactancia (ABF). Para cada indicador se muestra el porcentaje de casos perdidos y el valor de la estimación robusta (para los sujetos válidos)

Indicador	% de casos perdidos	Estimación
EBF al alta hospitalaria	14.24%	77.48%
EBF antes de los 15 días	40.15%	64.75%
EBF antes de los 6 meses	38.12%	27.30%
FBF antes de los 6 meses	13.31%	55.63%
Lactancia materna continua al año de vida	2.20%	23.86%
Mediana de duración de lactancia materna (ABF)	2.20%	7 meses
Mediana de duración de FBF	10.79%	3 meses
Mediana de duración de EBF	35.21%	1 mes

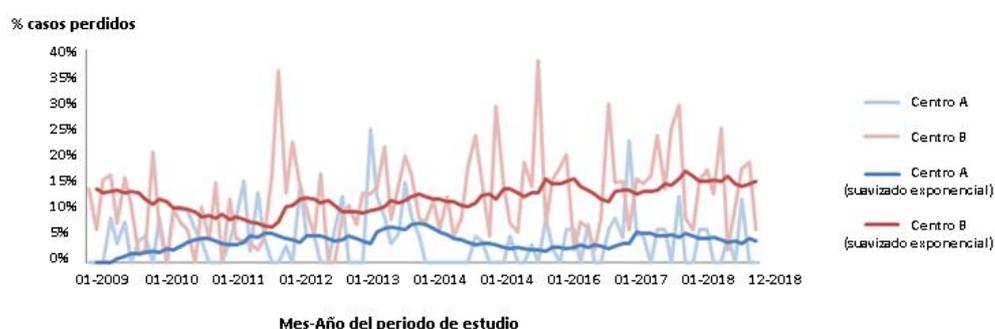
Por otro lado, presentamos la Figura 6.5 para mostrar el efecto de la DQ en las estimaciones para el ejemplo del indicador clave “Mediana de duración de la lactancia materna”. Mostramos el porcentaje de sujetos atendiendo a su nivel de DQ y la estimación del indicador para la subpoblación de sujetos con DQ estándar y sujetos con problemas de DQ tolerables. Se obtuvieron mejores tasas de lactancia para los sujetos con DQ estándar.



**Figura 6.5 Efecto de la calidad de datos (DQ) en las estimaciones del indicador “mediana de duración de lactancia materna”.** Mostramos un gráfico de columnas para el indicador de mediana de duración por tipo de lactancia: lactancia materna exclusiva (EBF), lactancia materna completa (FBF) y cualquier tipo de lactancia (ABF). Las columnas representan el porcentaje de sujetos atendiendo a su nivel de DQ junto a la estimación robusta para las subpoblaciones de casos válidos. Los sujetos con DQ estándar se muestran en verde y los sujetos con problemas tolerables de DQ en azul. Los casos perdidos se muestran en rojo. También destacamos el área de mejora de DQ

Así mismo, utilizamos el nivel de DQ “casos perdidos” sobre el indicador anterior (“mediana de duración de FBF”) para representar la evolución temporal respecto al total de casos, distinguiendo por centro de salud (figura 6.6).

**Evolución de casos perdidos para el indicador “mediana de duración de FBF” durante el periodo de estudio**



**Figura 6.6 Evolución del porcentaje de casos perdidos para el indicador “mediana de duración de lactancia materna completa (FBF)” a lo largo del tiempo, diferenciado entre dos centros de salud del Área.** En líneas azul y rojo claro se muestra el porcentaje de perdidos para dos centros de salud del área. En líneas azul y rojo oscuro se muestra el porcentaje de perdidos para los mismos centros utilizando suavizado exponencial.

## 6.5 Discusión

El cálculo de indicadores internacionales de alimentación infantil, como los indicadores de resultado de la BFHI, generalmente se basan en costosas encuestas ad-hoc, que se llevan a cabo durante períodos prolongados. Este trabajo introduce una estimación robusta de indicadores a través de un proceso de aseguramiento de DQ de HCE.

Tras el proceso de aseguramiento de DQ, la mayoría de las estimaciones robustas de los indicadores propuestos para los diferentes tipos de lactancia materna tuvieron alrededor del 10% de casos perdidos y coincidieron con lo esperado en el hospital de estudio, acreditado como hospital BFHI en tres ocasiones. Estudios previos basados en encuestas mostraron resultados de indicadores de lactancia materna para el Hospital Virgen del Castillo que destacaban dentro del contexto nacional ([García-de-León-González et al., 2011](#)). Esta tendencia se mantuvo en los indicadores que presentamos. Así, el indicador de EBF al alta hospitalaria tras el parto (77.8%) fue ligeramente superior al mínimo requerido para un hospital BFHI (75%) ([WHO, 2017](#)). Además, el indicador de FBF menor de 6 meses (55.63%) fue superior a su indicador equivalente incluido en la última Encuesta Nacional de Salud (39%), así como las medianas de FBF y ABF. Aunque no teníamos datos nacionales recientes para comparar los resultados de nuestros indicadores de EBF después del alta (EBF de menos de 15 días, EBF de menos de 6 meses y la mediana de duración de EBF), la diferencia con los indicadores de EBF disponibles para el hospital en estudios anteriores ([García-de-León-González et al., 2011](#)) reflejó los problemas sustanciales que habíamos detectado en los análisis de DQ para las variables clave en esos indicadores, con más del 30% de casos perdidos.

A pesar de la relevancia y las ventajas de utilizar los datos de HCE para la monitorización e investigación de la alimentación infantil ([Ajetunmobi et al., 2014](#); [American Academy of Pediatrics et al., 2008](#); [Euro-Peristat project with SCPE and Eurocat, 2010](#); [Halvorsen et al., 2015](#); [Nickel et al., 2017](#); [Paul et al., 2019](#); [Weng et al., 2012](#)), a nuestro entender, este es el primer estudio que se ocupa de la evaluación de DQ para obtener una estimación robusta de los indicadores IYCF sobre un IDR poblacional basado en datos procedentes de recolección rutinaria de la HCE. Hay dos ejemplos de estudios poblacionales de alimentación infantil que utilizan registros de HCE exclusivamente ([Halvorsen et al., 2015](#)), ([Bartsch et al., 2018](#)), pero ninguno de ellos menciona el uso de procedimientos de evaluación de DQ y uno de ellos necesita una revisión manual complementaria de los registros de HCE ([Bartsch et al., 2018](#)). No hemos encontrado referencias de otros repositorios poblacionales de alimentación infantil hasta los 18 meses con DQ asegurada basados en HCE para una monitorización fiable de indicadores de resultado de alimentación infantil.

Encontramos problemas de DQ en los contactos individuales para establecer el valor de IFS (IFS desconocido) en un 30.42% para EBF, 19.02% para FBF y 22.50% para ABF. Estos valores no se distribuyen homogéneamente a lo largo de la secuencia de contactos sino que aumentan según la edad de la revisión del Programa Niño Sano como podemos ver en la Figura 6.3 Para el primer contacto al alta hospitalaria, el porcentaje de IFS desconocidos es bajo debido a la obligatoriedad de las variables hospitalarias de la HCE relacionadas con el IFS.

Sin embargo, después del análisis longitudinal, como podemos ver en la Figura 6.4, el porcentaje de sujetos con patrón nulo para la serie de IFS (Longitudinal-DQ = 5) fue solo del 9% para FBF y del 2% para ABF. Para EBF el porcentaje fue del 33% debido a la importante proporción de IFS desconocidos que impidió encontrar más sujetos con patrones válidos para la serie IFS, junto con los mayores requisitos para este tipo de lactancia. Además, obtuvimos un porcentaje significativo de sujetos (33% para FBF y 39% para ABF, no así para EBF) donde los datos retrospectivos eran completos y consistentes con los datos transversales (Longitudinal-DQ = 1) para los cuales pudimos calcular una duración puntual precisa de lactancia materna basada en el evento de cese. Para el resto de sujetos (con Longitudinal-DQ = 2, 3 o 4), solo se pudo utilizar la información transversal para el cálculo de duración.

A pesar de los problemas de DQ encontrados, la detallada meta-información de DQ obtenida permitió, en primer lugar, definir y aplicar adecuadamente criterios de tolerancia al cálculo de la duración de lactancia materna. En concreto, de los casos con algún problema longitudinal de DQ (97% para EBF, 67% para FBF y 61% para ABF), la tolerancia aplicada permitió recuperar el 66%, 87% y 97% de casos respectivamente para cada tipo de lactancia. Así mismo, la meta-información de DQ también guió la estimación de indicadores y permitió establecer los criterios de aseguramiento para cada indicador. Obteniendo de esta forma una población válida para el seguimiento de la alimentación infantil, excepto para los indicadores de seguimiento de EBF, como podemos observar en la Tabla 6.4.

El problema para medir la EBF es que, a diferencia de la leche de fórmula o los sólidos, la introducción de líquidos puede ser discontinua sin tener que condicionar necesariamente un cambio de tipo de lactancia, lo que puede generar problemas de inconsistencia. Además, los profesionales no registraron los líquidos con tanta regularidad como la introducción de leche de fórmula o la alimentación complementaria. El alto porcentaje de casos perdidos no aplica al indicador “EBF al alta hospitalaria” debido a que solo le aplica la información perinatal, con obligatoriedad de las variables de alimentación infantil en la HCE hospitalaria, y a la amplia formación y experiencia en lactancia materna de las enfermeras del hospital.

Además, podemos ver el efecto de la DQ en las estimaciones de indicadores en la Figura 6.5 para el caso particular del indicador de mediana de duración de lactancia materna. Los casos con DQ estándar tuvieron mejores tasas de lactancia materna, para los cuales obtuvimos un valor puntual de duración, a diferencia de los casos con problemas tolerables de DQ. Sus peores tasas se relacionaron con las decisiones tomadas por el grupo de expertos para determinar la duración puntual en este tipo de casos (punto medio del intervalo de cese de la lactancia materna para patrones de duración no censurada y el mínimo del intervalo para patrones de duración censurada).

Finalmente, podemos observar la evolución de los casos perdidos para el indicador “mediana de duración de FBF” en la figura 6.6, a lo largo del periodo de estudio, diferenciado entre dos centros de salud del Área de Salud V – Altiplano. De esta forma, se pueden construir indicadores propios de DQ que permiten prestar atención a la variabilidad de la calidad del registro entre distintos centros, que también puede ser aplicada a profesionales.

Por lo tanto, obtuvimos un IDR para la monitorización e investigación de la alimentación infantil desde el nacimiento hasta los 18 meses de edad, con datos procedentes de recolección rutinaria de HCE tanto de la atención hospitalaria como de la atención primaria de 2009 a 2018. A pesar de los problemas de DQ encontrados en los datos brutos, el enfoque de aseguramiento de DQ por indicadores propuesto en este trabajo, permitió obtener una estimación robusta de la mayoría de los indicadores con un porcentaje significativo de sujetos con información válida para la monitorización e investigación.

Para controlar los posibles sesgos de selección por la exclusión de sujetos con problemas no tolerables (casos perdidos), además de los datos de alimentación, el repositorio incorpora suficiente información perinatal ([García-de-León-Chocano et al., 2016](#)) para comparar la distribución de la población de sujetos válidos y sujetos perdidos según los factores relacionados con el abandono de la lactancia materna descritos en la literatura ([Mangrío et al., 2018](#); [Moore et al., 2016](#)) (edad materna, contacto piel a piel precoz, valoración del inicio de la lactancia, tipo de parto, edad gestacional,...).

Finalmente, la estimación de los indicadores (Tabla 6.4) fue consistente con los resultados previamente publicados ([García-de-León-González et al., 2011](#)). Además, utilizamos datos de un estudio multicéntrico de Resultados Informados por Pacientes, Patient Reported Outcomes, (PI09 / 90899) realizado en nuestra área de salud para una muestra de 241 madres que dieron a luz en el período 2009 - 2011 para comparar sus resultados con los

resultados de nuestros indicadores, también para el período 2009 - 2011, y obtuvimos valores similares para los indicadores.

### 6.5.1 Limitaciones

Solo se pudo obtener una duración precisa para sujetos sin problemas de DQ, sin embargo, dada la baja proporción de estos sujetos fue necesario incorporar sujetos con DQ tolerable donde la duración se obtuvo como una estimación dentro del intervalo de cese de la lactancia materna.

En cuanto a la extensión a otros entornos, para generar la base de reglas, los formularios de seguimiento de alimentación de la HCE deben estar compuestas por: 1) un recuerdo de 24 horas con los elementos necesarios para poder calcular el estado de alimentación según las definiciones de la OMS; y 2) suficientes hitos de alimentación para determinar los eventos de cese. Además, en caso de no contar con datos al alta hospitalaria, se debe considerar un tratamiento específico, ya que no se dispone de los datos de inicio de la lactancia (datos truncados).

No tenemos variables sociodemográficas en el conjunto de datos extraídos de la HCE excepto la edad para discutir posibles generalizaciones de los resultados clínicos. Sin embargo, menos del 15% de los niños no aparecen en la muestra, por lo que los resultados son aplicables a la mayoría de la población de la zona.

### 6.5.2 Trabajos futuros

La meta-información detallada y sistemática obtenida durante la evaluación de DQ nos permitirá implementar ciclos DMAIC (del inglés, *Define, Measure, Analyze, Improve, Control*) dentro de la metodología de la Gestión de la Calidad de Datos Total (TDQM, del inglés Total Data Quality Management) a partir de la definición, seguimiento y evaluación de los indicadores de DQ como el empleado en la figura 6.5. La definición de los indicadores de DQ estará orientada a abordar la importante área de mejora de DQ (Figura 6.5), centrándonos en reducir la variabilidad entre profesionales y centros sanitarios para los principales problemas detectados (Figuras 6.3 y 6.4) y en particular los relacionados con la EBF y con la completitud de la ABF a partir del 6º mes.

