



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



ESCUELA TÉCNICA  
SUPERIOR INGENIEROS  
INDUSTRIALES VALENCIA

**TRABAJO FIN DE GRADO EN INGENIERÍA BIOMÉDICA**

# **APLICACIÓN EN UN HOSPITAL DEL ANÁLISIS DE PROCESOS SANITARIOS HACIENDO USO DE TECNOLOGÍAS DE MINERÍA DE PROCESOS**

AUTOR: Marta Mares García

TUTOR: Vicente Traver Salcedo

**Curso Académico: 2016-17**



## **Agradecimientos**

A todos los que han hecho posible este trabajo

A todos los que han estado conmigo estos 4 años



## Resumen

La importancia de la Sanidad Pública en el Estado del Bienestar hace que constantemente se busquen medidas políticas y técnica que mejoren la calidad de los servicios prestados y reduzcan su coste. Para ello, dentro de la variabilidad clínica inherente a la atención sanitaria, las investigaciones más recientes se esfuerzan en estandarizar los procesos de forma que se pueda atender a cada paciente de acuerdo a sus necesidades únicas, pero, a la vez, se sigan una serie de protocolos que optimicen los procedimientos sanitarios. A la hora de analizarlos, surge el problema de que los gestores no tienen conocimientos reales de cómo se desarrollan y cuál es la asistencia que están recibiendo de forma efectiva sus pacientes. Por este motivo, y dado que la cantidad de datos que almacenan y manejan los hospitales es cada vez más inabarcable a través de medios convencionales, surge la necesidad de aplicar métodos de vanguardia basados en las nuevas tecnologías. La minería de procesos permite a los gestores pasar de un corpus de datos desorganizados de un proceso concreto a una representación gráfica que, sintetizando la información, permite la comparación entre protocolo y atención real. En este proyecto, se ha utilizado la herramienta de minería de procesos PALIA para ayudar a los gestores del Hospital Universitario y Politécnico de la Fe a conocer y mejorar la gestión de las Consultas Externas del hospital. Tras estudiar el funcionamiento del área y los datos recogidos en ella, se obtuvieron los flujos de pacientes usuarios del área de Consultas Externas, permitiendo identificar anomalías y responder a preguntas concretas de sus gestores. Los resultados obtenidos manifiestan el potencial de las tecnologías de minería de procesos en la gestión de las Consultas Externas de los hospitales y abren paso a futuros proyectos en este u otros procesos sanitarios.

**Palabras Clave:** Minería de procesos, Consultas externas, Procesos sanitarios.



## Resum

La importància de la Sanitat Pública en l'Estat del Benestar fa que constantment es cerquen mesures polítiques i tècniques que milloren la qualitat dels serveis prestats i reduïsquen el seu cost. Per a açò, dins de la variabilitat clínica inherent a l'atenció sanitària, les recerques més recents s'esforcen a estandarditzar els processos de manera que es puga atendre a cada pacient d'acord amb les seues necessitats úniques, però, alhora, es segueixen una sèrie de protocols que optimitzen els procediments sanitaris. A l'hora d'analitzar-los, sorgeix el problema que els gestors no tenen coneixements reals de com es desenvolupen i quina és l'assistència que estan rebent de forma efectiva els seus pacients. Per aquest motiu, i atès que la quantitat de dades que registren i manegen els hospitals és cada vegada més inabastable a través de mitjans convencionals, sorgeix la necessitat d'aplicar mètodes d'avantguarda basats en les noves tecnologies. La mineria de processos permet als gestors passar d'un corpus de dades desorganitzades d'un procés concret a una representació gràfica que, sintetitzant la informació, permet la comparació entre protocol i atenció real. En aquest projecte, s'ha utilitzat l'eina de mineria de processos PALIA per a ajudar als gestors de l'Hospital Universitari i Politècnic de la Fe a conèixer i millorar la gestió de les Consultes Externes de l'hospital. Després d'estudiar el funcionament de l'àrea i les dades arreplegades en ella, es van obtenir els fluxos de pacients usuaris de l'àrea de Consultes Externes, permetent identificar anomalies i respondre a preguntes concretes dels seus gestors. Els resultats obtinguts manifesten el potencial de les tecnologies de mineria de processos en la gestió de les Consultes Externes dels hospitals i obren pas a futurs projectes en aquest o en altres processos sanitaris.

**Paraules Clau:** Minería de procesos, Consultes Externes, Processos sanitaris.





## Abstract

The Importance of Public Healthcare within the Welfare State asks for a constant search of political and technical measures that improve the quality of the service. With that in mind and taking into account the medical variability that healthcare inherently has to face, recent investigations try to standardize medical processes so every patients' unique needs are taken care of and, at the same time, a protocol can be followed to help optimize the process. When trying to analyze these proceedings it's difficult for the management branch to really know how their plans and protocols actually pan out and what assistance are effectively receiving the patients. Because of that, and taking into account that the sheer quantity of data that hospitals register and manage nowadays is simply unsurmountable by traditional means, the use of *avant-garde* information technology based methods are called for. Using process mining, managers are able to analyze a disorganized corpus of registered data in a graphical and meaningful way that allows a comparison between real practice and protocols. In this project, we use the process mining tool PALIA in order to help the managers of the *Hospital Universitario y Politécnio la Fe* to know and improve the organization of its External Consultations. After studying how this area works and the data registered in it, the patients' flows were produced, which surfaced anomalies and answered movements questions. The obtained results bring to light the potential of process mining technologies in the improvement of management of External Consultations and open the door for further projects and developments in this or other healthcare areas.

**Key Words:** Process Mining, External Consultations, Healthcare Processes.



# MEMORIA

APLICACIÓN EN UN HOSPITAL DEL ANÁLISIS DE PROCESOS SANITARIOS  
HACIENDO USO DE TECNOLOGÍAS DE MINERÍA DE PROCESOS

Marta Mares García



## Índice de la Memoria

1. INTRODUCCIÓN .....	3
2. OBJETIVOS .....	6
2.1 Objetivo principal .....	6
2.2 Objetivos secundarios .....	6
3. ESTADO DEL ARTE .....	7
3.1 Los procesos hospitalarios. Tipos .....	7
3.2 Análisis de procesos en salud .....	8
3.3 Minería de procesos en salud .....	10
<i>Tipos de minería de procesos</i> .....	11
<i>Presente y futuro de la minería de procesos</i> .....	12
<i>Herramientas de minería de procesos</i> .....	13
<i>Campos médicos de aplicación de la minería de procesos</i> .....	14
4. MATERIALES Y MÉTODOS .....	16
4.1 Entorno de trabajo .....	16
<i>Estructura del HUPLAFE y niveles de localización</i> .....	16
<i>Protocolo de las Consultas Externas</i> .....	17
4.2 Aplicación PALIA .....	20
<i>PALIA Web</i> .....	21
4.4 Fases del trabajo .....	26
<i>Metodología CRISP-DM</i> .....	26
<i>Fases del trabajo</i> .....	28
5. RESULTADOS Y DISCUSIÓN .....	31
5.1 Resultado del Análisis de los Datos .....	31
<i>El corpus de datos</i> .....	31
<i>Los campos del corpus</i> .....	31
<i>Clasificación de las citas</i> .....	36
<i>Calidad de los datos y errores</i> .....	39
5.2 Resultado de la Preparación de los Datos y el Modelado .....	47
<i>De cita a evento</i> .....	47
<i>Niveles de localización</i> .....	48
<i>Consultas Externas en PALIA Web</i> .....	49
6. CONCLUSIONES Y TRABAJO FUTURO .....	67
6.1 Trabajo Futuro .....	68
7. REFERENCIAS .....	69



## 1. INTRODUCCIÓN

El cuidado de la salud de las personas y la gestión de dicho cuidado ha evolucionado sensiblemente a lo largo del tiempo. De los retiros para tuberculosos a los hospitales o de los barberos a los médicos de familia. Estos cambios, que sirven como ejemplo de la enorme evolución de la sanidad, palidecen ante el mayor de todos ellos: El paso de la sanidad como lujo o caridad a la sanidad como derecho. El Estado del Bienestar, la forma en la que al menos nominalmente se organizan la mayoría de estados occidentales, surgió a finales del siglo XIX para paliar los efectos que sobre la población obrera había y estaba teniendo la industrialización. Seguros de accidentes, de jubilación y de desempleo, pero también de enfermedad y asistencia sanitaria que, aunque precarios y exclusivos en un principio, cristalizaron durante la Segunda Guerra Mundial. La experiencia de solidaridad tras el frente hizo que gran parte de las sociedades occidentales aceptasen corresponsabilidad en el bienestar de sus conciudadanos, pero, para bien o para mal, esta experiencia se ha diluido en el tiempo.

Hoy, más que nunca, los Estados del Bienestar se hallan en una encrucijada difícil de resolver. La fuerte crisis económica iniciada en 2007 unido al aumento del déficit hasta niveles más que peligrosos de los Estados, entre ellos el español, están poniendo a prueba los puntos de estrés de los Estados del Bienestar. Unos sistemas diseñados, hace más de medio siglo, con el modelo keynesiano de gasto público como motor del crecimiento como paradigma lo que pone en duda tanto su legitimidad como su idoneidad. Al fin y al cabo, lo público está cuanto menos demodé y como mínimo se exige una privatización endógena de la gestión pública aplicando modelos y estrategias privadas a la gestión pública. En otras ocasiones, se aboga directamente por la privatización y la reducción y salida del estado en todos aquellos bienes públicos de los que en principio se podría hacer cargo el mercado.

Sin embargo, creo que es importante destacar que el mercado jamás podrá hacerse cargo de la sanidad de forma completa. Numerosos estudiosos de los sistemas del bienestar (Pino y Lara, 2013) han señalado en términos cristalinos como la provisión de la sanidad se halla plagada de fallos de mercado de difícil solución. Por un lado, la complejidad y variabilidad de los numerosos procedimientos y condiciones médicas y la urgencia que puede llegar a tener en muchos casos, ponen de manifiesto que no existe información perfecta para el consumidor, un rasgo necesario para el mercado. Por ello, es necesaria la figura del médico que por esta asimetría de información tiene que tener necesariamente un carácter fiduciario de cara al paciente que sólo puede proporcionar mediante regulaciones, colegios profesionales sometidos al derecho público o su empleo directo por parte de un sistema nacional de salud. Si a esto le unimos los enormes costes de entrada y salida que rompen todo atisbo de competencia perfecta o las externalidades negativas que pueden tener, por ejemplo, una epidemia de gripe mal controlada, economistas y científicos sociales por igual manifiestan la dificultad de que el mercado provea por sí mismo la sanidad demandada y necesaria en el siglo XXI. Sobre todo, sus partes menos lucrativas o netamente deficitarias.