A partir de este repositorio se realizarán una serie de estudios epidemiológicos: 1) validez predictiva de las variables perinatales sobre la duración de diferentes tipos de lactancia materna; 2) correlación de la duración de la lactancia materna con la calidad de vida, la obesidad infantil, el asma infantil y el trastorno por déficit de atención con hiperactividad. Por último, otras fuentes de datos innovadoras, como las relacionadas con las aplicaciones de salud móvil, pueden proporcionar información valiosa para validar las evaluaciones de DQ y complementar los datos de HCE y deberían explorarse en investigaciones futuras (Bruun et al., 2016).

## 6.6 Conclusiones

La metodología proporcionada con este estudio permite una monitorización poblacional continua y fiable de los indicadores de alimentación infantil de la BFHI a partir de datos recolectados de HCE de manera rutinaria.

Los datos recolectados de forma rutinaria mostraron, en general, problemas de DQ, especialmente para EBF. Sin embargo, nuestro enfoque de aseguramiento de DQ por indicadores nos permitió incrementar el nivel de tolerancia a los errores de DQ y con ello, disminuir la probabilidad de sesgo de selección para su uso en el seguimiento de resultados. La mejora en la recopilación de datos para EBF debe desarrollarse en el hospital estudiado



# 7 El valor del repositorio en el Área de Salud V - Altiplano

En este capítulo se describe la utilidad del IDR en el Área de Salud V – Altiplano, como uno de los ejes centrales de la estrategia HQTQ para implementar las recomendaciones de la BFHI en toda el área de salud y mejorar la calidad de la atención materno-infantil durante los primeros 1000 días de vida. El diseño del repositorio se incluyó dentro de la definición de la propia estrategia HQTQ, permitiendo el adecuado seguimiento del PAAL y facilitando la obtención de resultados en salud, en la organización y la sociedad. Los resultados y el carácter innovador de la experiencia HQTQ posibilitaron la obtención del sello de Buenas Prácticas del Sistema Nacional de Salud (C8), además de propiciar el *scaling-up* del modelo de información perinatal a todo el Servicio Murciano de Salud (C9).

*El contenido de este capítulo es una adaptación y ampliación para la tesis de la publicación Pu3 –“El valor de la información para desarrollar e implementar Buenas Prácticas” para el Congreso Inforsalud 2016. Como resultado, el Área de Salud V - Altiplano obtuvo el Sello de Buenas Prácticas del Sistema Nacional de Salud en el marco de la Estrategia de Salud Sexual y Reproductiva, contribución C8. Adicionalmente, se inició el *scaling-up* del modelo de información perinatal a todo el Servicio Murciano de Salud, contribución extraordinaria C9. La publicación Pu3 y la realización de las contribuciones C8 y C9 fueron posibles gracias a los trabajos previos desarrollados en los proyectos de investigación HQTQ-1 y HQTQ-2.*

## 7.1 Introducción

La brecha existente entre las prácticas asistenciales y la evidencia clínica es una preocupación sobre la que actualmente existe un consenso internacional. Las buenas prácticas (BBPP) basadas en la evidencia son aquellas intervenciones o experiencias en el marco de líneas estratégicas de salud basadas en el mejor conocimiento científico disponible y que han demostrado ser efectivas, que puedan ser transferibles y representen un elemento innovador para el sistema sanitario. En la actualidad, el Sistema Nacional de Salud mantiene abiertas iniciativas para identificar y replicar BBPP. A nivel europeo existen convocatorias de proyectos de investigación para promover el *scaling-up* o ampliación a mayor escala de BBPP que hayan demostrado su efectividad en estudios piloto.

En el Área de Salud V – Altiplano, tras la acreditación del Hospital Virgen del Castillo como hospital BFHI en 2004 ([García-de-León-González et al., 2011](#)), se definió la experiencia “Hasta Que Tú Quieras” (HQTQ) como una experiencia de BBPP implementada como una intervención de mejora de la calidad asistencial en la atención a la alimentación infantil, que ampliaba el foco a la atención primaria (hasta los dos años) y adaptando la asistencia sanitaria a la mejor evidencia científica disponible, específicamente a los criterios de calidad propuestos por la estrategia internacional BFHI de la OMS-Unicef en toda el área de salud. Para su implementación, se definió el proceso de atención a la alimentación del lactante (PAAL) ([García-De-León-González et al., 2008](#)), en el marco de lo que actualmente se conocen como Procesos Asistenciales Integrados (PAI), cubriendo la asistencia a la alimentación del niño pequeño desde una perspectiva integradora y centrada en las necesidades de la madre y su hijo, garantizando la continuidad asistencial entre los dos niveles asistenciales.

La experiencia HQTQ, ha estado estrechamente relacionada desde su inicio con el diseño, construcción y explotación de la información del IDR de alimentación infantil. Precisamente, las necesidades informativas de HQTQ para el adecuado seguimiento del PAAL guiaron la definición y diseño del proceso de construcción de repositorios con calidad de datos asegurada. Así mismo, la información de calidad procedente del IDR se ha utilizado para el “aprendizaje de salud” de la organización, tal como lo propone el Instituto de Medicina de EE. UU. ([Institute of Medicine \(US\), 2007](#)) y otros, que utilizan datos recopilados de forma rutinaria para orientar la atención.

## 7.2 Metodología

Para dar el máximo valor a la información, el diseño, construcción y explotación del repositorio se incluyeron dentro de la definición de la propia estrategia asistencial HQTQ, a la que debía dar soporte. HQTQ fue diseñada durante los proyectos de investigación HQTQ-1 y HQTQ-2 como una estrategia “*bottom-up*” dentro del marco de la calidad total, basada en unos principios que enfatizan la importancia de orientar la organización a las necesidades de las personas (madres, bebés y familia), la gestión por procesos (el PAAL) y su orientación hacia los resultados. La metodología empleada en el diseño de HQTQ siguió un enfoque abierto y recursivo en el que confluyeron dos ámbitos de desarrollo. El primero, un ámbito clínico-estratégico que define el modelo de atención, su contenido y especificaciones. El segundo, un ámbito multidisciplinar, con representación de informáticos y clínicos, para dar valor a la información: prestando especial atención al modelado de la información y aportando una metodología para el aseguramiento de calidad de datos para el re-uso de datos de HCE en investigación y monitorización. Para desplegar este enfoque se desarrolló un marco de trabajo participativo formado por expertos del dominio clínico (grupo de mejora HQTQ), gestores, el de del servicio de informática del Área V de Salud de la Región de Murcia, autor de esta tesis, con un amplio conocimiento de los circuitos implementados en los sistemas de HCE del Área V, así como expertos en calidad de datos del Instituto ITACA de la UPV.

A continuación se detallan los elementos clave de la aportación de valor a la información procedente de HCE:

- En primer lugar, se definió el modelo de información para garantizar que la información clínica relevante estuviese disponible para los fines particulares de la reutilización de HCE, en este caso, para la monitorización de la atención integral al recién nacido y lactante y la investigación sobre la alimentación infantil hasta los dos años. El modelo de información se basó en los flujos de trabajo y especificaciones del proceso de atención (PAAL). Los responsables del grupo de mejora HQTQ y el responsable del servicio de informática del hospital, participaron activamente en el diseño de los circuitos y formularios de la HCE perinatal del sistema de información corporativo del hospital (Selene), así como en el diseño de los formularios de seguimiento infantil del sistema de información de atención primaria (OMI-AP). En ambos casos, se prestó especial atención a su adecuación a las directrices de calidad del PAAL y dar respuesta a sus necesidades informativas.
- En segundo lugar, a partir del modelo de información, se aplicó la metodología propia para la construcción de repositorios con calidad de datos asegurada definida en el capítulo 2 ([García-de-León-Chocano et al., 2015](#)), para obtener un IDR de alimentación infantil hasta los dos años con calidad asegurada. De este modo, el IDR habilitó una fuente de datos poblacional, continua y fiable para la monitorización del PAAL.
- En tercer lugar, se definió una propuesta de indicadores para la monitorización del PAAL, como un conjunto de indicadores de buenas prácticas de estrategias internacionales relacionadas con la atención infantil. Para asegurar la validez e idoneidad de los indicadores se realizó una revisión sistemática de la literatura y las prin-

cipales bases de indicadores. La propuesta de indicadores se presentó en el proyecto piloto para el MSSSI descrito en el capítulo anterior y se realizó en base a las siguientes características funcionales: 1) Su capacidad para detectar áreas de mejora así como para evaluar el impacto de intervenciones de mejora, 2) Su comparabilidad entre diferentes centros, autonomías y países, 3) Su comparabilidad a escala temporal para evaluar los avances y retrocesos de la BFHI y así avanzar en el conocimiento sobre el desarrollo de la implantación de la estrategia, 4) Su capacidad para identificar poblaciones de riesgo para diseñar intervenciones específicas, 5) Su factibilidad para su monitorización y evaluación.

### 7.3 Resultados

A continuación se describen los principales resultados de la experiencia de BBPP HQTQ en los distintos ámbitos donde la información obtenida del IDR ha tenido un papel clave:

En la organización:

- Sello de BBPP del Ministerio de Sanidad, Servicios Sociales e Igualdad (MSSSI) para la experiencia “HQTQ” en la convocatoria 2014 ([Área de Salud V - Altiplano, 2014](#)), en el marco de la Estrategia de Salud Sexual y Reproductiva. Obtuvo la calificación de “excelente”, debido, entre otros factores, a su novedoso enfoque basado en la información y en la participación.
- Re-acreditación BFHI del Hospital Virgen del Castillo en 2017
- Propuesta de indicadores de proceso y resultado: 128 indicadores que se agrupan en 11 aspectos del PAAL para la monitorización de la atención materno-infantil compuestos por: antecedentes familiares, entorno obstétrico, condiciones obstétricas, prácticas obstétricas, estado del bebé, contacto precoz, primera toma de lactancia materna, suplementación, problemas de lactancia, alimentación al alta y duración de la lactancia materna y alimentación complementaria.

En salud:

- Tras el mejora significativa de los principales indicadores de lactancia materna producida con la primera acreditación BFHI del Hospital Virgen del Castillo en 2004 ([García-de-León-González et al., 2011](#)), la experiencia HQTQ permitió mantener e incluso mejorar los importantes resultados en salud: 86.45% de recién nacidos con contacto precoz piel con piel tras el nacimiento (siendo del 73.50% para los casos de cesárea), 75.70% de recién nacidos recibieron la primera toma de lactancia materna durante la primera hora tras el nacimiento (siendo del 64.96% para los casos de cesárea), el 76.04% de los recién nacidos fueron alimentados con lactancia materna exclusiva desde el nacimiento hasta el alta, resultados que permitieron la re-acreditación BFHI del hospital en 2017. Respecto a los resultados de indicadores relacionados con la duración de lactancia de materna, como se detalló en el capítulo 6, el indicador de FBF menor de 6 meses (55.63%) fue superior a su indicador equivalente incluido en la última Encuesta Nacional de Salud (39%), así como las medianas de FBF y ABF. La mediana de duración de lactancia materna (ABF) se situó en 7 meses, manteniendo el nivel post-acreditación BFHI obtenido en 2004.
- Así mismo, los procesos de mejora-aprendizaje basados en la información quedaron impregnados en la organización y acabaron influyendo en otras áreas de la atención como el consumo de leches especiales (leches hidrolizadas para el trata-

miento de alergias a las proteínas de leche de vaca), con una diferencia en su consumo por 100 tarjetas sanitarias de 19.68 (IC 95% 20.13 – 19.24), una tercera parte de la media regional y el de consumo de antibióticos en población menor de dos años que se situó en 13 dosis diarias definidas por 1000 habitantes (DHD), cercano a la mitad de la media regional ([Área de Salud V - Altiplano, 2014](#)).

En la sociedad:

- La BFHI es una de las intervenciones más rentables ([Horton et al., 1996](#)), por lo que el mantenimiento de esta estrategia en el Área V de Salud desde 2010 hasta la actualidad, gracias a la experiencia de BBPP HQTQ, representa un importante coste-beneficio. Señalar a modo de ejemplo que si el consumo de leches especiales financiadas en el año 2012 en el conjunto de la Región de Murcia hubiera sido como en el Área de Salud V, la contención de gasto hubiese supuesto unos 676 000 €.
- Mejora en las medidas de percepción de la sociedad como la encuesta de calidad percibida en atención hospitalaria de la Región de Murcia en la que se evidencian niveles elevados de calidad percibida y satisfacción en las madres atendidas en el Hospital Virgen del Castillo durante el parto-nacimiento y hasta el momento del alta. Así mismo también ha habido un progresivo aumento de la participación comunitaria a través del grupo Vínculo, el grupo de madres de Yecla y Jumilla implicado en actividades de formación, investigación y especialmente en actividades de promoción de la lactancia, así como a través de las alianzas que se han establecido con instituciones sanitarias y municipales, tanto locales como regionales ([Área de Salud V - Altiplano, 2014](#)).
- Los resultados de la estrategia HQTQ junto con su importante componente participativo y comunitario, así como su importante red de alianzas con organismos de la salud y universidades, culminaron en la propuesta y aprobación de cambios legislativos en la Región de Murcia. En concreto: 1) “El derecho del bebé recién nacido a no ser separado de su madre tras el parto (contacto piel con piel inmediato e ininterrumpido), salvo por razones médicamente justificadas” y 2) “El derecho a la identidad sanitaria desde el momento de su nacimiento, como sujeto de la asistencia, con apertura de historia clínica.”

## 7.4 Discusión

En el contexto actual de escasez de recursos, es importante reforzar el modelo de las organizaciones y su capacidad para incorporar BBPP orientadas a mejorar su funcionalidad y sostenibilidad. Una de las claves del éxito de la BBPP HQTQ ha sido su capacidad de generar información poblacional con calidad de datos asegurada, procedente de datos de HCE. Esta información fue crucial para guiar los procesos de mejora-aprendizaje de la organización, tanto en el plano instrumental para evaluar su progreso e identificar oportunidades de mejora, como en el plano estratégico para impulsar alianzas, visibilizar sus resultados y compartir la experiencia en otros contextos. Para ello, la creación de grupos multidisciplinares de clínicos y expertos en tecnologías de la información para la salud permitió diseñar y construir el IDR con calidad de datos asegurada con un enfoque centrado en la mejora de la calidad asistencial. De este modo, la información obtenida del repositorio pudo guiar los procesos de mejora-aprendizaje de la organización, desplegando su información con calidad de datos asegurada para detectar áreas de mejora, evaluar el impacto de las intervenciones y evaluar los avances y retrocesos en la implementación de estrategias internacionales de salud infantil. Esta información es crucial para implementar y

monitorizar BBPP que mejoran la calidad de la atención, permiten trabajar con transparencia y promueven la autonomía de la familia en un campo de la salud con importantes lagunas informativas.

Así, el IDR de alimentación infantil se ha constituido como una fuente de datos poblacional, continua y fiable, que ha permitido que todo el proceso de evaluación y mejora del PAAL se apoye en los datos recogidos durante la asistencia sanitaria rutinaria, minimizando el impacto de sesgos o análisis parciales de la actividad de las organizaciones. El mantenimiento de los estándares de calidad de estrategias internacionales como la BFHI requieren un importante esfuerzo y compromiso por parte del hospital. Dado que un factor clave para el mantenimiento de la BFHI a lo largo del tiempo en una organización, es el seguimiento regular de los indicadores de proceso y de resultados (Pérez-Escamilla, 2007; Pérez-Escamilla & Chapman, 2012; Zakarija-Grković et al., 2018), es posible afirmar en este sentido, que el IDR ha sido un elemento clave en el mantenimiento de los estándares de la BFHI en el Hospital Virgen del Castillo desde 2004, con sucesivas re-acreditaciones (la última en 2017), así como un facilitador para su extensión a atención primaria, donde actualmente los 3 centros de salud del Área de Salud V – Altiplano, se encuentran en proceso de acreditación.

Por otro lado, el IDR de alimentación infantil ha sido fuente de información para trabajos de investigación sobre lactancia materna en el Área V, incluida la tesis en desarrollo “Lactancia Materna y calidad de vida relacionada con la salud en la población escolar”.

Por todo ello, podemos afirmar que la información de calidad obtenida del IDR ha sido un vector de cambio en la organización, contribuyendo en gran medida a la obtención del Sello de BBPP del Ministerio de Sanidad, Servicios Sociales e Igualdad en la convocatoria de 2014 (Área de Salud V - Altiplano, 2014).

Finalmente, el éxito de la experiencia HQTQ, donde la información obtenida del IDR ha sido uno de sus ejes centrales, propició la ampliación a escala (*scaling-up*) del modelo de información perinatal a todo el Servicio Murciano de Salud. La extensión del modelo de información perinatal, implementado en el HIS Selene, comenzó con un proyecto piloto en el Hospital Santa Lucía del Área de Salud II – Cartagena, donde se realizaron los ajustes necesarios para incorporar al modelo de información las necesidades específicas de hospitales de nivel 3. Posteriormente, desde dicho hospital se lideró su extensión al resto de hospitales con maternidad del Servicio Murciano de Salud, proporcionando con ello un modelo de información homogéneo para poder evaluar y explotar de forma exhaustiva la información poblacional sobre la salud de los recién nacidos.



# 8 Conclusiones y trabajo futuro

Este capítulo resume las principales conclusiones derivadas de esta tesis, así como los trabajos futuros que pueden ser de aplicación a partir de la misma.

## 8.1 Conclusiones

La generación de información de calidad, válida y fiable para evaluar y monitorizar las prácticas sanitarias supone un elemento gran valor para las organizaciones. Su obtención a partir de *real world data* procedente de recolección rutinaria de HCE implica importantes retos relacionados con la evaluación y aseguramiento de la DQ, pero proporciona una fuente poblacional y continua sobre la que investigar o monitorizar las prácticas sanitarias. A continuación, se detallan las conclusiones extraídas de la tesis:

**Co1 Ha sido posible diseñar un proceso general de construcción de repositorios con DQ asegurada a partir de datos de HCE.** Para ello, la adopción de un marco conceptual de DQ, basado en dimensiones de DQ, facilita la organización del proceso de la evaluación la calidad de datos biomédicos y la definición de los criterios de aseguramiento. Medir el efecto de los procedimientos sobre las distintas dimensiones permite una gestión estandarizada de la calidad de datos total (TDQM) del proceso, facilitando la evaluación de la configuración del proceso para la funcionalidad esperada para mejorar su diseño durante el proceso de construcción: En este estudio se comprobó que: 1) Medir el efecto de los procedimientos sobre las distintas dimensiones, midiendo métricas antes y después de cada procedimiento, permitió etiquetar cada caso con su calidad, especificando los problemas encontrados por dimensión y procedimiento, para un control preciso del proceso construcción sobre el repositorio final, 2) la incorporación de procedimientos secuenciales, sobre los que se aplicaron criterios de eficacia e integridad de los datos, ayudó a mejorar gradualmente la DQ, tanto en el repositorio perinatal como en el IDR completo, 3) los puntos de corte establecidos durante el proceso de construcción fueron útiles para incorporar las actividades de selección, análisis y corrección de datos, 4) los resultados de DQ permitieron identificar el importante área de mejora de DQ, lo que puede orientar acciones de mejora, tanto en el propio sistema de HCE, como en el proceso de registro de información de HCE realizado por los usuarios, y 5) la detallada meta-información de DQ obtenida permitió implementar los criterios de aseguramiento, así como establecer un adecuado nivel de tolerancia a errores de calidad de datos, aumentando la calidad final del repositorio de alimentación, al tiempo que se redujo la cantidad de casos descartados.

Mejorar gradualmente la DQ mejora la eficiencia del proceso de construcción y facilita el abordaje de los problemas de DQ más complejos como los derivados de la variabilidad y consistencia. En concreto, iniciar el proceso con el procedimiento de descarte de variables no informativas supone una mejora evidente en la eficiencia de los procedimientos posteriores y tratar en primer lugar los problemas de DQ relacionados con la integridad de la estructura permite generar una estructura canónica, homogénea y estable, la cual facilita el abordaje posterior de la integridad semántica del conjunto de datos. En este estudio, el descarte de variables no informativas fue cercano al 50%, con la consi-

guiente mejora de eficiencia procesos posteriores, así mismo, ha quedado de manifiesto que los problemas de contextualización y estabilidad temporal afectaron a buena parte de los casos. En concreto, al 100% de los casos perinatales y al 72.83 % de los casos de seguimiento en atención primaria. No abordarlos previamente dificultaría enormemente el análisis semántico posterior.

La definición de reglas para identificar determinados problemas de calidad de datos, especialmente los relacionados con la integridad semántica, así como la definición de los criterios para establecer un adecuado nivel de tolerancia a errores de DQ, requieren la participación de expertos del dominio clínico y expertos en tecnologías de la información con conocimientos de los formularios y circuitos de trabajo definidos en la HCE. En este estudio, la definición de reglas y criterios del repositorio de alimentación del Área de Salud V – Altiplano, se realizó con la participación el Jefe de Servicio de Pediatría, una estudiante de enfermería con beca en el hospital y el autor y co-directores de esta tesis.

*Esta conclusión responde a la pregunta de investigación PR1, cubre el objetivo O1 y fue obtenida a partir de las publicaciones Pu1, Pu2 y Pu6*

**Co2 El proceso de construcción definido inicialmente permitió construir el repositorio perinatal con DQ asegurada a partir de datos del sistema de HCE del hospital, Selene. Posteriormente, la aplicación del proceso de construcción sobre una nueva fuente de datos, las revisiones de salud infantil hasta los dos años, registradas en el sistema de HCE de atención primaria OMI-AP, permitió ampliar el repositorio generando un nuevo IDR con información integrada de ambos sistemas.**

El IDR integra información tanto de los aspectos clínicos que definen la condición de salud de los sujetos en cada contacto asistencial como los datos del contexto organizativo necesarios para evaluar el grado de adecuación de las prácticas asistenciales a los objetivos de calidad de la organización.

Los procedimientos de aseguramiento de la integridad semántica del proceso de construcción definido inicialmente siguieron una perspectiva transversal que se mostró útil para evaluar y asegurar la DQ del repositorio perinatal. La ampliación del proceso con procedimientos para el aseguramiento longitudinal de DQ se mostró eficaz para evaluar el registro longitudinal completo de alimentación infantil y proporcionó la meta-información necesaria para una estimación robusta de indicadores de alimentación infantil.

De esto modo, se obtuvo un IDR para la monitorización e investigación de la alimentación infantil desde el nacimiento hasta los 18 meses de edad, con datos procedentes de recolección rutinaria de HCE multi-fuente, tanto de la atención hospitalaria como de la atención primaria de 2009 a 2018.

*Esta conclusión responde a la pregunta de investigación PR1, cubre los objetivos O1, O2 y O3 y fue obtenida a partir de las publicaciones Pu1, Pu2 y Pu6*

**Co3 Las evaluaciones realizadas a las distintas fuentes de HCE relacionadas con la alimentación infantil durante esta tesis reflejaron problemas de DQ. A pesar de estos problemas, la detallada meta-información de DQ generada durante el proceso permitió al grupo de expertos definir criterios de aseguramiento para, en primer lugar, asegurar la integridad estructural del repositorio (criterios para la homogeneización, estabilización del repositorio y de-duplicación), y posteriormente asegurar la integridad semántica, ajustando el adecuado nivel tolerancia a errores de DQ. Esto permitió elevar el porcentaje de casos válidos hasta situarlo en torno al 90%, tanto de casos válidos pe-**

rinatales, como de población válida para el seguimiento de la mayoría de indicadores de alimentación infantil. Obteniendo de esta forma un IDR con un nivel de calidad de datos adecuado para su re-uso en investigación y monitorización de la asistencia sanitaria. En concreto, para el caso perinatal, el nivel de tolerancia establecido permitió recuperar el 58.35% de casos (de los 25.82% de casos con inconsistencia multivariante); por otro lado, para el seguimiento de la alimentación infantil, el nivel de tolerancia permitió recuperar el 66% de casos para EBF, el 87% para FBF y el 97% para ABF.

*Esta conclusión responde a la pregunta de investigación PR2, cubre los objetivos O2 y O3 y fue obtenida a partir de las publicaciones Pu1, Pu2 y Pu6*

**Co4 El uso de la norma UNE-EN ISO 13 940 se ha mostrado suficiente para dar el soporte necesario a la continuidad asistencial entre los dos niveles asistenciales del repositorio de alimentación del Área V, asegurando un sistema de conceptos común entre ambos niveles dentro del mismo repositorio.** Por otro lado, el uso de modelos clínicos estandarizados UNE-EN ISO 13 606, con uso concurrente de conceptos de la norma UNE-EN ISO 13 940, permitió integrar los repositorios del Área de Salud V – Altiplano y Hospital 12 de Octubre de Madrid en un único IDR creado a partir de instancias XML UNE-EN ISO 13 606, asegurando su interoperabilidad semántica. A partir de estos modelos fue posible definir un conjunto único de criterios y métricas para evaluar la DQ de manera conjunta y también fue posible construir indicadores de alimentación infantil basados en los elementos de los modelos clínicos estandarizados. Su adopción generalizada permitiría la evaluación y monitorización del grado de implementación de estrategias de salud materno-infantiles como la BFHI a partir de los indicadores definidos.

Sin embargo, a nivel local de una misma organización sanitaria, las ventajas de la estandarización ISO13 606 se contrarrestan la mayor complejidad de las instancias jerárquicas que genera como repositorio frente a la sencillez de formatos planos más eficientes y versátiles para el análisis de datos. Por otro lado, las herramientas de modelado ISO13 606 presentan limitaciones para establecer criterios de integridad multivariante de alto nivel como los definidos en esta tesis.

*Esta conclusión responde a la pregunta de investigación PR3, cubre los objetivos O4 y O5 y fue obtenida a partir de la publicación Pu4*

**Co5 Ha sido posible realizar una estimación robusta de los indicadores de alimentación infantil definidos por la OMS: “indicadores para evaluar las prácticas de alimentación del lactante y del niño pequeño (IYCF)”.** La estimación se realizó según el método actuarial, utilizando el cálculo de duración de lactancia materna realizado a partir del análisis longitudinal de DQ. El factor clave de la estimación robusta ha sido este análisis longitudinal de DQ que permitió extraer una detallada meta-información de DQ de la serie de datos transversales y retrospectivos de la HCE de alimentación infantil. A partir de esta información fue posible la identificación de patrones de duración válidos, con distintos niveles de tolerancia a problemas de DQ y ajustando la precisión del cálculo de la duración de lactancia materna de cada patrón a su DQ. Para este caso de estudio, el análisis longitudinal permitió identificar como afectaban los problemas de DQ encontrados durante la evaluación de las revisiones de salud de manera individual (entre el 19% y 30% de contactos según el tipo de lactancia), a la serie completa de revisiones de salud de cada sujeto, encontrando que la mayoría de sujetos tenía algún pro-

blema longitudinal de DQ (97% para EBF, 67% para FBF y 61% para ABF). A pesar de ello, el nivel de tolerancia establecido permitió recuperar el 66%, 87% y 97% de casos respectivamente para cada tipo de lactancia. Finalmente, la meta-información de DQ obtenida del análisis longitudinal facilitó el establecimiento de los criterios de aseguramiento, que permitieron obtener una población válida (cerca al 90%) para la mayoría de los indicadores IYCF implementados, excepto para los de seguimiento de EBF.

Para controlar los posibles sesgos de selección en la estimación de indicadores de alimentación infantil debidos a la exclusión de casos con problemas no tolerables (casos perdidos), además de los datos de alimentación, el repositorio incorpora suficiente información perinatal para comparar la distribución de la población de sujetos válidos y sujetos perdidos según los factores relacionados con el abandono de la lactancia materna.

Las estimaciones de los indicadores fueron consistentes con resultados previamente publicados, que destacan dentro del contexto nacional y coinciden con lo esperado en el hospital de estudio, acreditado como hospital BFHI en tres ocasiones.