Ahora bien, esto no quiere decir que el sector público tenga carta blanca a la hora gestionar la sanidad y que pueda estar libre de toda crítica. Al contrario, en una sociedad de cambio acelerado y con nuevas técnicas médicas y de gestión que podrían mejorar el sistema sanitario, la gestión pública debe cuestionarse sus métodos y mejorarlos constantemente. En el caso de

un sistema nacional de salud como el español, en el que la atención sanitaria está garantizada bajo criterios de residencia, el combate se libra, sobre todo, en la reducción de los costes sanitarios sin perder la calidad del sistema. Esto quiere decir que no es ni suficiente, ni deseable, ni útil únicamente recortar el gasto, sino que es necesario optimizarlo, racionalizarlo y hacerlo eficiente y eficaz. Para ello, es necesario una aproximación directa a los problemas a los que se enfrenta la sanidad pública, haciendo cambios de pincel fino y no políticas de brocha gorda.

Uno de los problemas a los que se enfrentan tanto gestores como médicos a la hora de hacer más eficaz y eficiente el gasto público sanitario es el de la **variabilidad clínica**. La sanidad pública no es un servicio que se pueda producir y ofrecer en masa ya que, como veíamos anteriormente, los casos médicos son muchos y muy diferentes. Sin embargo, la atención sanitaria debe ser de igual calidad para todos y cada uno de los pacientes, institucionalizando las prácticas clínicas para que no dependa la calidad de esta atención ni de la hora, ni del médico, ni del hospital. Esto se consigue a través del establecimiento de **protocolos médicos** que estandaricen, en la medida de lo posible, tanto la práctica médica como la gestión del tiempo y los espacios. Esto no quiere decir que se reduzca al paciente a ser “un caso más” sino que libera tiempo y recursos preciosos para destinarlos a la medicina personalizada, un elemento de la sanidad pública cada vez más necesario cuando el propio paciente se considera a sí mismo un cliente con unas necesidades únicas a satisfacer.

Precisamente uno de los aspectos de mayor queja y con mayor margen de mejora en la sanidad española es el apartado de las esperas. Estos tiempos de espera o retrasos directamente afectan a varios aspectos de la sanidad pública, desde las citas y los retrasos en las salas de hospitales y ambulatorios, a los meses que pasan al pedir cita de un especialista o esperar una intervención. Si tenemos en cuenta el estrés psicológico y emocional de la enfermedad y que el riesgo de una mala toma de decisiones o de una toma de decisiones tardía o la tardanza en la realización de una prueba u operación puede ser enorme, en algunos casos, incluso la muerte, podemos entender cómo la reducción de los tiempos de espera en absolutamente todos los ámbitos de la sanidad debe ser una prioridad.

El **Hospital Universitario la Fe de Valencia**, cuyos datos se han utilizado en la elaboración de este trabajo, atiende una media de 4.000 pacientes diarios o, lo que es lo mismo, casi un millón y medio de pacientes anualmente ("La Fe – Hospital Universitari i Politècnic", 2017). Reducir el tiempo que pasan los pacientes en el hospital es, por lo tanto, un problema logístico importante que compete tanto a médicos como gestores. En este trabajo se ha analizado el tiempo que pasan los pacientes de las **Consultas Externas** esperando a ser atendidos, entre otras cosas. Este es un indicador importante de la calidad del sistema y una reducción de los mismos implicaría una mejora en la calidad de vida de los pacientes teniendo en cuenta que la atención en dichas consultas implica además un desplazamiento del paciente que le aleja de sus obligaciones cotidianas.

Para ello emplearé la **minería de procesos**, una metodología posibilitada y basada en la aplicación de tecnologías de la información y la comunicación de vanguardia para el análisis de procesos de cualquier tipo, en este caso, sanitarios. Siempre que ha habido una revolución tecnológica de enjundia, los avances en la productividad han sido tan dramáticos que han llevado a toda la industria y a toda la sociedad a una nueva era. Sin embargo, los avances tecnológicos del último siglo han generado una dualidad enorme entre la industria, que ha



aumentado de forma exponencial su productividad, y el sector servicios que, aún mejorando claro está, no ha llegado nunca a cumplir las expectativas de productividad generadas. En mi humilde opinión, la minería de procesos puede ser la herramienta crucial para llevar a la gestión sanitaria a esa nueva era, aprovechando al máximo las potencialidades de la revolución tecnológica, al dar un sentido y un propósito a la miríada de datos que estas mismas tecnologías nos permiten recoger. Espero que este trabajo sirva para mostrar, en la medida de lo posible, como puede hacerlo.

## 2. OBJETIVOS

Se diferencian a continuación el objetivo principal de este proyecto y los objetivos secundarios que se marcan en el camino hacia cumplir el primero.

### 2.1 Objetivo principal

El principal objetivo de este proyecto es mejorar la gestión de las Consultas Externas del Hospital Universitario y Politécnico de la Fe (HUPLAFE). Sus gestores establecen unos protocolos y marcan un orden, pero no saben realmente lo que está sucediendo en la práctica. Sí que recogen los datos y pueden ver dónde se encuentran los fallos más significativos, pero lo más costoso para ellos es ver el porqué de esas anomalías. No tienen una visión global del proceso que relacione servicios y pacientes y les permita establecer causas y efectos. Para ayudar a solucionar estos problemas de gestión, se hará una prueba de uso de una herramienta y se valorará si puede obtener información nueva de los datos, que ayude a los gestores a mejorar las Consultas Externas (CEX).

### 2.2 Objetivos secundarios

En el camino hacia alcanzar el objetivo principal, aparecen nuevos retos:

- Entender y conocer las CEX, sin duda, es fundamental para poder emprender el proyecto. Por una parte, experimentar el proceso como paciente, ya que es el centro del mismo, y, por otra parte, saber cuáles son los pasos que forman el proceso y quién y cuándo los realiza.
- Entender y comprender las necesidades y peticiones de los gestores del HUPLAFE para poder ofrecerles una herramienta que les sea útil y cumpla su propósito.
- Debido a la gran cantidad de datos generados por las CEX del HUPLAFE y su heterogeneidad, es primordial para su análisis entenderlos y familiarizarse con la base de datos. Lo cual facilitará la óptima explotación de los datos clínicos, extraer el mayor conocimiento de ellos y que éste pueda repercutir en la práctica clínica.
- Validar la herramienta de minería de procesos que emplearemos en el análisis de las CEX. Al ser una herramienta todavía en etapa de investigación, se quiere probar su comportamiento en entornos reales y conocer su utilidad en el manejo de procesos asistenciales. Adaptando la herramienta a este tipo de datos, podremos mejorarla de cara a futuras aplicaciones en el sector salud.
- El proyecto requiere una comunicación entre los funcionarios del HUPLAFE y los investigadores de la UPV. Debido a esta dualidad, se marca como objetivo favorecer la interacción entre ambos sectores y el entendimiento de los mismos.
- Durante el desarrollo del proyecto se necesitará hacer uso de otras herramientas a parte de la propia de minería de procesos, por ello, se precisa saber identificarlas y aprovecharlas en la etapa del proyecto que sea conveniente.

### 3. ESTADO DEL ARTE

Para poder entender la gestión de procesos y el análisis de los mismos, primero se ha de conocer en qué se fundamenta un proceso y las características que justifican el interés de su estudio.

Diversos autores coinciden en definir los **procesos** como el conjunto de actuaciones, actividades interrelacionadas, decisiones y tareas que requieren recursos determinados e implican valor añadido, con miras a obtener ciertos resultados que satisfagan plenamente los requerimientos del usuario y las metas de la organización, a la vez que se consideran el punto de concreción de los indicadores diseñados para el control y la gestión (Hernández Nariño y Nogueira Rivera, 2012).

Los procesos poseen dos **características** relevantes que justifican la necesidad de estudiarlos:

- La variabilidad. Cada vez que se repite el proceso, las actividades que lo componen se ejecutan de forma ligeramente distinta, provocando la variabilidad de los resultados. Dicho de otra forma, dos outputs del proceso nunca son exactamente iguales.
- La repetitividad. Los procesos se crean como modelo para producir un resultado y poder obtener ese resultado una y otra vez. Esta característica permite trabajar sobre el proceso y mejorarlo, ya que, después de cada cambio en el proceso, la evolución de los resultados dará información relativa al éxito del mismo.

Un **hospital**, como toda organización, busca ofrecer **servicios de calidad** que mantengan un cierto nivel de calidad de vida de sus usuarios. En el caso de un hospital, tanto la práctica clínica como la organización del centro influirán en la salud del paciente. La prestación de servicios hospitalarios de calidad depende de la ejecución adecuada y eficiente de los procesos (Rojas y otros, 2016). Para ello es importante que todos los procesos que se llevan a cabo en el hospital estén definidos y sigan un control por parte de los expertos, con el objetivo de analizar si se cumplen los protocolos y si éstos pueden mejorarse.

#### 3.1 Los procesos hospitalarios. Tipos

Los procesos hospitalarios son una serie de actividades dirigidas a diagnosticar, tratar y prevenir cualquier enfermedad con el fin de mejorar la salud de los pacientes (Rojas y otros, 2016). Es una definición general que coincide con el propósito principal del hospital y su función. No obstante, los procesos hospitalarios no son solo la prestación médica que se recoge en las guías clínicas, son apoyados por **actividades clínicas y no clínicas**. Éstas están ejecutadas por diferentes tipos de personal (médicos, enfermeros, especialistas técnicos, celadores, administrativos) y pueden variar de una organización a otra. Se sabe que los procesos de salud son **altamente dinámicos**, complejos, ad hoc y cada vez más multidisciplinarios.