A pesar de la relevancia y las ventajas de utilizar los datos de HCE para la monitorización e investigación de la alimentación infantil, a nuestro entender, este es el primer estudio que se ocupa de la evaluación de DQ para obtener una estimación robusta de los indicadores IYCF sobre un IDR poblacional basado en datos procedentes de recolección rutinaria de la HCE.

*Esta conclusión responde a la pregunta de investigación PR4, cubre los objetivos O2; O3 y O6 y fue obtenida a partir de las publicaciones Pu1, Pu2 y Pu6*

**Co6 Se ha evaluado la usabilidad e impacto del repositorio.** Por un lado, ha quedado demostrada la mejora de DQ durante los sucesivos procedimientos del proceso de construcción, primero en el repositorio perinatal y posteriormente en el de seguimiento de la alimentación hasta los dos años, al tiempo que se ha podido realizar la estimación robusta de indicadores de alimentación infantil sobre el IDR completo. Esto ha permitido la constitución de un repositorio poblacional, continuo y fiable para la investigación poblacional, que ha sido fuente de datos para trabajos de investigación sobre lactancia materna en el Área V, incluida la tesis en desarrollo “Lactancia Materna y calidad de vida relacionada con la salud en la población escolar”. Así mismo, la estimación robusta de indicadores de alimentación infantil a partir del registro longitudinal de HCE ha permitido habilitar un sistema de monitorización poblacional y continuo, con DQ asegurada. A partir del mismo, se ha podido evaluar el grado de implementación de la BFHI en el Área de salud V – Altiplano y monitorizar sus resultados a partir de datos de uso rutinario de HCE.

Por otro lado, el IDR y la habilitación del sistema de monitorización de indicadores han tenido un notable impacto en el Área de Salud V – Altiplano, estando íntimamente relacionado con el desarrollo de la buena práctica HQTQ y sus resultados en salud, en la organización y la sociedad. La información de calidad obtenida a partir del mismo ha sido un vector de cambio en la organización: Ha guiado los procesos de mejora-aprendizaje del Área V a partir de la detección de áreas de mejora y la evaluación del impacto de las intervenciones en salud, al tiempo que ha permitido evaluar los avances y retrocesos en la implementación de estrategias internacionales de salud infantil como la BFHI. Todo ello, propició la obtención del Sello de BBPP del Ministerio de Sanidad, Servicios Sociales e Igualdad en la convocatoria de 2014.

Del mismo modo, la información de calidad, poblacional y continua proporcionada por el IDR ha sido un factor clave para el mantenimiento de los estándares de la BFHI a lo largo del tiempo en el Área V, incluyendo la última re-acreditación BFHI del hospital en 2017, así como un facilitador para su extensión a atención primaria, donde actualmente

los 3 centros de salud del Área de Salud V – Altiplano, se encuentran en proceso de acreditación. En el plano estratégico permitió impulsar alianzas, visibilizar los resultados de HQTQ y compartir la experiencia en otros contextos. En este sentido, el éxito de la experiencia HQTQ propició la ampliación a escala (scaling-up) del modelo de información perinatal a todos los hospitales del Servicio Murciano de Salud, proporcionando un modelo de información homogéneo para poder evaluar y explotar de forma exhaustiva la información poblacional sobre la salud de los recién nacidos.

Todo ello, junto con el importante componente participativo y comunitario de la estrategia HQTQ y su importante red de alianzas con organismos de la salud y universidades, culminó en la propuesta y aprobación de cambios legislativos en la Región de Murcia. En concreto: 1) “El derecho del bebé recién nacido a no ser separado de su madre tras el parto (contacto piel con piel inmediato e ininterrumpido), salvo por razones médicamente justificadas” y 2) “El derecho a la identidad sanitaria desde el momento de su nacimiento, como sujeto de la asistencia, con apertura de historia clínica.”

*Esta conclusión responde a la pregunta de investigación PR5, cubre los objetivos O2; O3 y O6 y fue obtenida a partir de las publicaciones Pu1, Pu2, Pu3, Pu5 y Pu6*

## 8.2 Trabajos futuros

A continuación, se describen algunas de las líneas futuras de investigación relacionadas con los resultados de esta tesis:

- La construcción de los repositorios ha seguido el enfoque de la Gestión de la Calidad de Datos Total (*TDQM*), donde la mejora de la calidad de datos se ha limitado a las posibilidades de mejora de la DQ del propio proceso de construcción. Tras constatar el importante área de mejora de la DQ y la variabilidad de la misma entre profesionales y centros sanitarios, la cual se puede observar en indicadores de DQ como el diseñado en el último capítulo, se pretende seguir avanzando en la mejora de la DQ de los repositorios elaborando ciclos DMAIC (del inglés *Define, Measure, Analyze, Improve, Control*) orientados a reducir la variabilidad y mejorar la DQ desde el proceso de registro en la HCE. Para ello, se definirán indicadores de DQ que podrán ser medidos de manera continua. El análisis de sus resultados guiará la propuesta de acciones de mejora orientadas al proceso de registro en la HCE que realizan los profesionales sanitarios para, posteriormente, controlar y comprobar si generan mejoras en la DQ. De esta forma, el sistema de monitorización contará con una doble monitorización de indicadores clínicos y de DQ.
- A partir del IDR se realizarán una serie de estudios epidemiológicos: 1) validez predictiva de las variables perinatales sobre la duración de diferentes tipos de lactancia materna; 2) correlación de la duración de la lactancia materna con la calidad de vida, la obesidad infantil, el asma infantil y el trastorno por déficit de atención con hiperactividad. Adicionalmente, el IDR se está utilizando como una de las fuentes de datos para la tesis en desarrollo “Lactancia Materna y calidad de vida relacionada con la salud en la población escolar”.
- Para posibilitar estudios sobre la influencia de la lactancia materna en la población, como el descrito en el punto 2) del apartado anterior, se pretende ampliar el IDR con nuevas fuentes de datos, para completar la información de alimentación con información de episodios mórbidos (procedentes tanto del HIS hos-

pitalario Selene como del sistema de información de atención primaria OMI-AP) y consumo farmacéutico (procedente del sistema de receta electrónica).

- También se valorará la ampliación del IDR con fuentes de datos procedentes de aplicaciones de salud móvil, dado que pueden proporcionar información valiosa para validar las evaluaciones de DQ y complementar los datos de HCE con información proporcionada por el propio usuario o, en este caso, su familia.
- Por último, se pretende extender la metodología a otros ámbitos clínicos para la construcción de nuevos IDR para la investigación y monitorización de distintos ámbitos de la salud.

# Bibliografía

- Aizpuru, F., Latorre, A., Ibáñez, B., & others. (2012). Variability in the detection and monitoring of chronic patients in primary care according to what is registered in the electronic health record. *Fam Pract*, 29, 696-705.
- Ajetunmobi, O., Whyte, B., Chalmers, J., Fleming, M., Stockton, D., & Wood, R. (2014). Informing the «early years» agenda in Scotland: Understanding infant feeding patterns using linked datasets. *J Epidemiol Community Health*, 68, 83-92.
- American Academy of Pediatrics, Steering Committee on Quality Improvement and Management, American Academy of Pediatrics Committee on Practice and Ambulatory Medicine, & Hodgson ES, Simpson L, Lannon CM. (2008). Principles for the development and use of quality measures. *Pediatrics*, 121(2), 411-418.
- Área de Salud V - Altiplano. (2014). *Buenas prácticas en la Estrategia de Atención al Parto y Salud Reproductiva del SNS. 2014. Murcia 1. Hasta Que Tú Quieras (HQTQ). Lactancia Materna. Hospital Yecla.*  
[http://www.sanidad.gob.es/organizacion/sns/planCalidadSNS/pdf/BBPP\\_2014/BBPP\\_APySR\\_Murcia\\_1.pdf](http://www.sanidad.gob.es/organizacion/sns/planCalidadSNS/pdf/BBPP_2014/BBPP_APySR_Murcia_1.pdf)
- Bai, D. L., Wu, K. M., & Tarrant, M. (2013). Association between intrapartum interventions and breastfeeding duration. *J Midwifery Womens Health*, 58(1), 25-32.
- Bartsch, E., Park, A. L., Young, J., Ray, J. G., & Tu, K. (2018). Infant Feeding Practices Within a Large Electronic Medical Record Database. *BMC Pregnancy Childbirth*, 18, 1.
- Box, T. L., McDonnell, M., Helfrich, C. D., & others. (2010). Strategies from a nationwide health information technology implementation: The VA CART story. *J Gen Intern Med Suppl*, 1, 72-76.

- Bruun, S., Wedderkopp, N., Mølgaard, C., Kyhl, H. B., Zachariassen, G., & Husby, S. (2016). Using text messaging to obtain weekly data on infant feeding in a Danish birth cohort resulted in high participation rates. *Acta Paediatr*, 105(6), 648-654.
- Bundy, D. G., Marsteller, J. A., Wu, A. W., & others. (2012). Electronic health record-based monitoring of primary care patients at risk of medication-related toxicity. *JtComm J Qual Patient Saf*, 38, 216-223.
- Burnum, J. (1989). The misinformation era: The fall of the medical record. *Ann Intern Med*, 110, 482-484.
- Busck-Rasmussen, M., Villadsen, S., Norsker, F., Mortensen, L., & Andersen, A. (2014). Breastfeeding practices in relation to country of origin among women living in Denmark: A population-based study. *Matern Child Health J.*, 18(10), 2479-2488.
- Callendret, M., Gelbert-Baudino, N., Raskovalova, T., Piskunov, D., Schelstraëte, C., Durand, M., & others. (2015). Hospital practices and breastfeeding cessation risk within 6 months of delivery. *Arch Pediatr*, 22(9), 924-931.
- Cattaneo, A. (2012). Academy of Breastfeeding Medicine Founder's Lecture 2011: Inequalities and Inequities in Breastfeeding: An International Perspective. *Breastfeed Med*, 7(1), 3-9.
- Cattaneo, A., Burmaz, T., Arendt, M., & others. (2010). Protection, promotion and support of breast-feeding in Europe: Progress from 2002 to 2007. *Public Health Nutr*;, 13(6), 751-759.
- Chantry, C. J., Dewey, K. G., Peerson, J. M., Wagner, E. A., & LA, N.-R. (2014). In-hospital formula use increases early breastfeeding cessation among first-time mothers intending to exclusively breastfeed. *J Pediatr*, 164(6), 1339-1345.
- Chapman, D. J., & Pérez-Escamilla, R. (2009). US National Breastfeeding Monitoring and Surveillance: Current Status and Recommendations. *J Hum Lact*, 25(2), 139-150.

- Coloma, P. M., Schuemie, M. J., Trifiró, G., & others. (2011). Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: The EU-ADR Project. *Pharmaco epidemiol Drug Saf*, 20, 1-11.
- Cruz-Correia, R. J., Pereira Rodrigues, P., Freitas, A., Almeida, F. C., Chen, R., & Costa-Pereira, A. (2010). Data Quality and Integration Issues in Electronic Health Records. En *Information Discovery On Electronic* (pp. 55-96).
- Dentler, K., Cornet, R., ten Teije, A., Tanis, P., Klinkenbijn, J., Tytgat, K., & et al. (2014). Influence of data quality on computed Dutch hospital quality indicators: A case study in colorectal cancer surgery. *BMC Med Inform DecisMak*, 14-32.
- E. U. Project on Promotion of Breastfeeding in Europe. (2004). Protection, promotion and support of breastfeeding in Europe: A blueprint for action. *European Commission, Directorate Public Health and Risk Assessment, Luxembourg*.  
[http://europa.eu.int/comm/health/ph\\_projects/2002/promotion/promotion\\_2002\\_18\\_en.htm](http://europa.eu.int/comm/health/ph_projects/2002/promotion/promotion_2002_18_en.htm)
- European Centre for Disease Control Prevention. (2014). Data quality monitoring and surveillance system evaluation: A handbook of methods and applications. *Stockholm: ECDC*.  
<https://www.ecdc.europa.eu/sites/default/files/media/en/publications/Publications/Data-quality-monitoring-surveillance-system-evaluation-Sept-2014.pdf>
- European Committee for Standardization Health Informatics. (2007). System of Concepts to Support Continuity of Care. Part 1: Basic concepts. *CEN, European Committee for Standardization. CONTSYS. (EN ISO 13940-1)*.
- Euro-Peristat project with SCPE and Eurocat. (2010). European Perinatal health report. The health of pregnant women and babies in Europe in 2010. *European Perinatal Health Report*.  
[https://www.europeristat.com/images/doc/EPHR2010\\_w\\_disclaimer.pdf](https://www.europeristat.com/images/doc/EPHR2010_w_disclaimer.pdf)

- Foglia, R. P., Alder, A. C., & G, R. (2013). Improving perioperative performance the use of operations management and the electronic health record. *J Pediatr Surg*, 48, 95-98.
- García-de-León-Chocano, R., Muñoz-Soler, V., Sáez, C., García-de-León-González, R., & García-Gómez, J. M. (2016). Construction of quality-assured infant feeding process of care data repositories: Construction of the perinatal repository (Part 2). *Comput Biol Med*, 71, 214-222.
- García-de-León-Chocano, R., Sáez, C., Muñoz-Soler, V., García-de-León-González, R., & García-Gómez, J. M. (2015). Construction of quality-assured infant feeding process of care data repositories: Definition and design (Part 1). *Comput Biol Med*, 67, 95-103.
- García-de-León-Chocano, R., Sáez, C., Muñoz-Soler, V., Oliver-Roig, A., García-de-León-González, R., & García-Gómez, JM. (2021). Robust estimation of infant feeding indicators by data quality assessment of longitudinal electronic health records from birth up to 18 months of life. *Comput Methods Programs Biomed*, 207:106147.
- García-De-León-González, R., Bas-Peña, E., Oliver-Roig, A., Castillo-Gómez, C., Hernández-Martínez, M., García-De-León-Chocano, R., & et al. (2005). *Memoria del proyecto de investigación: Hasta Que Tu Quieras (HQTQ)*. [https://www.murciasalud.es/recursos/ficheros/302353-PIENCA05\\_13.pdf](https://www.murciasalud.es/recursos/ficheros/302353-PIENCA05_13.pdf)
- García-De-León-González, R., Oliver-Roig, A., Armando-Silva, A., Castillo-Gómez, C., Hernández-Martínez, M., García-De-León-Chocano, R., & et al. (2008). *Memoria del proyecto de investigación: Diseño de la atención integral al proceso de la alimentación del lactante. HQTQ-2*. [https://www.murciasalud.es/recursos/ficheros/302355-Informe\\_Final\\_Marzo.pdf](https://www.murciasalud.es/recursos/ficheros/302355-Informe_Final_Marzo.pdf)
- García-de-León-González, R., Oliver-Roig, A., Hernández-Martínez, M., Mercader-Rodríguez, B., Muñoz-Soler, V., Maestre-Martínez, M. I., & others. (2011). Becom-

- ing baby-friendly in Spain: A quality-improvement process. *Acta Paediatr*, 100(3), 445-450.
- Grol, R., & Grimshaw, J. (2003). From best evidence to best practice: Effective implementation of change in patients' care. *Lancet*, 362, 1225-1230.
- Grupo de trabajo de la Guía de Práctica Clínica sobre atención al parto normal. (2010). Guía de Práctica Clínica sobre la atención al parto normal. Plan de Calidad para el Sistema Nacional de Salud del Ministerio de Sanidad y Política Social. Agencia de Evaluación de Tecnologías Sanitarias del País Vasco (OSTEBA). Agencia de Evaluación de Tecnologías Sanitarias de Galicia (Avalia-t). *Guías de Práctica Clínica en el SNS, OSTEBA Nº 2009/01*.
- Guidry, V. T., Gray, C. L., Lowman, A., Hall, D., & Wing, S. (2015). Data quality from a longitudinal study of adolescent health at schools near industrial livestock facilities. *Ann Epidemiol*, 25(7), 532-538.
- Häggkvist, A. P., Brantsæter, A. L., Grijbovski, A. M., Helsing, E., Meltzer, H. M., & Haugen, M. (2010). Prevalence of breast-feeding in the Norwegian Mother and Child Cohort Study and health service-related correlates of cessation of full breast-feeding. *Public Health Nutr*, 13(12), 2076-2086.
- Halvorsen, M. K., Langeland, E., Almenning, G., Haugland, S., Irgens, L. M., Markestad, T., & others. (2015). Breastfeeding surveyed using routine data. *Tidsskr Nor Lægeforen*, 135(3), 236-241.
- Horton, S., Sanghvi, T., Phillips, M., Fiedler, J., Pérez-Escamilla, R., Lutter, C., & others. (1996). Breastfeeding promotion and priority setting in health. *Health Policy Plan*, 10(1093), 156-168.
- Institute of Medicine (US). (2007). Round table on Evidence-Based Medicine. *The Learning Healthcare System: Workshop Summary. Washington (DC): National Academies Press (US)*. <https://www.ncbi.nlm.nih.gov/books/NBK53494>

- Instituto de Salud Carlos III. (2013). *Manual Práctico de interoperabilidad semántica para entornos sanitarios basada en arquetipos*.  
<http://gesdoc.isciii.es/gesdoccontroller?action=download&id=29/11/2013-45c9ee530c>
- ISO. (2008). *ISO 13606:2008—Health informatics—Electronic health record communication*.
- Jeusfeld, M. A., Quix, C., & Jarke, M. (1998). Design and Analysis of Quality Information for Data Warehouses. *Proc of the 17th International Conference on Conceptual Modeling*, 349-362.
- Kahn, G. M., Raebel, M. A., Glanz, J. M., Riedlinger, K., & Steiner, J. F. (2012). A Pragmatic Framework for Single-site and Multisite Data Quality Assessment in Electronic Health Record-based Clinical Research. *Med Care*, 50:S21–9.
- Kahn, M. G., Callahan, T. J., Barnard, J., Bauck, A. E., Brown, J., Davidson, B. N., & others. (2016). A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *EGEMS (Wash DC)*, 4(1), 1244.
- Kahn, M. G., & Weng, C. (2012). Clinical research informatics: A conceptual perspective. *J Am Med Inform Assoc*, 19, e36-e42.
- Karr, A. F., Sanil, A. P., & Banks, D. L. (2006). Data quality: A statistical perspective. *Stat Meth*, 3, 137-173.
- Khare, R., Utidjian, L., Ruth, B. J., Kahn, M. G., Burrows, E., Marsolo, K., & others. (2017). A Longitudinal Analysis of Data Quality in a Large Pediatric Data Research Network. *J Am Med Inform Assoc*, 24(6), 1072-1079.
- Klann, J. G., Anand, V., & SM, D. (2013). Patient-tailored prioritization for a pediatric care decision support system through machine learning. *J Am Med Inform Assoc*, 20 (e2), e267-74.

- Klazinga, N., Fischer, C., & Ten Asbroek, A. (2011). Health services research related to performance indicators and benchmarking in Europe. *J Health Serv Res Policy*, 16 (Suppl 2), 38-47.
- Labbok, M. H., & Starling, A. (2012). Definitions of Breastfeeding: Call for the Development and Use of Consistent Definitions in Research and Peer-Reviewed Literature. *Breastfeed Med*, 7(6), 397-402.
- Labbok, M., & Krasovec, K. (1990). Toward Consistency in Breastfeeding Definitions. *Stud Fam Plann*, 21(4), 226-230.
- Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: a methodology for information quality assessment. *Inform Manage*, 40, 133-146.
- Li, R., Abela, L., Moore, J., Woods, L. M., Nur, U., Rachet, B., & others. (2014). Control of Data Quality for Population-Based Cancer Survival Analysis. *Cancer Epidemiol*, 38(3), 314-320.
- Liaw, S. T., Rahimi, A., Ray, P., & others. (2013). Towards an ontology for data quality in integrated chronic disease management: A realist review of the literature. *Int J Med Inform*, 82(1), 10-24.
- MacKenzie, S. L., Wyatt, M. C., Schuff, R., Tenenbaum, J. D., & Anderson, N. (2012). Practices and perspectives on building integrated data repositories: Results from a 2010 CTSA survey. *J Am Med Inform Assoc*, 19(1):119-24.
- Madnick, S. E., Wang, R. Y., Lee, Y. W., & Zhu, H. (2009). Overview and Framework for Data and Information Quality Research. *ACM J Data Inform Quality*, 1, 1, Article 2.
- Maldonado, J. A., Bosca, D., Moner, D., & Robles, M. (2009). LinkEHR-Ed: A multi-reference model archetype editor based on formal semantics. *Int J Med Inf*, 78;8:559-70.

- Mangrio, E., Persson, K., & Bramhagen, A. C. (2018). Sociodemographic, Physical, Mental and Social Factors in the Cessation of Breastfeeding Before 6 Months: A Systematic Review. *Scand J Caring Sci*, 32(2), 451-465.
- Martínez-Costa, C., & Schulz, S. (2014). Ontology content patterns as bridge for the semantic representation of clinical information. *Appl Clin Inform*, 5(3), 660-669.
- Mathias, J. S., Agrawal, A., Feinglass, J., & others. (2013). Development of a 5 year life expectancy index in older adults using predictive mining of electronic health record data. *J Am Med Inform Assoc*, (e1), e118-24.
- McMurry, A. J., Murphy, S. N., MacFadden, D., Weber, G., & others. (2013). SHRINE: Enabling Nationally Scalable Multi-Site Disease Studies. *Carter KW, editor. PLoS ONE*, 7;8(3):e55811.
- Meedya, S., Fahy, K., & Kable, A. (2010). Factors that positively influence breastfeeding duration to 6 months: A literature review. *Women Birth*, 23, 135-145.
- Melnik, BC. (2014). The potential mechanistic link between allergy and obesity development and infant formula feeding. *Allergy Asthma Clin Immunol*, 10(1), 37.
- Menárguez-Tortosa, M., Martínez-Costa, C., & JT, F.-B. (2012). A generative tool for building health applications driven by ISO 13606 archetypes. *J Med Syst*, 36(5), 3063-3075.
- Moore, E. R., & Bergman, N. (2007). Early skin-to-skin contact for mothers and their healthy newborn infants. *Cochrane Database Syst Rev* 2007, (3):CD003519.
- Moore, E. R., Bergman, N., Anderson, G. C., & Medley, N. (2016). Early Skin-To-Skin Contact for Mothers and Their Healthy Newborn Infants. *Cochrane Database Syst Rev*, 11(11).
- Murphy, S. N., Weber, G., Mendis, M., Gainer, V., Chue, H. C., Churchil, S., & Kohane, I. (2010). Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc*, 17(2):124–30.

- Nickel, N. C., Warda, L., Kummer, L., Chateau, J., Heaman, M., Green, C., & others. (2017). Protocol for establishing an infant feeding database linkable with population-based administrative data: A prospective cohort study in Manitoba, Canada. *BMJ Open*, 7(10): e017981.
- Observatorio de Salud de la Mujer y del Sistema Nacional de Salud. MSSSI. (2008). *Estrategia de atención al parto normal y nacimiento en el Sistema Nacional de Salud*. Ministerio de Sanidad, Servicios Sociales e Igualdad.
- Oliveira, P., Rodrigues, F., & Henriques, P. (2005). A formal definition of data quality problems. *MIT IQ Conference*, 10-12.
- Patterson, E. S., Zhang, J., Abbott, P., & others. (2013). Enhancing electronic health record usability in pediatric patient care: A scenario-based approach. *Jt Comm J Qual Patient Saf*, 39, 129-135.
- Paul, J. A., Chateau, J., Green, C., Warda, L., Heaman, M., Katz, A., & others. (2019). Evaluating the Manitoba Infant Feeding Database: A Canadian infant feeding surveillance system. *Can J Public Health*, 110, 649.
- Pérez-Escamilla, R. (2007). Evidence based breast-feeding promotion: The Baby-Friendly Hospital Initiative. *J Nutr*, 137(2), 484-487.
- Pérez-Escamilla, R., & Chapman, D. J. (2012). Breastfeeding protection, promotion, and support in the United States: A time to nudge, a time to measure. *J Hum Lact*, 28(2), 118-121.
- Pérez-Escamilla, R., Martínez, J. L., & Segura-Pérez, S. (2016). Impact of the Baby-friendly Hospital Initiative on breastfeeding and child health outcomes: A systematic review. *Matern Child Nutr*, 12(3), 402-417.
- Pérez-Escamilla, R., & Moran, V. H. (2016). Scaling Up Breastfeeding Programmes in a Complex Adaptive World. *Matern Child Nutr*, 12(3), 375-380.

- Perrine, C. G., Scanlon, K. S., Li, R., Odom, E., & LM, G.-S. (2012). Baby-Friendly hospital practices and meeting exclusive breastfeeding intention. *Pediatrics*, *130*(1), 54-60.
- Persell, S. D., Kaiser, D., Dolan, N. C., & others. (2011). Changes in performance after implementation of a multifaceted electronic-health-record-based quality improvement system. *Med Care*, *49*, 117-125.
- Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Commun ACM*, *45*(4), 212-218.
- Reed, M., Huang, J., Graetz, I., & others. (2012). Outpatient electronic health records and the clinical care and outcomes of patients with diabetes mellitus. *Ann Intern Med*, *157*, 482-489.
- Reimer, A. P., Milinovich, A., & Madigan, E. A. (2016). Data quality assessment framework to assess electronic medical record data for use in research. *Int. J. Med. Inform*, *90*, 40-47.
- Rollins, N. C., Bhandari, N., Hajeerbhoy, N., Horton, S., Lutter, C. K., Martines, J. C., & others. (2016). Why invest, and what it will take to improve breastfeeding practices. *Lancet*, *387*, 491-504.
- Röthlin, M. (2010). Management of Data Quality in Enterprise Resource Planning Systems. En *Josef Eul Verlag GmbH (ed.)*.
- Sáez, C., Gutiérrez-Sacristán, A., Kohane, I., García-Gómez, J. M., & Avillach, P. (2020). EHRtemporalVariability: Delineating temporal data-set shifts in electronic health records. *GigaScience*, *Volume 9, Issue 8*.
- Sáez, C., Martínez-Miranda, J., Robles, M., & García-Gómez, J. M. (2012). Organizing data quality assessment of shifting biomedical data. *Stud Health Technol Inform*, *180*, 721-725.
- Sáez, C., Moner, D., García-De-León-Chocano, R., Muñoz-Soler, V., García-De-León-González, R., Maldonado, J. A., & others. (2017). A Standardized and Data Quality

- Assessed Maternal-Child Care Integrated Data Repository for Research and Monitoring of Best Practices: A Pilot Project in Spain. *Stud Health Technol Inform*, 235, 539-543.
- Sáez, C., Pereira Rodrigues, P., Gama, J., Robles, M., & García-Gómez, JM. (2015). Probabilistic change detection and visualization methods for the assessment of temporal stability in biomedical data quality. *Data Min. Knowl. Discov.*, 29(4), 950-975.
- Sáez, C., Robles, M., & García-Gómez, JM. (2017). Stability metrics for multi-source biomedical data based on simplicial projections from probability distribution distances. *Stat Methods Med Res*, 26(1), 312-336.
- Sáez, C., Romero, N., Conejero, J. A., & García-Gómez, J. M. (2021). Potential limitations in COVID-19 machine learning due to data source variability: A case study in the nCov2019 dataset. *Journal of the American Medical Informatics Association, Volume 28, Issue 2*, 360-364.
- Sáez, C., Zurriaga, O., Pérez-Panadés, J., Melchor, I., Robles, M., & García Gómez, J. M. (2016). Applying probabilistic temporal and multi-site data quality control methods to a public health mortality registry in Spain: A systematic approach to quality control of repositories. *J Am Med Inform Assoc*, 23:1085-95.
- Safran, C., Bloomrosen, M., Hammond, W. E., & others. (2007). Toward a national framework for the secondary use of health data: An American Medical Informatics Association White Paper. *J Am Med Assoc*, 14, 1-9.
- Scott, A., Laam, L. A., Wary, A. A., & others. (2011). ProvenCare Perinatal: A model for delivering evidence/guideline-based care for perinatal populations. *JtComm J Qual Patient Saf*, 37(5), 229-239.
- Shankaranarayanan, G., & Cai, Y. (2006). Supporting data quality management in decision-making. *Decis Support Syst*, 42, 302-317.