La mejora de los procesos sanitarios podría tener un alto impacto en la vida de los pacientes. Sin embargo, mejorarlos no es tarea fácil ya que hay numerosos **factores** a tener en cuenta. Siempre hay la necesidad de reducir el coste de los servicios y mejorar la capacidad de satisfacer la demanda, reducir los tiempos de espera de los pacientes, mejorar la productividad de los recursos y aumentar la transparencia de los procesos.

Dumas y Aalst (2005), Rebuge y Ferreira (2012) y Kaymak y Mans (2012) especifican la existencia de **dos tipos de procesos hospitalarios**: Procesos de Tratamiento Médico (parte clínica) y Procesos Organizativos (parte administrativa).

Los **Procesos de Tratamiento Médico** son los procesos clínicos responsables de la gestión de la salud de los pacientes, incluyendo tareas que van desde el diagnóstico hasta la ejecución de acciones para curar o aliviar la enfermedad de cada paciente. Muchos de estos procesos están recogidos en las Guías de Práctica Clínica o son un conjunto de varias de ellas.

Los **Procesos Organizacionales** se centran en la comprensión organizacional de los procesos, recogiendo información colaborativa de los profesionales y sus unidades organizativas. Por ejemplo, esto incluye asignar tareas por turnos y transferir información y conocimientos médicos entre los diferentes tipos de personal o servicios del hospital.

**Otra clasificación** paralela es la que proponen Mans y van der Aals (2015). Dejan de lado los procesos compuestos por las decisiones individuales que toman los profesionales médicos respecto al diagnóstico y tratamiento de los pacientes (Procesos de Tratamiento Médico de la clasificación anterior), y se concentra en los **procesos operativos** (Procesos Organizacionales de la clasificación anterior). Estos procesos se ocupan de la logística de los procesos de trabajo, es decir, se centran en la orquestación y la gestión de los procesos de atención en lugar de las actividades individuales.

Los procesos operativos se subdividen en dos clases principales. La **Atención Electiva**, se refiere al cuidado y al tratamiento que se puede retrasar durante días o semanas sin dejar de ser médicamente correcto (procedimientos estándar, rutinarios y no rutinarios programados). Por el contrario, la **Atención No Electiva** representa los tratamientos médicos inesperados para el paciente y que deben planificarse con poca antelación (atención urgente y emergencias médicas) (Rojas y otros, 2016).

### 3.2 Análisis de procesos en salud

Hasta ahora se está exponiendo la idea de que para ofrecer servicios hospitalarios de calidad la gestión se ha de centrar en los procesos, no obstante, los procesos no han sido siempre el foco de la gestión hospitalaria.

Anterior a la gestión por procesos existía en los hospitales una **gestión vertical hospitalaria clásica**. Una estructura organizativa basada en una visión compartimental, especializada, orientada a los servicios, que no tenía una visión integral de la asistencia hospitalaria (Cabo Salvador, 2017). Se centraba en el funcionamiento de cada departamento de forma aislada, sin tener en cuenta que el flujo de pacientes que recibía provenía de otro servicio y las decisiones de uno podían influir en el otro, al estar relacionados. Como consecuencia se podían generar demoras, duplicidades e incremento de costes.

En el lado contrario se presenta la **gestión por procesos** como una gestión horizontal y matricial que tiene como centro el paciente y la asistencia que recibe dentro del hospital. La cadena asistencial está constituida por diferentes actividades que forman el proceso, en las que

intervienen profesionales de distintos servicios que necesitan estar coordinados. Para conocer cómo se diseña y ejecuta, se necesita una gestión por procesos que tenga en cuenta la relación entre los subprocesos del hospital: subprocesos de gestión (administrativos, alta...), subprocesos de diagnóstico y tratamiento (protocolos médicos y de enfermería...), subprocesos de apoyo (hostelería, lavandería, limpieza...), etc.

En la actualidad, centrándose las miradas de los gestores en los procesos, el potencial de la minería de procesos la sitúa como una herramienta prometedora en la mejora de la gestión por procesos. No obstante, antes de centrarnos en estas técnicas, se exponen a continuación **estrategias anteriores a la minería de procesos** en el análisis de procesos del sector salud.

#### MBE (Medicina Basada en la Evidencia)

Sharma y Boeckmann (2016) definen la MBE como la aplicación de la mejor evidencia clínica disponible y actual a las decisiones de atención sanitaria sobre pacientes individuales. MBE consta de **cinco pasos principales**: plantear una pregunta clínica, recuperación sistemática de la mejor evidencia disponible, valoración crítica de la evidencia, aplicación de los resultados y la evaluación del funcionamiento. Este enfoque sistemático permite al clínico evaluar y calificar la validez de un estudio y su aplicabilidad a la clínica en cuestión.

Grol y Grimshaw (1999) utilizan la MBE para analizar procesos hospitalarios. Declaran que, para mejorar la aplicación de la evidencia en la práctica, las técnicas utilizadas deben basarse en un entendimiento de los diferentes modelos y estrategias de implementación de la evidencia científica. Por lo que, la MBE debe complementarse con una implementación basada en la evidencia.

#### Lean

La filosofía principal de Lean es mejorar continuamente un proceso eliminando pasos sin valor añadido o “residuos” (en el transporte, desplazamientos, tiempos de espera, superproducción...), generales del mundo de la industria pero que se han adaptado a los procesos hospitalarios. Declaran que las organizaciones están formadas por procesos y al involucrarse con los cinco principios de Lean (Radnor y otros, 2012) de una manera escalonada y secuencial, pueden trabajar para agregar valor, reducir los residuos y mejorar continuamente el proceso.

Lean apareció por primera vez en el Sistema de Salud del Reino Unido en 2001 y en EE.UU. en 2002 (Radnor y otros, 2012). Sin embargo, la literatura manifiesta una variabilidad considerable en la implementación de Lean con diferencias en el enfoque y el alcance. En la mayoría de los casos se tiende hacia pequeños proyectos cerrados que **mejoran una parte de la asistencia del hospital**, en lugar de adoptar un enfoque global de la organización. El Royal Bolton NHS Foundation Trust se encuentra como el hospital más cercano a una aplicación completa de Lean en el Reino Unido.

### 3.3 Minería de procesos en salud

La minería de procesos (*process mining*, PM, en inglés) es una disciplina de investigación relativamente reciente. Los primeros artículos que se conocen, o al menos los más destacados, son los de van der Aalst y otros de 2003 y 2004. La minería de procesos se centra en extraer conocimiento de datos generados y almacenados en bases de datos de sistemas de información, con el objetivo de construir **registros de eventos** (*event logs*, en inglés). Los registros de eventos se pueden representar visualmente como un conjunto de trazas/caminos. Cada uno contiene todas las **actividades** ejecutadas por una instancia particular del proceso, en el caso de tratarse de un proceso hospitalario, la mayoría de veces corresponderá a un paciente.

Los **flujos de trabajo** (*workflows*, en inglés) de los procesos, son el conjunto de las trazas/caminos que se comentan. Contienen de forma ordenada las actividades por las que pasan los pacientes, ya sean estados de una enfermedad, localizaciones o pruebas. Los flujos representan la evolución espacial y/o temporal de los pacientes dentro del proceso hospitalario. Por este motivo, se define el flujo de trabajo como **flujo de pacientes** (*patient flow*, en inglés). Las herramientas de minería de procesos comienzan a emplearse en el estudio del flujo de pacientes. Perimal-Lewis y otros (2012) usaron técnicas de PM para obtener información de los datos del recorrido de los pacientes por el hospital e identificar problemas que podían causar bloqueos en el acceso o congestión en los servicios.

En lo relativo a los sistemas que contienen las bases de datos, los **sistemas de información** de procesos, con conocimiento del proceso u orientados a procesos (*Process-Aware Information Systems*, en inglés) son los sistemas capaces de producir registros de eventos. Los Sistemas de Información Hospitalarios (HIS, del inglés) son un ejemplo de ello.

La figura 1 siguiente muestra un **esquema general de la aplicación de la minería de procesos** en la asistencia sanitaria (Rojas y otros, 2016). Normalmente, cualquier actividad dirigida a la atención de un paciente, ejecutada por un médico, enfermero, técnico o cualquier otro facultativo del hospital, se almacena en el HIS. Estas actividades son las que se reflejan en los registros de eventos para el control y análisis posterior de los procesos.

**Los modelos de los procesos** se crean para especificar el orden en el que los diferentes actores deben realizar sus actividades dentro del proceso, o también para analizar críticamente el diseño del proceso. Por otra parte, los modelos de los procesos se utilizan en el desarrollo del HIS. Al tener un modelo preciso del comportamiento real del proceso, éste puede indicarle al HIS los requisitos de los procesos y especificarle alguna implementación concreta o requisitos que añadir. De esta forma el HIS podrá contribuir al análisis del proceso a partir de los datos que haya almacenado.

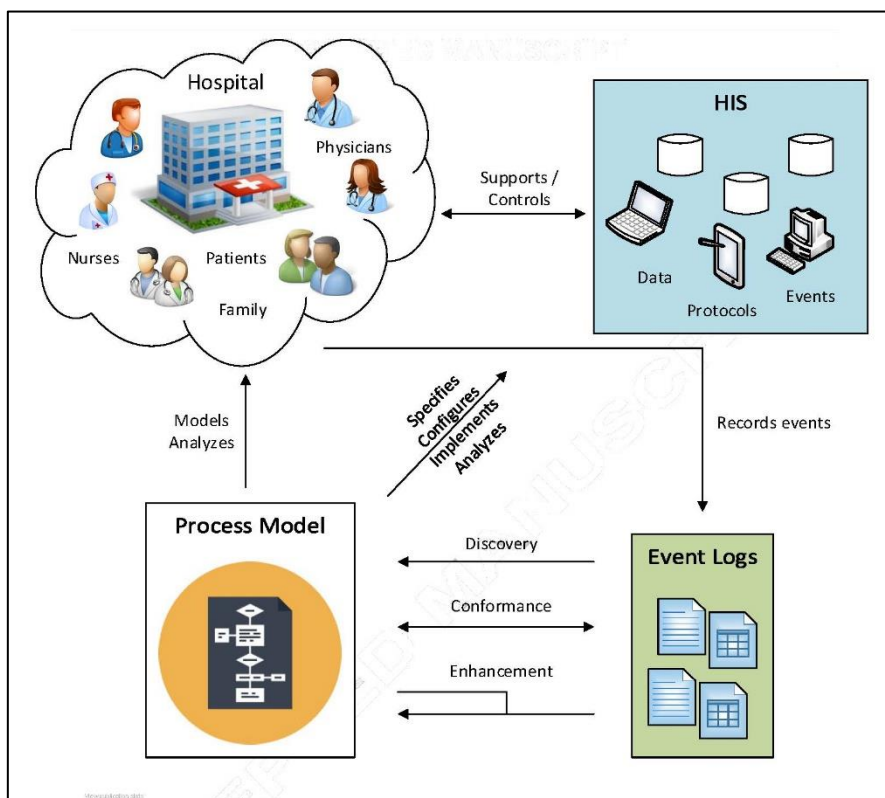


Figura 1. Esquema general de la aplicación de la minería de procesos en la asistencia sanitaria. De Rojas y otros (2016).