- Toubiana, L., & Cuggia, M. (2014). Big Data and Smart Health Strategies: Findings from the Health Information Systems Perspective. *IMIA Yearbook*, 9(1):125-7.
- UNICEF. (2020). *UNICEF data: Monitoring the situation of children and women. Access the data: Infant and young child feeding*. <http://data.unicef.org/topic/nutrition/infant-and-young-child-feeding>
- United Nations Children's Fund, World Health Organization, World Bank, Early Childhood Development Action Network, The Partnership for Maternal, Newborn and Child Health. (2018). *Nurturing care framework for early childhood development: A framework for helping children survive and thrive to transform health and human potential*. Geneva. [https://nurturing-care.org/resources/Nurturing\\_Care\\_Framework\\_en.pdf](https://nurturing-care.org/resources/Nurturing_Care_Framework_en.pdf)
- Venancio, S. I., Saldiva, S. R., Escuder, M. M., & ER, G. (2012). The Baby-Friendly Hospital Initiative shows positive effects on breastfeeding indicators in Brazil. *J Epidemiol Community Health*, 66(10), 914-918.
- Verheij, R. A., Curcin, V., Delaney, B. C., & McGilchrist, M. M. (2018). Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. *J Med Internet Res*, 20(5), e185.
- Verschuuren, M., Achterberg, P. W., Gijzen, R., Harbers, M. M., & Vijge, K. P., E. ., Wilk EA Van D. (2012). *ECHI indicator development and documentation. Joint Action for ECHIM Final Report Part II*. ECHIM project partners;
- Victora, C. G., Bahl, R., Barros, A. J., França, G. V., Horton, S., Krasevec, J., & others. (2016). Breastfeeding in the 21st century: Epidemiology, mechanisms, and lifelong effect. *Lancet*, 387(475-90), 1024-1027.
- Walker, K. L., Kirillova, O., Gillespie, S. E., Hsiao, D., Pishchalenko, V., & others. (2014). Using the CER Hub to ensure data quality in a multi-institution smoking cessation study. *J Am Med Inform Assoc*, 21(6):1129–35.

- Wang, R. Y. (1996). Strong DM Beyond Accuracy: What Data Quality Means to Data Consumers. *J Manag Inform Syst*, 12, 5-34.
- Watson, A. J., O'Rourke, J., Jethwani, K., & others. (2011). Linking electronic health record-extracted psychosocial data in real-time to risk of readmission for heart failure. *Psychosomatics*, 52, 319-327.
- Weber, G. M., Barnett, W., Conlon, M., Eichmann, D., Kibbe, W., & others. (2011). Direct2Experts: A pilot national network to demonstrate interoperability among research-networking platforms. *J Am Med Inform Assoc*, 18:157–60.
- Weiner, M. G., & PJ, E. (2009). Toward reuse of clinical data for research and quality improvement: The end of the beginning? *Ann Intern Med*, 151, 359-360.
- Weiskopf, N. G., & Weng, C. (2013). Methods and dimensions of electronic health record data quality assessment: Enabling reuse for clinical research. *J Am Med Inform Assoc*, 20, 144-151.
- Weng, C., Appelbaum, P., Hripcsak, G., Kronish, I., Busacca, L., Davidson, K. W., & others. (2012). Using EHRs to integrate research with patient care: Promises and challenges. *J Am Med Inform Assoc*, 19(5), 684-687.
- Whitney, C. W., Lind, B. K., & Wahl, P. W. (1998). Quality Assurance and Quality Control in Longitudinal Studies. *Epidemiol Rev*, 20(1), 71-80.
- WHO. (1998). *Evidence for the Ten Steps to Successful Breastfeeding*. Geneva.
- WHO. (2017). *National implementation of the Baby-friendly Hospital Initiative*. Geneva. <http://apps.who.int/iris/bitstream/10665/255197/1/9789241512381-eng.pdf?ua=1>
- WHO, UNICEF. (1989). *Protecting promoting and supporting breastfeeding: The special role of maternity services*. WHO.
- WHO, UNICEF. (2007). Indicators for assessing infant and young child feeding practices. Part 1 Definitions. Washington: WHO/UNICEF.

- [https://www.unicef.org/nutrition/files/IYCF\\_updated\\_indicators\\_2008\\_part\\_1\\_definitions.pdf](https://www.unicef.org/nutrition/files/IYCF_updated_indicators_2008_part_1_definitions.pdf)
- WHO UNICEF. (2009). BFHI Section 1: Background and implementation. *WHO. Baby-friendly hospital initiative: revised, updated and expanded for integrated care. Geneva, 3, 2015.*
- WHO, UNICEF. (2018). *Implementation guidance: Protecting, promoting and supporting breastfeeding in facilities providing maternity and newborn services—The revised Baby-friendly Hospital Initiative. Geneva.*  
<https://apps.who.int/iris/bitstream/handle/10665/272943/9789241513807-eng.pdf?ua=1>
- Wise, M. E., & Lovell, C. (2013). Public health surveillance in the dialysis setting: Opportunities and challenges for using electronic health records. *Semin Dial, 26, 399-406.*
- Wood, N. K., & Woods, N. F. (2018). Outcome measures in interventions that enhance breastfeeding initiation, duration, and exclusivity: A systematic review. *MCN Am J Matern Child Nurs, 43(6), 341-347.*
- World Health Organization. (2002). *Innovative Care for Chronic Conditions: Building Blocks for Action. Geneva, World Health Organization.*  
<http://www.who.int/chp/knowledge/publications/icccglobalreport.pdf>
- World Health Organization. (2003). *Global Strategy for Infant and Young Child Feeding. World Health Organization, Geneva.* [http://www.who.int/child-adolescent-health/New\\_Publications/NUTRITION/gs\\_iycf.pdf](http://www.who.int/child-adolescent-health/New_Publications/NUTRITION/gs_iycf.pdf)
- Yih, W. K., Kulldorff, M., Fireman, B. H., Shui, I. M., Lewis, E. M., Klein, N. P., & others. (2011). Active surveillance for adverse events: The experience of the Vaccine Safety Datalink project. *Pediatrics, 127 (Suppl 1), S54-64.*

Zakarija-Grković, I., Boban, M., & S. Janković A. Čuže, T. B. (2018). Compliance With WHO/UNICEF BFHI Standards in Croatia After Implementation of the BFHI. *J Hum Lact*, 34(1), 106-115.



# Apéndice A

## Material suplementario del capítulo 3.

Tabla A-0.1 Variables informativas del repositorio perinatal

Variable	Descripción de la variable	Concepto clínico
VI_DeliveryType	Tipo de parto (vaginal, cesárea)	Intervenciones obstétricas
VI_DeliveryBeginning	Inicio de parto (espontáneo, inducido, etc)	Intervenciones obstétricas
VI_Anaesthesia1	Tipo de anestesia 1	Intervenciones obstétricas
VI_Anaesthesia2	Tipo de anestesia 2	Intervenciones obstétricas
VI_Tobacco	Tabaco (cigarrillos/día)	Hábitos tóxicos durante la gestión
VI_Alcohol	Consumo de alcohol (nunca, ocasional, etc)	Hábitos tóxicos durante la gestión
VI_Drugs	Consumo de drogas	Hábitos tóxicos durante la gestión
VI_Apgar1	Apgar al minuto	Reanimación neonatal
VI_Apgar5	Apgar a los 5 minutos	Reanimación neonatal
VI_Oxygen	Necesidad de oxígeno durante la reanimación neonatal (minutos)	Reanimación neonatal
VI_Ambu	Necesidad de bolsa de ambú durante la reanimación neonatal (minutos)	Reanimación neonatal
VI_Intubation	Necesidad de intubación durante la reanimación neonatal (minutos)	Reanimación neonatal
VI_CPR	Necesidad de masaje cardíaco durante la reanimación neonatal	Reanimación neonatal
VI_GastricLavage	Necesidad de lavado gástrico durante la reanimación neonatal	Reanimación neonatal
CTC_Time	Momento del primer contacto precoz piel con piel madre-hijo	Contacto precoz piel con piel
CTC_Place	Lugar del primer contacto precoz piel con piel madre-hijo	Contacto precoz piel con piel
CTC_NoContactReason_Combo	Justificación de por qué el primer contacto piel con piel se realiza después de la 2ª hora de vida	Contacto precoz piel con piel
CTC_NoContactReason_TXT	Justificación de por qué el primer contacto piel con piel se realiza después de la 2ª hora de vida (Texto libre)	Contacto precoz piel con piel
CTC_NB_Destination	Destino del recién nacido después del parto (con la madre, ingreso, etc)	Contacto precoz piel con piel
CTC_Together	¿La madre y el recién nacido ingresan juntos en maternidad?	Contacto precoz piel con piel
FEED_1BF_Time	Momento de la primera toma de lactancia materna	Inicio temprano de lactancia materna
FEED_1BF_Place	Lugar de la primera toma de lactancia materna	Inicio temprano de lactancia materna

1BF_Assesment	Valoración de la primera toma de lactancia materna	Inicio temprano de lactancia materna
ASSES_Breast	Valoración del pecho	Evaluación del pecho
ASSES_Breast_Injury1	Evaluación de lesiones en el pecho: herida abierta / fisura	Evaluación del pecho
ASSES_Breast_Injury2	Evaluación de lesiones en el pecho:: costras / cicatrización	Evaluación del pecho
ASSES_Breast_Injury3	Evaluación de lesiones en el pecho: ampollas	Evaluación del pecho
ASSES_Breast_Injury4	Evaluación de lesiones en el pecho:sangrado	Evaluación del pecho
ASSES_Breast_Injury5	Evaluación de lesiones en el pecho: eritema	Evaluación del pecho
ASSES_Breast_Pain1	Evaluación de dolor en el pecho durante el enganche	Evaluación del pecho
ASSES_Breast_Pain2	Evaluación de dolor en el pecho durante la toma	Evaluación del pecho
ASSES_Breast_Pain3	Evaluación de dolor en el pecho después de la toma	Evaluación del pecho
ASSES_Breastfeeding_ND1	Diagnóstico enfermería 1: Patrones de comunicación m-h adecuados	Diagnósticos enfermeros de lactancia
ASSES_Breastfeeding_ND2	Diagnóstico enfermería 2: Satisfacción materna con el proceso de amamantamiento	Diagnósticos enfermeros de lactancia
ASSES_Breastfeeding_ND3	Diagnóstico enfermería 3: Avidez del lactante por mamar	Diagnósticos enfermeros de lactancia
ASSES_Breastfeeding_ND4	Diagnóstico enfermería 4: Deglución de la mama sostenida y eficaz	Diagnósticos enfermeros de lactancia
ASSES_Breastfeeding_ND5	Diagnóstico enfermería 5: El niño aparece satisfecho tras las tetadas	Diagnósticos enfermeros de lactancia
FEED_Glucose	Lactancia materna suplementada con glucosa	Alimentación al alta (Recuerdo 24h)
FEED_Formula	Lactancia materna suplementada con formula artificial	Alimentación al alta (Recuerdo 24h)
FEED_AdminSupplement	Método de administración de la suplementación	Alimentación al alta (Recuerdo 24h)
FEED_Supplement_TXT	Motivo de administración de los suplementos (Texto libre)	Alimentación al alta (Recuerdo 24h)
FEED_Discharge1	Cantidad de lactancia materna al alta	Alimentación al alta (Recuerdo 24h)
FEED_Discharge2	Tipo de alimentación al alta	Alimentación al alta (Recuerdo 24h)
VI_MotherAge	Edad de la madre en el momento del parto	
VI_Parity	Paridad	
VI_Gestation	Número de fetos	
VI_Diabetes	Diabetes materna	
VI_Hypertension	Hipertensión arterial materna	
VI_HBV	Infección materna HB	
VI_HCV	Infección materna HC	
VI_HIV	Infección materna HIV	
VI_PregnancyVisits	Nº de revisiones durante la gestación	
VI_MaternalEducation	Asistencia a las clases de educación maternal	

VI_BF_Intention	Intención de lactancia materna	
VI_BF_IntentionDuration	¿Cuánto tiempo desea amamantar a su bebé?	
VI_Anaesthesiologist	Anestesiista	
DeliveryComments	Comentarios del parto (Texto libre)	
VI_GestationalAge	Edad gestación	
VI_Sex	Sexo del recién nacido	
VI_BirthWeight	Peso al nacimiento del recién nacido	Somatometría
VI_MaternityWeight	Peso al ingreso de la madre	Somatometría
VI_DischargeWeight	Peso al alta del recién nacido	Somatometría
VI_LossWeight	Perdida fisiológica de peso del recién nacido	Somatometría
VI_NB_Disease	Enfermedades del recién nacido al nacer	
VI_SocialRisk	Evaluación de riesgo social al alta del recién nacido	
VI_NutritionalRisk	Evaluación de riesgo nutricional al alta del recién nacido	
VI_NeurologicalRisk	Evaluación de riesgo neurológico al alta del recién nacido	
VI_NB_Death	Nacido muerto (feto cruz)	
VI_NB_Death_Data	Nacido muerto (feto cruz): fecha	
VI_NB_Death_Hour	Nacido muerto (feto cruz): hora	
VI_NB_Death_Sex	Nacido muerto (feto cruz): sexo	
VI_NB_Death_Weight	Nacido muerto (feto cruz): peso	
VI_NB_Death_Comments	Nacido muerto (feto cruz): comentarios (texto libre)	

#### Material suplementario del capítulo 4.

**Tabla A.2 Variables informativas del episodio de seguimiento**

Variable	Descripción	Concepto clínico
APESOPED	Peso actual (Kg)	Somatometría
ATALLAPE	Talla actual (cm)	Somatometría
IMCPEDIA	IMC Pediátrico	Somatometría
PPCEFRN	P.C. al nacer (cm)	Somatometría
PPESOALT	Peso al alta (kg)	Somatometría
PPESORN	Peso al nacer (Kg)	Somatometría

PTALLARN	Longitud al nacer (cm)	Somatometría
CFORMULA	Fórmula artificial	Hitos de alimentación
CLECVAC	Leche de vaca	Hitos de alimentación
CLIQUIDI	Líquidos	Hitos de alimentación
CCEREALE	Cereales sin gluten	Hitos de alimentación
CCERECEG	Cereales con gluten	Hitos de alimentación
CCLARA	Huevo completo	Hitos de alimentación
CFRUTAS	Frutas	Hitos de alimentación
CLEGUMBR	Legumbres	Hitos de alimentación
CPESCADO	Pescado	Hitos de alimentación
CVACUNOP	Pollo	Hitos de alimentación
CVERDURA	Verduras	Hitos de alimentación
CYEMAHUE	Yema de huevo	Hitos de alimentación
CYOGUR	Yogur	Hitos de alimentación
CEDADABA	Edad de abandono de la LM	Hitos de alimentación
CMOTABLM	Motivos de abandono LM	Abandono Lactancia Materna
LACTMAT	Tipo de lactancia materna	Recuerdo 24h
U24LMLIQ	24 h. ¿Ha tomado LM + agua, zumos, infusiones? (S/N)	Recuerdo 24h
U24PS	24 h. ¿Ha tomado alimentos sólidos o semisólidos? (S/N)	Recuerdo 24h
U24TLF	24 h. ¿Ha tomado leche de fórmula? (S/N)	Recuerdo 24h
U24TLM	24 h. ¿Ha tomado leche materna? (S/N)	Recuerdo 24h
ANTPERIN	Antecedentes perinatales	Antecedentes
PAPGAR1	Apgar al minuto	Antecedentes
PAPGAR5	Apgar a los 5 minutos	Antecedentes
PEDADGE	Edad gestacional al nacer (semanas)	Antecedentes
PFETOUNI	Feto único	Antecedentes
PINCINGR	Incidencias / Ingresos	Antecedentes
PLUGARP	Lugar de parto	Antecedentes
PREANEO	Reanimación	Antecedentes
PTIOPAR	Tipo de parto	Antecedentes
PDERTSOC	Derivación a trabajador social	Derivación
PEDERATE	Derivación a atención temprana	Derivación

PEDERESP	Derivación a especializada	Derivación
CHABINT	Hábito deposiciones	Hábitos
CHABSUE	Hábitos de sueño	Hábitos
CONLACA	Consejo lactancia artificial	Promoción
CONLACM	Consejo lactancia materna	Promoción
HNOGLUT	Consejo alimentación correcta	Promoción
HPRETPAS	Consejo de prevención de tabaquismo pasivo	Promoción
CPSINO	Conflicto parental	Riesgo Familiar
HIJONODE	Hijo no deseado	Riesgo Familiar
PFAMOES	Progenitor familia monoparental	Riesgo Familiar
PFAMONO	Familia monoparental	Riesgo Familiar
PHIADOPT	Adopción	Riesgo Familiar
CGUARDE	Va a guardería	Riesgo Guardería
COFERNR	Suplementos de Fe oral (niños de riesgo)	Riesgo Nutricional
FRIESGOD	Factores de riesgo Desarrollo Psicomotor	Riesgo Psicomotor
PDISMINU	Minusvalía (física, psíquica, sensorial)	Riesgo Psicomotor
PEINTEA	¿Se inicia protocolo de TEA?	Riesgo Psicomotor
PSIGATEA	Presenta signos de ALARMA DE TEA	Riesgo Psicomotor
CRIESGOP	Valoración del riesgo psicosocial	Riesgo Psicosocial
FALLOREV	No acude repetidamente a revisión de programa	Riesgo Psicosocial
PINMIGRA	Inmigración	Riesgo Psicosocial
CVALALP	Valoración de la alimentación	Valoración
CVALHIGI	Valoración de la higiene y cuidados	Valoración
DDESPSM	Valoración de la adquisición de habilidades	Valoración



# Apéndice B

## Material suplementario del capítulo 4.

Tabla B-0.1. Detalle de los arquetipos utilizados

Clase del modelo de referencia	Arquetipo
COMPOSITION	CEN-EN13606-COMPOSITION.InformeAlimentacion.v2
COMPOSITION	CEN-EN13606-COMPOSITION.InformeRecienNacido.v1
COMPOSITION	CEN-EN13606-COMPOSITION.RevisionNinoSano.v2
SECTION	CEN-EN13606-SECTION.EstadoBolsas.v1
SECTION	CEN-EN13606-SECTION.PatologiasGestacionales.v1
SECTION	CEN-EN13606-SECTION.PruebasAnaliticas.v1
ENTRY	CEN-EN13606-ENTRY.AlimentacionAlta.v1
ENTRY	CEN-EN13606-ENTRY.Alumbramiento.v1
ENTRY	CEN-EN13606-ENTRY.AntecedentesObstetricos.v1
ENTRY	CEN-EN13606-ENTRY.Apgar.v1
ENTRY	CEN-EN13606-ENTRY.ContactoPrecoz.v1
ENTRY	CEN-EN13606-ENTRY.ContextoFamiliar.v1
ENTRY	CEN-EN13606-ENTRY.ContextoHereditario.v1
ENTRY	CEN-EN13606-ENTRY.ControlEcografico.v1
ENTRY	CEN-EN13606-ENTRY.DestinoNeonato.v1
ENTRY	CEN-EN13606-ENTRY.EvaluacionProblemasLactanciaMaterna.v1
ENTRY	CEN-EN13606-ENTRY.Excrecion.v1
ENTRY	CEN-EN13606-ENTRY.ExploracionRecienNacido.v1
ENTRY	CEN-EN13606-ENTRY.FinParto.v1
ENTRY	CEN-EN13606-ENTRY.InicioParto.v1
ENTRY	CEN-EN13606-ENTRY.Isoinmunizacion.v1
ENTRY	CEN-EN13606-ENTRY.IsoinmunizacionMaterna.v1
ENTRY	CEN-EN13606-ENTRY.LactanciaMaterna.v1
ENTRY	CEN-EN13606-ENTRY.LlegadaPlanta.v1
ENTRY	CEN-EN13606-ENTRY.MadurezGestacional.v1

ENTRY	CEN-EN13606-ENTRY.Presentacion.v1
ENTRY	CEN-EN13606-ENTRY.ProblemaDiagnostico-PrincipalRN.v1
ENTRY	CEN-EN13606-ENTRY.ProblemaDiagnostico-SecundarioRN.v1
ENTRY	CEN-EN13606-ENTRY.ProcedimientosRN.v1
ENTRY	CEN-EN13606-ENTRY.Reanimacion.v1
ENTRY	CEN-EN13606-ENTRY.RevisionRecienNacido.v1
ENTRY	CEN-EN13606-ENTRY.Riesgo.v1
ENTRY	CEN-EN13606-ENTRY.Riesgo-InfeccionNeoantal.v1
ENTRY	CEN-EN13606-ENTRY.Riesgo-Metabolico.v1
ENTRY	CEN-EN13606-ENTRY.Riesgo-Neurologico.v1
ENTRY	CEN-EN13606-ENTRY.RiesgosNeonatales.v1
ENTRY	CEN-EN13606-ENTRY.SeguimientoGestacion.v1
ENTRY	CEN-EN13606-ENTRY.Somatometria.v2
ENTRY	CEN-EN13606-ENTRY.SuplementoLactanciaMaterna.v1
ENTRY	CEN-EN13606-ENTRY.TipoGestacion.v1
ENTRY	CEN-EN13606-ENTRY.TratamientoAltaRN.v1

**Tabla B-2. Detalle de los arquetipos re-utilizados del CMDIC**

Clase del modelo de referencia	Arquetipo
CLUSTER	CEN-EN13606-CLUSTER.Muestra.v1
CLUSTER	CEN-EN13606-CLUSTER.ResultadoPruebaDiagnosticaCualitativa.v1
CLUSTER	CEN-EN13606-CLUSTER.ResultadoPruebaDiagnosticaCuantitativa.v1
ENTRY	CEN-EN13606-ENTRY.DatosDemograficos.v1
ENTRY	CEN-EN13606-CLUSTER.Medicacion.v1
ENTRY	CEN-EN13606-CLUSTER.PlanDeActuacion.v1
ENTRY	CEN-EN13606-CLUSTER.ProblemaDiagnostico.v1
ENTRY	CEN-EN13606-CLUSTER.Procedimiento.v1
ENTRY	CEN-EN13606-CLUSTER.PruebaDiagnosticaDeLaboratorio.v1
ENTRY	CEN-EN13606-CLUSTER.Recomendacion.v1
ENTRY	CEN-EN13606-CLUSTER.Tratamiento.v1

En las figuras B.1 a B.8 se muestra una representación gráfica de los principales arquetipos utilizados. Dado el tamaño de la imagen resultante, se muestran diferentes imágenes resultado de la expansión de las distintas secciones del arquetipo principal

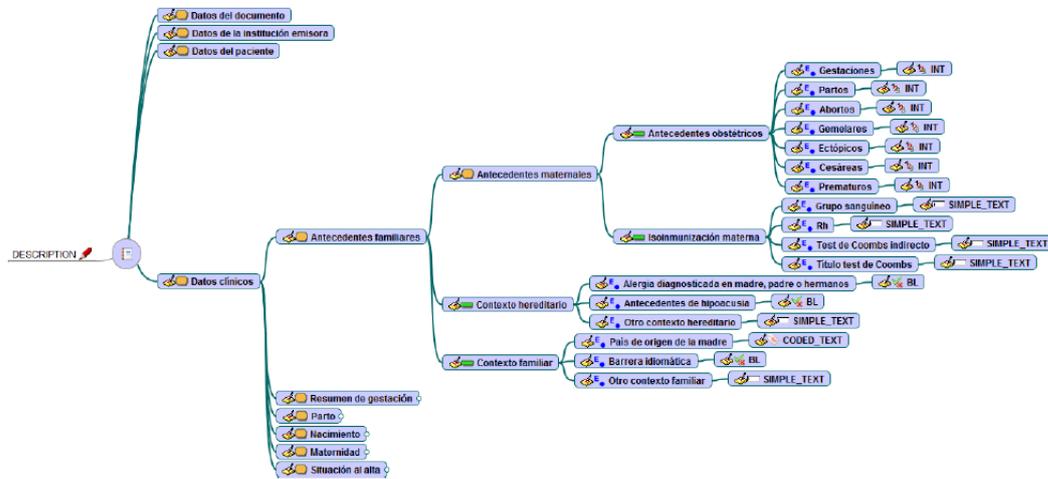


Figura B.1 Representación gráfica del arquetipo principal “Informe del recién nacido”, expansión de la sección Antecedentes familiares

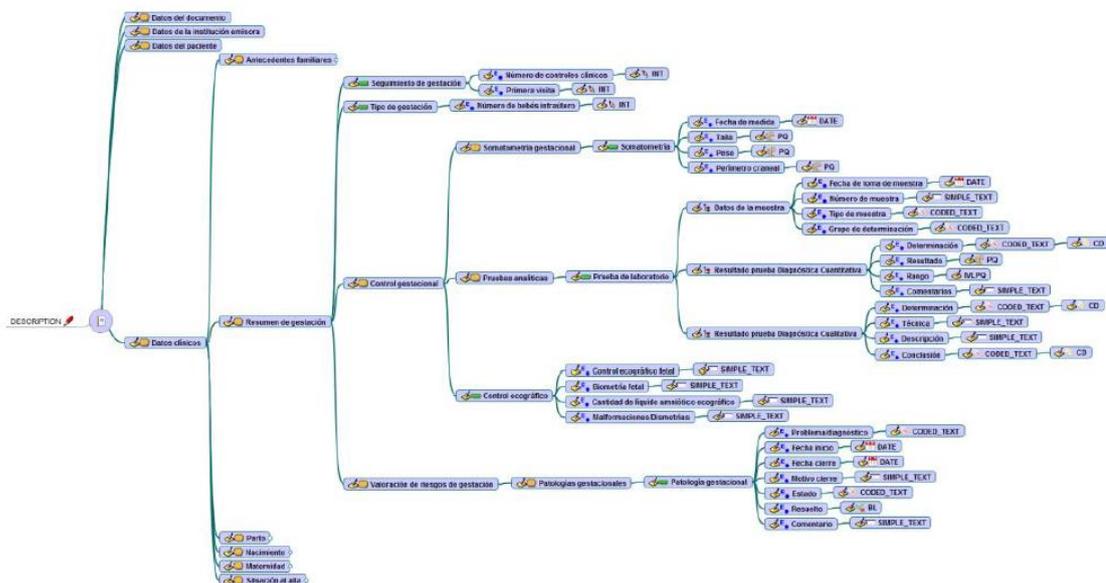


Figura B.2 Representación gráfica del arquetipo principal “Informe del recién nacido”, expansión de la sección Resumen de la Gestación