### Tipos de minería de procesos

En la figura 1 también se representan los tres tipos principales de minería de procesos, explicados en van der Aalst (2014):

1. **Descubrimiento de procesos** (*process discovery*). El descubrimiento automático de procesos permite extraer de los registros de eventos los modelos de los procesos. Este tipo de minería de procesos se utiliza cuando se cuenta con datos registrados de diferentes eventos del hospital, pero se desconoce realmente cómo se relacionan y cuál es el flujo que define el proceso. Este tipo de PM puede servir para crear un primer modelo del proceso.

Mans y otros (2008) aplican este tipo de técnicas de minería de procesos para obtener conocimiento significativo sobre los flujos de los procesos asistenciales. Por ejemplo, para descubrir los caminos típicos seguidos por grupos particulares de pacientes, dentro de un caso real de un proceso de oncología ginecológica en un hospital holandés.

2. **Verificación de conformidad** (*conformance checking*). Permite detectar y hacer un seguimiento de las desviaciones y anomalías del proceso, comparando un modelo dado con los registros de eventos. En este caso ya se cuenta con un modelo previo que describe el proceso, o al menos cómo debería ser. Si los expertos quieren conocer la realidad práctica del proceso y qué actividades no se desarrollan como está escrito en la teoría, utilizando este tipo de PM pueden ver dónde están los fallos y lo que provocan al resto del proceso. Esto se puede gracias a la visualización del flujo de pacientes que van avanzando por las etapas del proceso.

Zhou (2009), siendo el proceso de estudio el flujo asistencial de pacientes con artritis reumatoide, comparó los caminos de asistencia predefinidos con los flujos reales de pacientes descubiertos a partir de los datos.

3. **Mejora** (*enhancement*). Se puede mejorar o ampliar un modelo existente utilizando la información que recogen los registros de eventos sobre el proceso real. En el tipo anterior se comparaban modelo y registros, pero no se modificaba el modelo. En este caso, la información extraída de los registros de eventos se puede aplicar sobre el modelo para realizar cambios que mejoren el proceso. De esta forma, cada vez que se cambie una parte del modelo, se podrá conocer el impacto que ha tenido en la ejecución real del proceso a través de los datos registrados.

Bose y van der Aalst (2011) obtienen registros de eventos de un hospital académico holandés para ser analizados utilizando técnicas de minería de procesos. Los eventos están relacionados con el tratamiento y los pasos de diagnóstico de los pacientes con cáncer. Ponen en práctica los dos primeros tipos de PM y, tras descubrir numerosos hallazgos interesantes, manifiestan que podrían utilizarse para mejorar los procesos de asistencia subyacentes.

#### Presente y futuro de la minería de procesos

Una vez presentados los tipos, las técnicas de PM están mostrando la capacidad de descubrir los modelos de los procesos y de analizar su ejecución, lo cual proporciona grandes oportunidades de aprovechamiento de la información almacenada en los registros de eventos del HIS. No obstante, no solo se aseguran que puedan ser comprendidos con firmeza los procesos hospitalarios, sino que también pueden generar un impacto positivo en la **eficiencia de los procesos**. Como se comentaba anteriormente, para ofrecer unos servicios hospitalarios de calidad se debía hacer hincapié en este aspecto.

La minería de procesos se manifiesta en la literatura como una disciplina con potencial para mejorar la gestión de los centros médicos y la calidad de los servicios prestados, pero también se plantean **beneficios adicionales** de la aplicación de PM en la asistencia sanitaria. A continuación, se recogen los expuestos por Rojas y otros (2016):

- Ayudar a identificar y comprender el comportamiento real de los recursos y los pacientes.
- Proponer sugerencias para rediseñar el proceso.
- Reducir los tiempos de espera y de prestación de los servicios.
- Obtener información y mejorar la colaboración entre facultativos.
- Predecir el comportamiento de los pacientes basándose en casos previos.
- Agregar información adicional a las actividades, tales como datos clínicos o demográficos de los pacientes.
- Identificar cuáles son las actividades que causan cuellos de botella en el proceso.
- Ayudar a reconocer las reglas de decisión aplicadas en los diferentes casos que se dan dentro del proceso.



Muchas de estas funcionalidades de la minería de procesos ya se han probado en casos prácticos con datos reales, pero otras todavía están por estudiar en el futuro.

### Herramientas de minería de procesos

Existe una amplia gama de algoritmos y técnicas de minería de procesos disponibles, y herramientas comerciales y no comerciales con las cuales implementarlas y generar modelos, tablas e información para analizar. La **selección de una herramienta** fácil de usar y accesible para los profesionales, que pueda integrarse en su proceso diario, es clave para su éxito. La tecnología utilizada no debe ser una barrera para el funcionamiento adecuado del sistema. De hecho, mejorar la comprensibilidad y la usabilidad de las técnicas de minería de procesos es un reto actual todavía en desarrollo (van der Aalst, 2012).

Después de la revisión bibliográfica publicada por Rojas y otros (2016), concluyeron que las dos herramientas más utilizadas en salud eran **ProM** (van Dongen y otros, 2005) y **Disco** (Mans y van der Aals, 2015). No obstante, en este proyecto se empleará una herramienta llamada **PALIA** (Fernandez-Llatas y otros, 2015), que se presenta más adelante.

### ProM

En salud, la herramienta más comúnmente utilizada es ProM, que consiste en un software de código abierto de descarga gratuita y de uso mayoritariamente académico. Es una herramienta grande y poderosa por las funcionalidades que ofrece. Ha implementado un gran número de técnicas y algoritmos (van Dongen y otros, 2005) porque no solo contiene implementaciones de los algoritmos de minería de procesos más clásicos, sino que también permite la creación de complementos para agregar nuevos algoritmos a la herramienta (Figura 2). Se ha utilizado a lo largo de diferentes estudios de casos (en el 42% de los 74 casos estudiados por Rojas y otros (2016)).

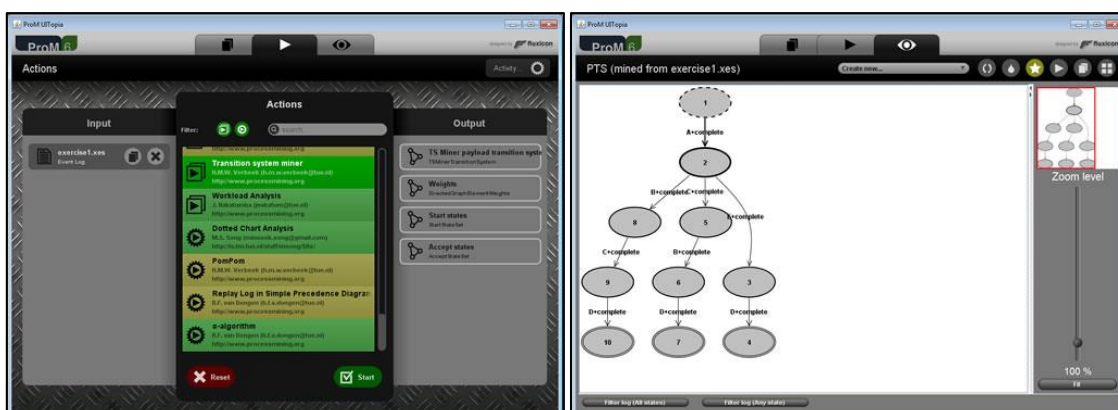


Figura 2. Dos capturas de pantalla de la interfaz de la herramienta ProM. Extraída de: <http://www.promtools.org/doku.php>

### Disco

Disco, por otra parte, es una aplicación comercial con una interfaz para los modelos de los procesos mucho más visual y amigable que ProM. Además, ofrece la posibilidad de aplicar de forma sencilla múltiples y variables opciones de filtrado en los registros de eventos (Figura 3). Disco nació después de ProM cuando se dieron cuenta de que los analistas de procesos en la práctica lo que necesitaban era una herramienta que sobre todo hiciera que la minería de procesos fuera fácil y rápida (Gunther y Rozinat, 2012). Ese es el objetivo principal de Disco. Sin embargo, todavía su aplicación en el sector salud está por explotar. Ha sido utilizado en un 10% de los casos estudiados por Rojas y otros (2016).

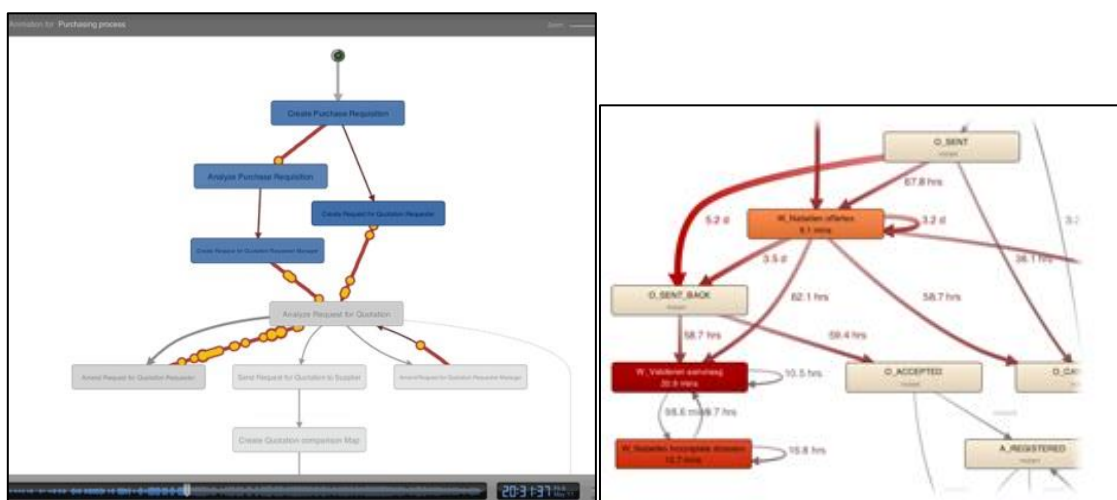


Figura 3. Mapa animado de un proceso visto con Disco (izq.) y el flujo de otro proceso en Disco (dcha.).  
Extraída de: <http://fluxicon.com/blog>.

### Campos médicos de aplicación de la minería de procesos

Dado que el estudio principal de este proyecto se centra en el análisis de las Consultas Externas del HUPLAFE, es interesante conocer **otros campos donde ya se haya aplicado la minería de procesos** con anterioridad. La figura 4 muestra las 22 áreas diferentes que se han registrado y el número de estudios en cada una de ellas. Los campos con mayor número de casos estudiados con PM son Oncología y Cirugía (Rojas y otros, 2016). La figura 4 también deja ver dos aspectos:

- La **multidisciplinariedad** de la minería de procesos en salud y el potencial que tiene para acabar aplicándose en todas las áreas médicas.
- Las **Consultas Externas** se encontrarían dentro de la Atención Ambulatoria (*Outpatient Care*). Si bien ya hay estudios en este campo, el uso de técnicas de PM para mejorar la gestión de las CEX del HUPLAFE se presenta como un reto ambicioso y real que puede aportar información relevante a las investigaciones previas y las venideras, dado el tamaño del hospital y la cantidad de servicios y personal involucrado en estas tareas.

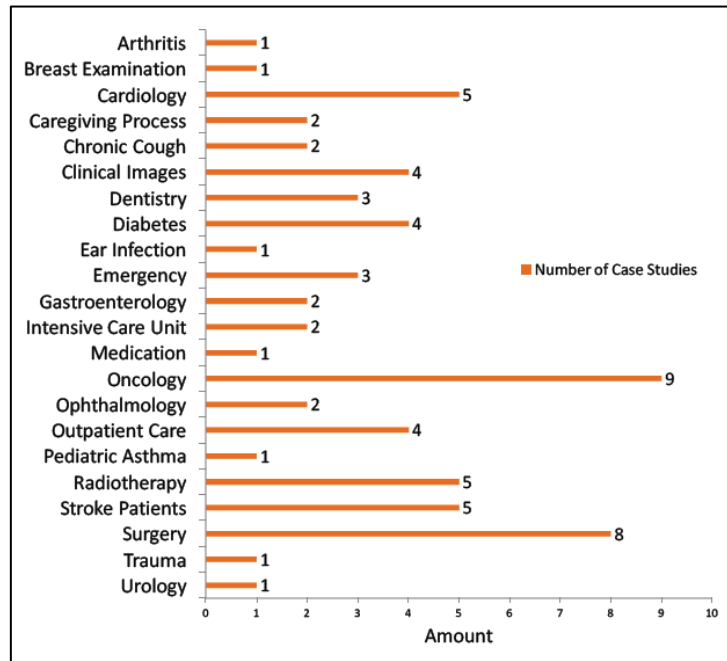


Figura 4. Los estudios de casos reales agrupados por su campo médico. De Rojas y otros (2016).

## 4. MATERIALES Y MÉTODOS

En este capítulo se muestra la metodología empleada y se describe el entorno de trabajo donde se aplicará la herramienta de minería de procesos. Se presenta el funcionamiento de la misma (PALIA) y cuáles son las posibilidades que ofrece para el análisis llevado a cabo en el proyecto.

### 4.1 Entorno de trabajo

El proceso objeto de estudio son los flujos de pacientes de las Consultas Externas del HUPLAFE a lo largo de un día entero. Se firmó un convenio entre la UPV y el Instituto de Investigación Sanitaria La Fe, por el cual nos daban acceso al corpus de **datos del día 14 de noviembre del 2016**. Como resultado de los análisis llevados a cabo por los gestores del HUPLAFE, se conoce que, tanto el mes de noviembre como el de febrero son de los meses más regulares, en el sentido de que son meses con pocas festividades y están sumidos en la rutina de trabajo, lo que hace que sean representativos del funcionamiento normal y cotidiano del hospital.

A continuación, se presenta cómo se organiza espacialmente el hospital y cuál es el protocolo que se ha de seguir en las consultas externas, es decir, cómo debería transcurrir idealmente el proceso en la práctica.

#### Estructura del HUPLAFE y niveles de localización

El HUPLAFE está formado por 9 edificios (Figura 5):

- Torre A: Instituto de Investigación Sanitaria La Fe y los laboratorios.
- Torre B y C: Consultas externas.
- Torres D, E, F y G: Hospitalización.
- Edificio H: Dirección y Administración. En este edificio se encuentra el Departamento de Calidad, y es donde trabajan los encargados de gestionar y evaluar las consultas externas.
- Edificio de instalaciones técnicas.

Las **dos torres** dedicadas expresamente a consultas externas tienen siete plantas cada una. En ellas, cada una de estas plantas se divide en dos espacios: consultas y técnicas. **Consultas** es donde se lleva a cabo la anamnesis entre el médico y el paciente, y el primero rellena un informe con los datos recogidos durante la misma. En **Técnicas** se realizan diversas pruebas a los pacientes y se registra la información de la misma forma. Éstas dependen del servicio en el que se encuentren, pero, por ejemplo, en neumología están las pruebas funcionales respiratorias y en cardiología les registran el electrocardiograma (ECG).



Figura 5. Estructura del HUPLAFE. Extraído de <http://ufpelafe.webcindario.com>

El área de consultas está formada por una sala de espera y 15 consultas propiamente dichas, donde el médico atiende al paciente. Estos espacios se llaman **locales**, término que sirve tanto para las consultas, como para las técnicas. El área de técnicas también tiene una sala de espera, pero tiene menos locales, 11 esta vez. De esta forma ya podemos adelantar cómo será la numeración de los locales. Por ejemplo, B412 corresponderá a un local de la torre B, planta 4 y consulta 12. Si este último número está entre 1-15, corresponde a un local de la parte de consultas, por el contrario, si está dentro del intervalo 16-26, será un local de la parte de técnicas.

Se muestran en la tabla 1 los **niveles de localización** de los pacientes, es decir, los diferentes niveles que tenemos para situar a los pacientes dentro del hospital.

TORRE							
B	C	D	E	F	G		
PLANTA							
0	1	2	3	4	5	6	7
TIPO 2							
Consultas				Técnicas			

Tabla 1. Niveles de localización de los pacientes en Consultas Externas.

### Protocolo de las Consultas Externas

El paciente está citado para acudir el 14 de noviembre al HUPLAFE. Por otro lado, el médico de la consulta tiene en su ordenador una tabla con una fila para cada paciente y cuatro casillas asociadas al mismo, dentro del **programa de llamada de pacientes**. Cada casilla corresponde al registro de una etapa del proceso de CEX:

1. **Hora de llegada.** Le aparece al médico la hora de llegada del paciente en la primera casilla.
2. **Hora de llamada.** El médico utiliza esta casilla para llamar al paciente a que entre al local, y se registra la hora.
3. **Hora de entrada.** Cuando el paciente entra por la puerta del local, el médico marca en la tercera casilla la hora. Si no es así, se pone automáticamente la hora de llamada como hora de entrada.
4. **Hora de salida.** Se puede registrar el momento de salida del paciente de tres formas:
  - El médico indica en el sistema que la prestación ha finalizado y se guarda la hora como salida.
  - El médico marca la opción “Envío a Citación” cuando quiere que el paciente vaya a Citación para programar la siguiente visita. Introduce las pautas de citación y se guarda la hora como salida.
  - Si el médico no marca que la prestación ha finalizado ni envía al paciente a Citación, se registrará como hora de salida el momento en el que firme el informe de la consulta.

Debemos tener en cuenta que, en los registros de citas del día disponible, no solo habrá pacientes que estén citados para ir ese día, también habrá citas de **personas sin cita previa** que necesitaban ser atendidas. Estos pacientes pueden venir de urgencias o acudir directamente a CEX.

Mediante el siguiente esquema se representa el **camino que recorre el paciente** desde que llega al hospital hasta que sale del local donde ha sido atendido:

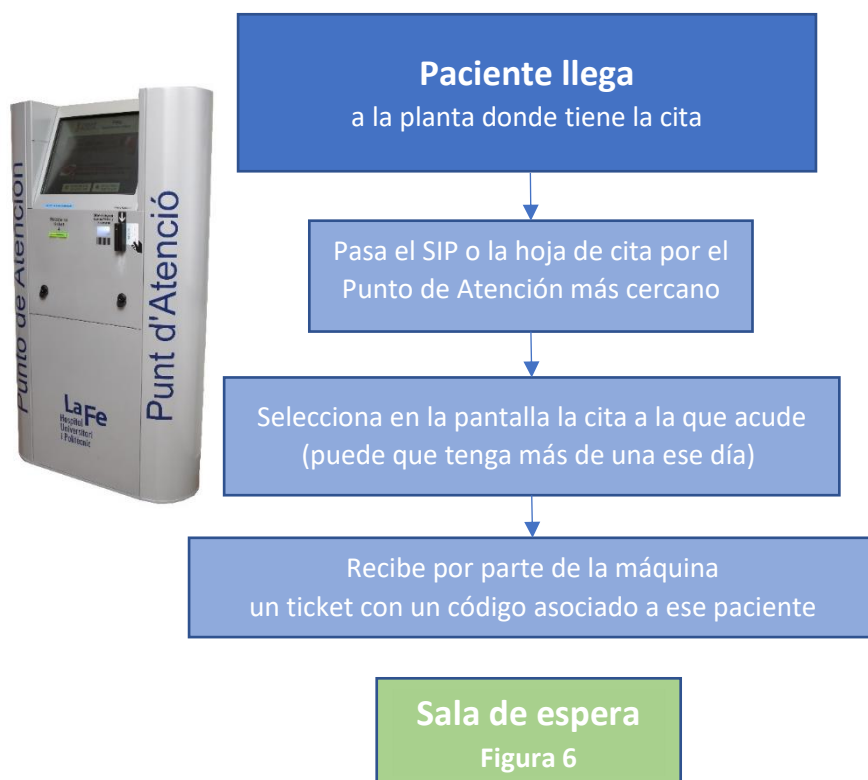






Figura 6. Sala de espera en Consultas Externas. Extraída de: <https://paufferro.es/proyecto/hospital-la-fede-valencia/>



El facultativo del local puede mandar al paciente a **citación** o **finalizar** la visita sin más. Si se trata de una prestación a la que tiene que asistir periódicamente, el paciente acudirá al servicio de citación para programar su siguiente visita, pero si el médico le quiere dar de alta marcará la prestación como finalizada y el paciente se marchará sin más. Por otra parte, después de pasar por citación o salir del local, el paciente puede **ir a otra cita donde se repetirá todo el proceso** de nuevo.