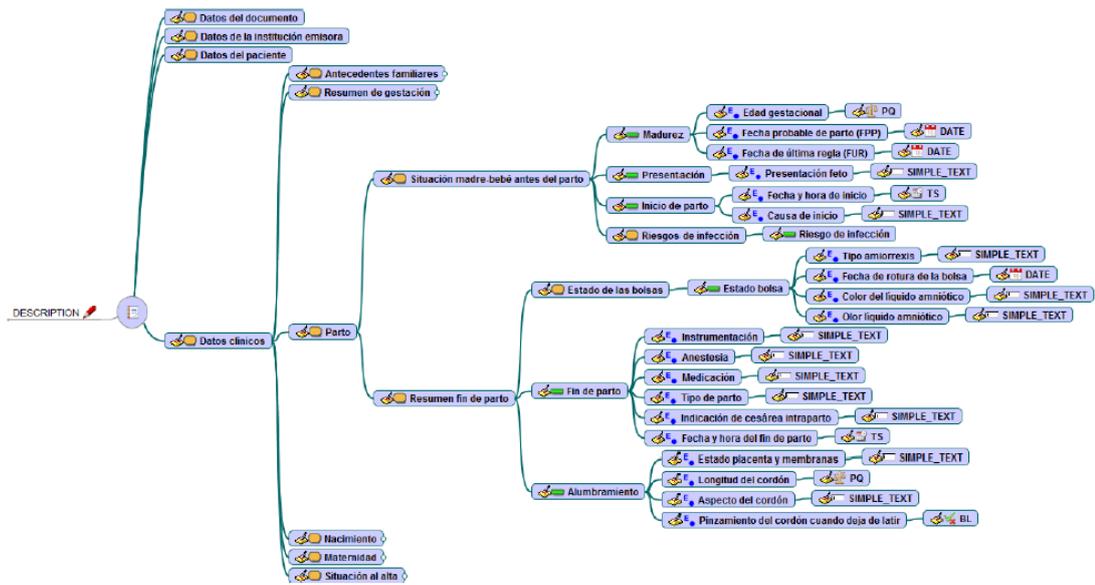


Figura B.3 Representación gráfica del arquetipo principal “Informe del recién nacido”, expansión de la sección Parto

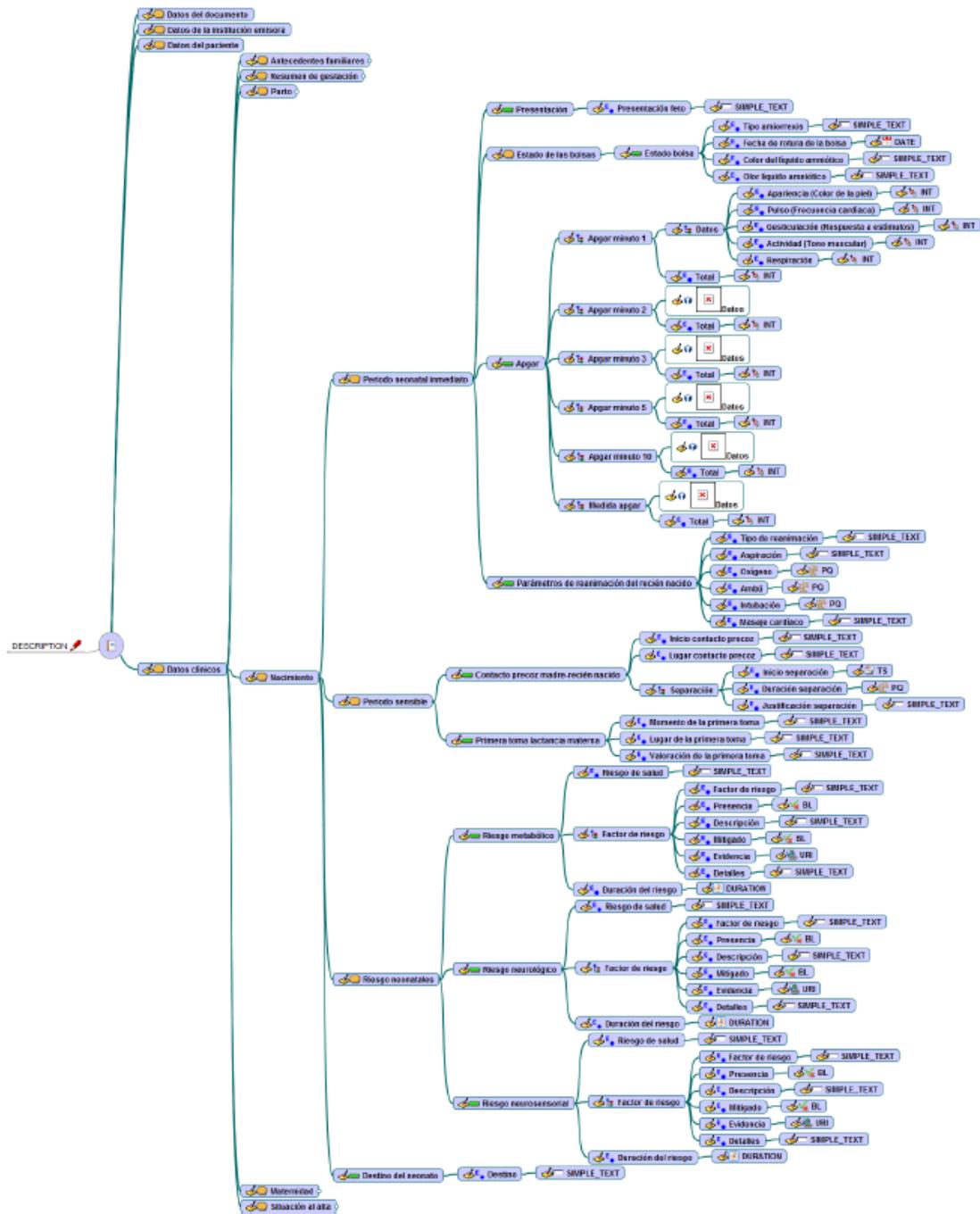


Figura B.4 Representación gráfica del arquetipo principal “Informe del recién nacido”, expansión de la sección Nacimiento

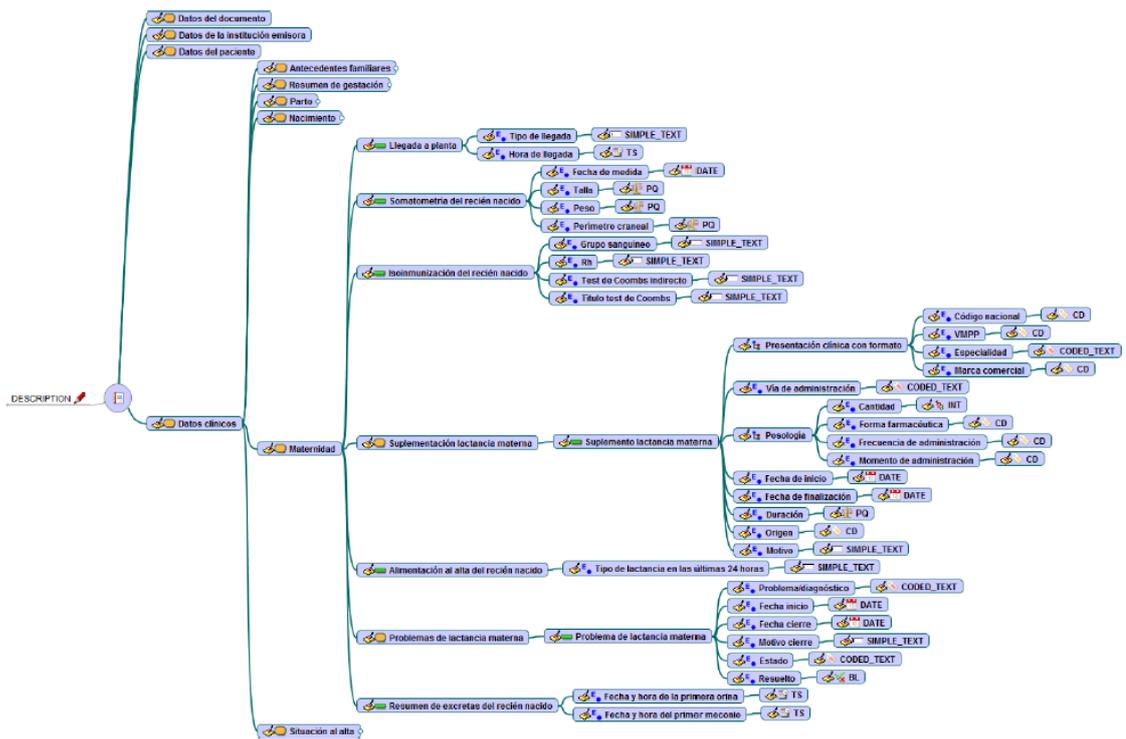


Figura B.5 Representación gráfica del arquetipo principal “Informe del recién nacido”, expansión de la sección Maternidad



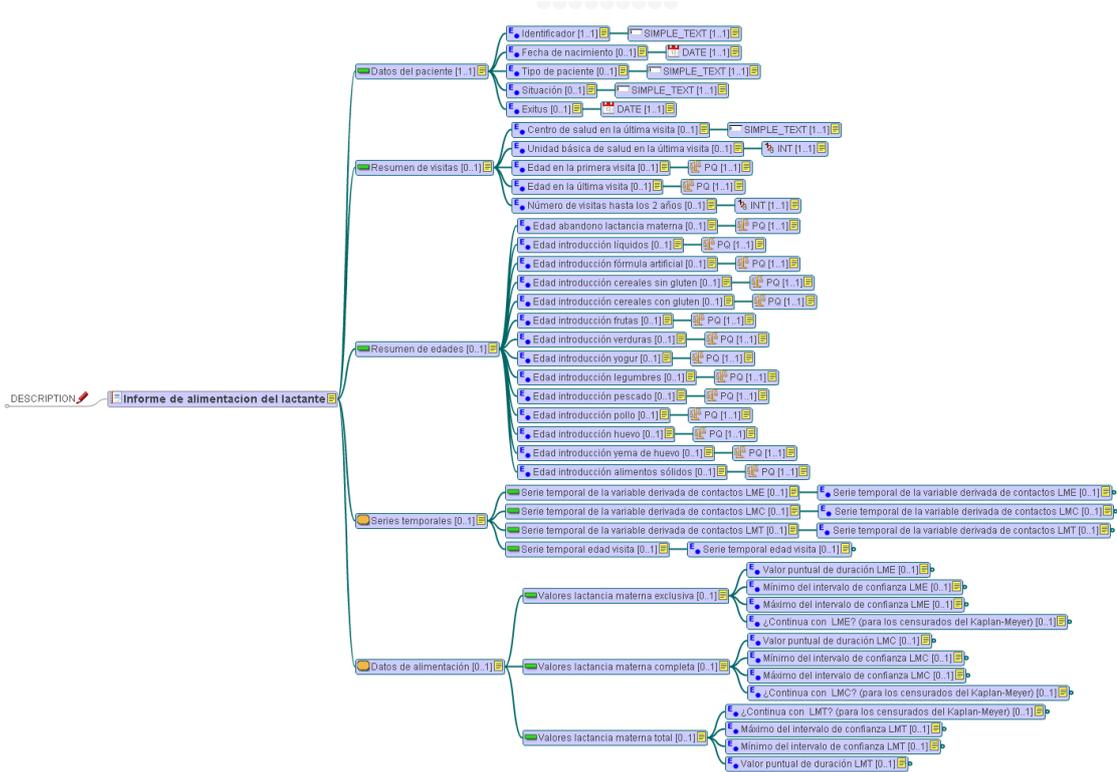


Figura B.7 Representación gráfica del arquetipo principal “Informe de Alimentación del Lactante”

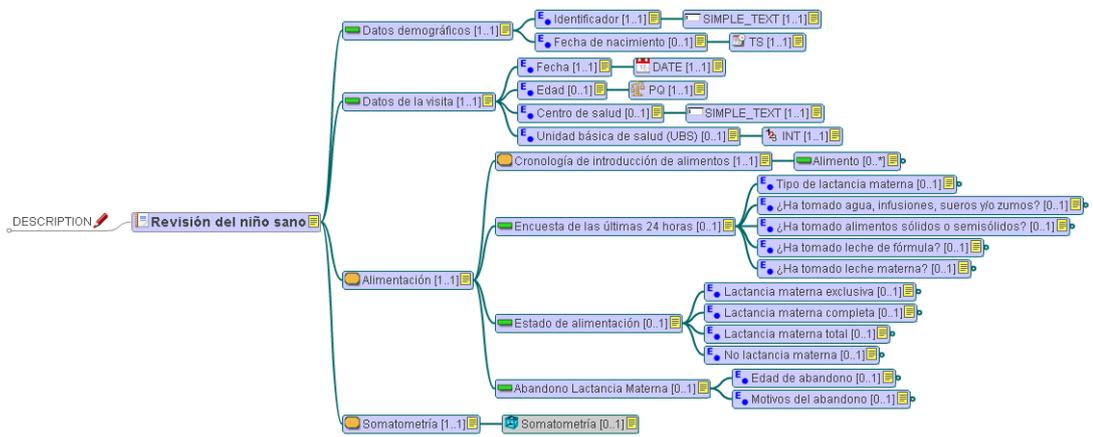


Figura B.8 Representación gráfica del arquetipo “Revisión de salud infantil” (Revisión de salud infantil del PANA)

# Apéndice C

## Material suplementario del capítulo 6.

**Tabla C.1 Tipos de lactancia materna de acuerdo a las 3 variables del estado de alimentación infantil (IFS). La primeras tres columnas representan la combinación de los diferentes valores de las 3 variables de IFS (ABF, FBF y EBF). La cuarta columna indica el tipo de lactancia materna resultante de esa combinación.**

ABF	FBF	EBF	Type of breastfeeding
No	No	No	El niño no estaba siendo amamantado (no lactancia materna)
Sí	Desconocido	Desconocido	El niño estaba recibiendo cualquier tipo de lactancia materna ( <i>Any breastfeeding</i> )
Sí	No	No	El niño estaba recibiendo lactancia materna parcial
Sí	Sí	Desconocido	El niño estaba recibiendo lactancia materna completa ( <i>Full breastfeeding</i> )
Sí	Sí	No	El niño estaba recibiendo lactancia materna predominante
Sí	Sí	Sí	El niño estaba recibiendo lactancia materna exclusiva
Desconocido	-	-	El tipo de lactancia materna era completamente desconocido