## 4.2 Aplicación PALIA

Para poder cumplir el objetivo marcado en capítulos anteriores, necesitamos una tecnología con la que inferir los procesos reales a partir de registros de ejecuciones reales, y que sirva de apoyo a los gestores del HUPLAFE en el diseño y control de calidad de los procesos de atención sanitaria. Entre las herramientas de PM nombradas en el apartado anterior, escogemos PALIA, un **algoritmo de inferencia de registros basados en actividades paralelas** (del inglés, Parallel Activity-based Log Inference Algorithm). Esta tecnología ha sido creada por el grupo de investigación SABIEN, dentro del ITACA (Instituto Universitario de Tecnologías de la Información y Comunicaciones), donde se ha desarrollado este proyecto.

PALIA recopila un conjunto de técnicas de reconocimiento de patrones sintácticos para construir un autómata formal, llamado autómata paralelo temporizado (del inglés Timed Parallel Automaton, TPA) (Fernandez-Llatas y otros, 2011), que es equivalente a una red segura de Petri (representación gráfica de un sistema como eventos discretos en el cual se puede describir la topología del sistema) para representar el proceso. PALIA puede manejar acciones paralelas porque realiza un análisis del tiempo en la creación de las dependencias entre nodos (Fernandez-Llatas y otros, 2015). Además, aunque puede tratar con éxito los datos basados en **eventos**, está diseñada específicamente para hacer frente a las **actividades** (Fernandez-Llatas y García-Gómez, 2015). Una actividad proporciona, además del evento inicial y final del registro del flujo de trabajo (workflow log), el resultado de la actividad. Esto es importante en los procesos del sector salud, ya que proporciona una visión sintetizada de la decisión tomada en los protocolos hospitalarios. Permite a PALIA definir una dependencia entre las acciones en función de sus resultados, que no está disponible en otros algoritmos.

De esta forma, queda remarcada PALIA como una técnica de PM con potencial para analizar procesos hospitalarios. Por este motivo, y por los que se muestran a continuación, hemos seleccionado este algoritmo. En experimentos pasados en los que se ha aplicado, se obtuvieron buenos resultados y su comportamiento fue generalmente mejor en comparación con otros algoritmos (Fernandez-Llatas y otros, 2011; Fernandez-Llatas y García-Gómez, 2015). Además, el uso de PALIA permite ser más flexibles en el caso del **enriquecimiento de los datos** de entrada en el futuro, mezclando el corpus actual con datos hospitalarios nuevos. Por último, al llevarse a cabo el proyecto junto a los desarrolladores de PALIA, se presenta la posibilidad de modificar y **adaptar la herramienta** a las necesidades que aparezcan durante el mismo.

Para el desarrollo del proyecto se hará uso de la herramienta PALIA en su **versión web**. Para conocer su potencial y las funcionalidades que ofrece, a continuación, se describe en profundidad:



## PALIA Web

### *Cargar el corpus de datos*

El corpus de datos que se carga en PALIA tiene que ser un archivo “.csv” que contenga en la primera fila la cabecera. El resto de filas del corpus son entradas llamadas **eventos**. Cada **paciente** puede tener uno o varios eventos, en el caso de estudio, un paciente puede tener una o varias citas registradas. El camino que recorre el paciente en el hospital se representa en PALIA como una **traza**. La herramienta carga todos los ID’s del corpus y guarda cada uno como una **muestra**. De esta forma, el número de muestras que detecta es igual al número de pacientes que hay en ese corpus.

Cuando se carga el archivo del corpus de datos, también hay que indicarle a PALIA qué tipo de elementos hay en la **cabecera** del archivo. Los tipos de elementos posibles son los siguientes:

- ID: identificador de la muestra. Si dos entradas tienen el mismo ID (mismo paciente) enlaza sus actividades en el tiempo quedando solo una traza/camino para ese ID.
- ACTIVITY: Referido a la actividad asociada al evento. La actividad indica el estado del paciente en el evento. En el caso de que se esté analizando el desplazamiento del paciente, su localización, cada actividad especificará un lugar. Por ejemplo, la torre y la planta pueden ser dos actividades porque representan dos localizaciones y han de estar una comprendida dentro de la otra para que ese evento indique que está en la planta X de la torre Y.
- START: Indica el día, mes, año, hora, minutos y segundos de inicio del evento.
- END: Indica el día, mes, año, hora, minutos y segundos de finalización del evento.
- SAMPLE: Es una característica que se atribuye a un ID, en nuestro caso a un paciente. Si un paciente tiene varias citas/eventos, el campo tipo *Sample* ha de ser igual en todas ellas.
- EXCLUDE: Se coloca cuando se quiere que esa columna no sea cargada en PALIA.

Un ejemplo se muestra en la tabla 2.

ID	ACTIVITY	ACTIVITY	START	END	SAMPLE	EXCLUDE
ID paciente	Torre	Planta	Inicio prestación	Fin prestación	Edad	Fecha Cita

Tabla 2. Elementos de las cabeceras. La primera fila corresponde a la cabecera que necesita PALIA para cargar los datos y la segunda es la cabecera del corpus.

### *Estructura gráfica de la herramienta*

La interfaz gráfica de la herramienta se divide en **cinco zonas** (Figura 7):

1. Menú superior: Permite el acceso a funciones comunes de la aplicación como cargar y guardar el corpus de trabajo o iniciar la minería de datos.

2. Zona de filtros: Permite seleccionar y filtrar qué datos del corpus inicial se utilizarán en la minería.
3. Zona de minería de datos: Una vez realizada la minería permite trabajar con la inferencia obtenida aplicando diversos tipos de visualización o filtrados.
4. Zona de representación gráfica de los flujos de trabajo (*workflow*): Visualiza la inferencia o inferencias obtenidas como resultado de aplicar el algoritmo de minería de datos.
5. Zona de información: Muestra información extra relacionada con el estado de las tareas aplicadas en las diferentes secciones.

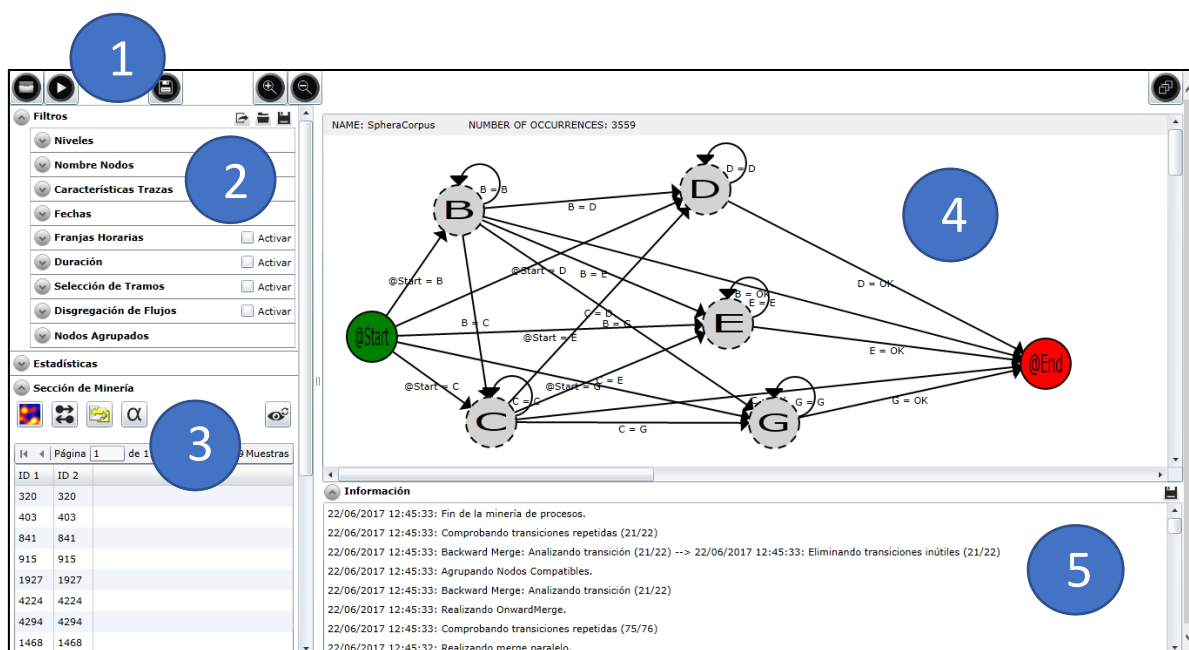


Figura 7. Zonas de la interfaz gráfica de PALIA.

### Representación gráfica de los flujos de trabajo

Las actividades de una muestra/paciente, en nuestro caso ubicaciones o localizaciones, se representan como un **nodo** (Figura 8). Los nodos tienen forma circular y contienen el nombre en su interior. Hay dos nodos especiales añadidos a todas las muestras, el nodo @Start (en verde) indica el inicio de cualquier muestra/camino y el nodo @End (en rojo) recogerá todos los finales de muestras/caminos. Los nodos normales aparecen en color gris y los que son el final de una muestra/camino están dibujados con el borde a trazos discontinuos. Los nodos del flujo de trabajo pueden cambiarse de ubicación para hacer el proceso más visual.

El resultado de una actividad indicará a que nodo destino pasa, es decir, el resultado genera una **transición** (Figura 8). Las transiciones se representan con una flecha desde el nodo origen al destino y con un nombre en el centro con el siguiente formato "Nombre nodo origen = Resultado Actividad", en nuestro caso eso equivale a "Nombre nodo origen = Nombre nodo destino". Se puede seleccionar una transición del flujo de trabajo para que aparezcan los ID's de los pacientes que contienen esa transición.

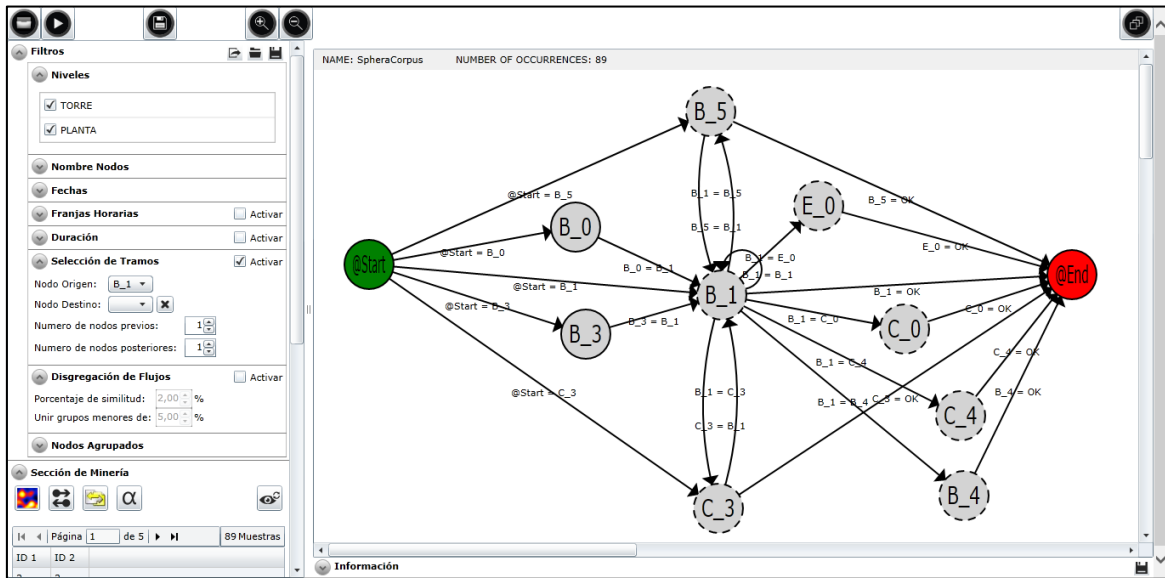


Figura 8. Ejemplo de inferencia en PALIA.

En los casos en los que como resultado del filtrado hay varios corpus, y por tanto se infieren **varios flujos de trabajo**, se visualizan en forma de tabla apilada (Figura 9). Se añade sobre cada uno para identificarlo una cabecera con el nombre, número de muestras que lo componen (ocurrencias) y el porcentaje que éstas suponen sobre el total de muestras del corpus original.

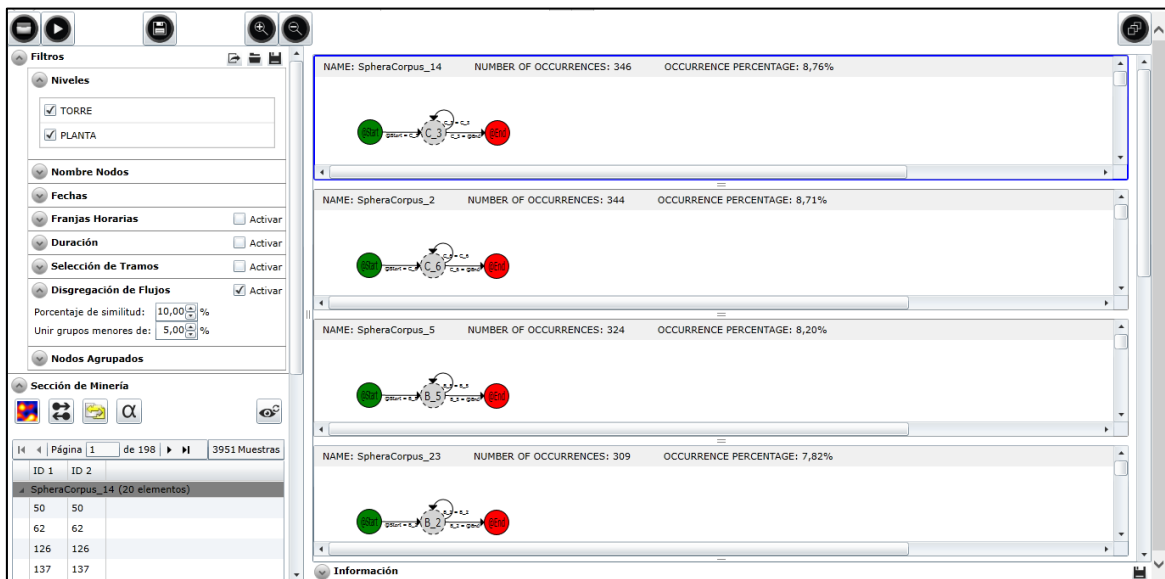


Figura 9. Visualización en PALIA de varios flujos de trabajo

### Zona de filtros

Los filtros son los encargados de adecuar los datos del corpus de entrada al formato del módulo de minería. En esta zona (Figura 10) podemos encontrar los diferentes filtros que permiten modificar y seleccionar qué datos del corpus inicial queremos que lleguen al motor de minería. Según el corpus elegido, se cargarán los diferentes filtros aplicables al mismo.



Figura 10. Zona de filtros

### **Filtro: Niveles**

Hay tantos niveles como campos tipo *Activity*. En el desplegable se puede seleccionar todos o solo alguno de ellos. De esta forma, el nombre de los nodos será uno u otro. Por ejemplo, en el caso de tener dos campos tipo *Activity*, llamados Torre y Planta, si solo seleccionamos el primero, el nombre de los nodos/eventos será el de la torre correspondiente y si seleccionamos ambos, el nodo se llamará como la Torre\_Planta asociada. Este filtro nos permitirá definir con qué granularidad de área dentro de las categorías existentes queremos trabajar.

### **Filtro: Nombre nodos**

En el desplegable aparece una lista con todos los nombres de los nodos que se han cargado y cada uno de ellos se puede deseleccionar para visualizar solo los otros. Este filtro es útil cuando queremos ver las relaciones existentes entre unas localizaciones en concreto. En el ejemplo expuesto anteriormente con los nodos tipo Torre\_Planta, este filtro nos puede servir para visualizar únicamente las relaciones que se establecen entre las plantas de una torre, si dejamos seleccionados los nombres de nodos que comiencen por la letra de la torre de interés. También permite renombrar el área de localización (nodo), pudiendo asignar un nuevo nombre a un área o incluso agrupar dos áreas dándoles el mismo nombre.

### **Filtro: Características Trazas**

Este filtro solo aparece si en el corpus de datos hay algún campo tipo *Sample*. Es un filtro en el que se pueden seleccionar las características que interesen de los pacientes. Si este campo fuera edad, como en el ejemplo, podemos quedarnos únicamente con los flujos de los pacientes mayores de 65 años.

### **Filtro: Selección de Tramos**

El filtro permite quedarse solo con los pacientes que pasan por un nodo origen o que pasan por un nodo origen y a continuación por otro nodo destino. A demás es posible indicar cuántos nodos antes del origen y del destino se quieren. El resto de datos de las muestras/caminos se descarta.

En el ejemplo de la figura 11, dejando esa configuración del filtro, nos quedaremos con los caminos de los pacientes que pasan por el nodo B\_1.



Figura 11. Filtro selección de tramos

### **Filtro: Disgregación de Flujos**

Con este filtro se divide/disgrega el corpus de datos en varios corpus agrupando las muestras por su porcentaje de similitud (Figura 12). Todos aquellos nuevos corpus cuyo número de muestras no supere el porcentaje “unir grupos menores de” sobre el número total de muestras del corpus inicial, se unificarán en un mismo corpus. Como resultado de aplicar este filtro, normalmente, el corpus inicial se convertirá en varios corpus y por tanto en varias inferencias calculadas.

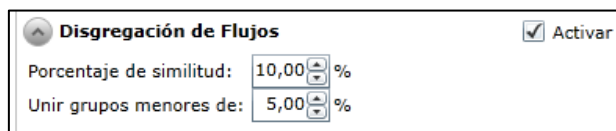
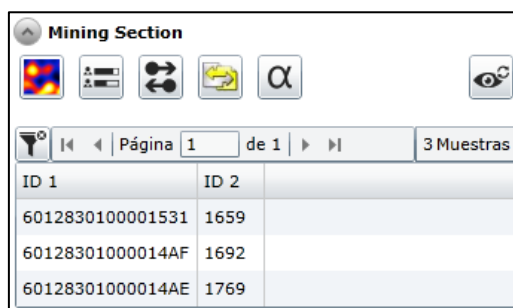


Figura 12. Filtro disgregación de flujos

### *Zona de minería de datos*

Tras haber realizado la minería de datos, desde esta zona (Figura 13) se puede interactuar con la inferencia obtenida. En la parte superior hay un conjunto de filtros para cambiar como se visualizan los datos u obtener vistas con datos extra. En la parte inferior aparece el número y el listado de muestras del corpus que han pasado los filtros. Se puede seleccionar una muestra/paciente para resaltar su camino sobre el resto.



ID 1	ID 2
6012830100001531	1659
60128301000014AF	1692
60128301000014AE	1769

Figura 13. Zona de minería de datos

### **Filtros de minería**

Los distintos botones de la parte superior permiten alterar como se visualiza el flujo de trabajo inferido o mostrar datos extra sobre la visualización.

De izquierda a derecha se encuentran:

1. Mapa de frecuencia: Cambia la visualización del flujo de trabajo inferido y lo muestra en forma de mapa de calor, desde el rojo, pasando por el naranja y llegando al verde, para indicar en el caso de las transiciones con qué frecuencia se dan y en el caso de los nodos durante cuánto tiempo ocurren. En verde aparecen los elementos con menor ocurrencia/duración y en rojo los de mayor ocurrencia/duración.
2. Mapa de ocupación: Muestra el flujo de trabajo en forma de mapa de calor indicando qué nodos/localizaciones están más ocupadas en un momento del tiempo, respecto a un valor de ocupación máxima.
3. Buscador de saltos: Resalta en el flujo de trabajo la ida y vuelta desde un nodo al mismo nodo origen (saltos), de un tiempo acotado.
4. Comparar inferencia con un patrón guardado: Permite comparar el flujo de trabajo de la inferencia visualizada con otra guardada previamente.
5. Valor de transparencia alfa: Permite cambiar el valor de transparencia de los objetos que se vuelven transparentes en algunas de las visualizaciones previas.
6. Restaurar: Restaura la visualización a su estado original.

## 4.4 Fases del trabajo

A la hora de cumplir el objetivo principal de este proyecto, los objetivos secundarios sirven de pauta para alcanzarlo, no obstante, los temas a abordar son muchos y se manifiesta la necesidad de seguir una **metodología** que clarifique las etapas del trabajo y su ejecución temporal.

### Metodología CRISP-DM

En la asignatura Sistemas de Información y Telemedicina II (SIT II) estudiamos un modelo de referencia para el desarrollo de proyectos de minería de datos llamado **CRISP-DM** (Sistemas de ayuda a la decisión basados en datos biomédicos para clínica asistencial y salud pública, 2017). Detecté que podía ser un **método aplicable a este proyecto** y que podría orientar el planteamiento del trabajo. De esta forma, pese a estar enfocado a la minería de datos, se usará para mejorar la gestión de las Consultas Externas del HUPLAFE haciendo uso de técnicas de minería de procesos.

Sin embargo, esta no es la primera vez que se utiliza esta metodología en minería de procesos, Zhou (2009) estudió un caso de un proceso hospitalario real con el sistema CRISP-DM y lo evaluó. Concluyó que esta metodología era muy **útil en los proyectos de minería de procesos** a pesar de las diferencias encontradas en la fase de modelado, que atribuyó a problemas concretos de su estudio.

### *Fases principales de la metodología CRISP-DM*

La figura 14 ilustra las seis fases principales de CRISP-DM, adaptadas a la minería de procesos. Las flechas indican las dependencias más importantes y frecuentes entre fases. El círculo exterior muestra la naturaleza cíclica de la minería de procesos en sí. Las lecciones aprendidas durante el método pueden originar nuevas preguntas que hagan volver a la etapa inicial para generar modelos más robustos. Las **seis fases principales** de CRISP-DM son las que siguen:

1. Comprensión del negocio. Análisis de la organización sanitaria y el problema a resolver.
  - a. Comprender los objetivos del proyecto
  - b. Analizar los requisitos clínicos del proceso
  - c. Diseñar un plan para lograr los objetivos propuestos
2. Análisis de los datos
  - a. Recolección de datos
  - b. Familiarizarse con los datos
  - c. Identificar problemas de calidad de los datos
  - d. Describir los datos
  - e. Detectar subconjuntos de variables u observaciones interesantes
  - f. Establecer hipótesis iniciales
3. Preparación de los datos
  - a. Selección de datos: criterios de inclusión y exclusión
  - b. Selección/Transformación de variables
  - c. Limpiar datos
  - d. Construir el conjunto de datos que introducir en la herramienta de PM
4. Modelado
  - a. Seleccionar y aplicar una herramienta de PM
  - b. Escoger los filtros y las vistas que respondan a las preguntas planteadas por los expertos
  - c. Obtener modelos del proceso a partir del flujo de pacientes
5. Evaluación clínica
  - a. Revisar los pasos llevados a cabo para obtener el modelo del proceso
  - b. Evaluar si se alcanzan los objetivos planteados
  - c. Probar que las vistas obtenidas estén en concordancia con la realidad
  - d. Estudiar las razones que justifiquen determinados comportamientos vistos en los flujos de pacientes
  - e. Evaluar el nivel de aceptación de la herramienta que tienen los expertos
  - f. Determinar la utilidad de los modelos obtenidos
6. Implantación de las soluciones
  - a. Sugerir cambios en el modelo del proceso en función de anomalías y desviaciones detectadas con la herramienta
  - b. Plantear adquisición diferente de los datos en función de la fiabilidad y el formato de los actuales
  - c. Proponer el registro de nuevos datos que añadan actividades al proceso para tener más información del mismo
  - d. Considerar la utilización de la herramienta de PM en el análisis diario que realicen los expertos

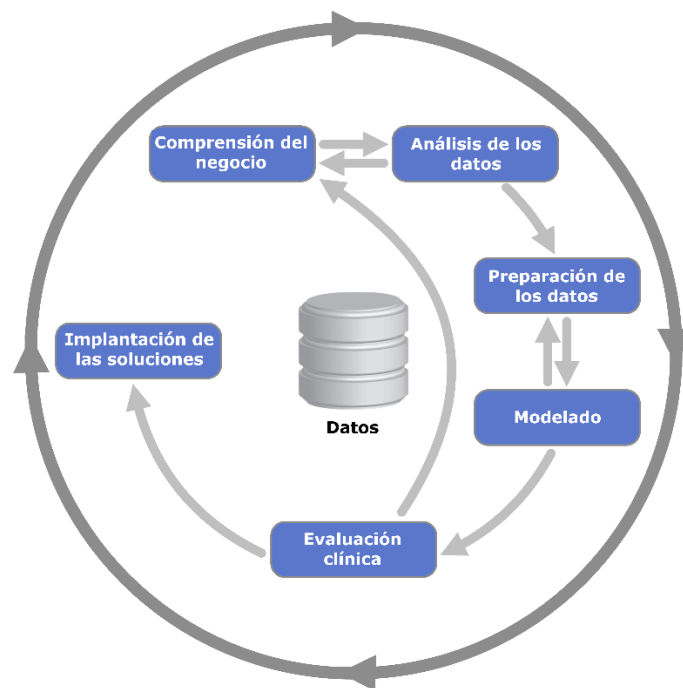


Figura 14. Diagrama de las seis fases principales de la metodología CRISP-DM. De los apuntes de la asignatura SIT II.

### Fases del trabajo

Una vez presentada la teoría de la metodología que se ha seguido en el desarrollo de este proyecto, se expone a continuación cómo se ha llevado a la práctica y cuáles han sido las etapas del trabajo.

#### *Comprensión del negocio. Análisis de la organización sanitaria y el problema a resolver.*

En primer lugar, se ha centrado el trabajo en **conocer y entender el funcionamiento de las Consultas Externas** en el HUPLAFE. Recorriendo el circuito que siguen los pacientes para visualizar la distribución de los espacios, siendo conscientes de la cantidad de personas que se mueven en el hospital, y las colas que se forman en algunas ocasiones, cómo utiliza el médico el programa de llamada de pacientes y cómo es la adquisición de los datos a analizar, entre otras cosas. En definitiva, todo aquello plasmado en el apartado *Entorno de trabajo*, que se considera crucial para alcanzar la máxima explotación de los datos.

Por otra parte, pasando a la gestión del hospital, es importante conocer el trabajo de sus **gestores en el Departamento de Calidad**, saber cuáles son sus necesidades a la hora de gestionar las Consultas Externas y cuáles son las preguntas que pueden hacer aflorar conductas en los datos que desconocían. Para poder conseguirlo, hace falta por el otro lado descubrir en profundidad la herramienta PALIA y sus funcionalidades, para luego poder aplicarlas una vez cargados los datos reales.



### *Análisis de los datos*

Para poder acceder a los datos reales y trabajar con ellos, justificar mi presencia en el hospital es el primer paso. Una vez identificada y teniendo todas las autorizaciones, se podrá comenzar la familiarización con los datos: **caracterizar el corpus**, conocer los campos y su formato, los valores posibles dentro de cada uno, qué parte del proceso representan, reconocer el número de pacientes y de citas, etc. Todo lo necesario para aprovechar lo que los datos puedan ofrecer.

De forma consecutiva se pasa a identificar los problemas de **calidad de los datos**, se detectan datos erróneos y cuáles son más fiables para obtener modelos visuales y lo más cercanos a la realidad. **Detectar subconjuntos** de variables u observaciones interesantes, como tipos de pacientes, puede enfocar el análisis a grupos concretos en función de las preguntas que se quieran responder con los modelos. Descubiertos estos grupos se pueden establecer razones que justifiquen su existencia y su comportamiento en la ejecución del proceso.

Para el desarrollo de esta fase se han utilizado **dos programas**:

- **Excel**: Para visualizar y modificar el archivo “.xlsx” original con los datos registrados de Consultas Externas en el día 14 de noviembre de 2016. Con el propósito de conocer el rango de valores de cada campo y los recorridos de los pacientes por el proceso (según qué datos tengan registrados), se han usado principalmente los filtros de este programa. También se han hecho los cálculos para detectar datos erróneos, se ha cambiado el nombre de algunos campos y se han empleado diversas funciones propias de Excel.
- **Matlab**: Se han creado varias funciones y scripts en Matlab para calcular el número de consultas por paciente, eliminar citas duplicadas, corregir citas solapadas y depurar la hora de salida, entre otras cosas.

### *Preparación de los datos*

Para obtener unos modelos u otros se han seleccionado los datos para formar un corpus específico cada vez. Para ello, se han seguido, a la hora de obtener todas las vistas, unos criterios de inclusión y exclusión de los datos. Sin duda para que la herramienta pueda entender y analizar estos datos se precisa la transformación de los mismos para cambiar formatos o para obtener combinaciones de ellos. En pocas palabras, se trata de **construir los registros de eventos** a partir del corpus analizado en función de la pregunta que se quiera responder.

Para el desarrollo de esta fase se han utilizado los **dos programas** anteriores:

- **Excel**: Para visualizar y modificar el archivo .xlsx original de los datos, ahora ya, analizados. Se ha utilizado para cambiar el formato de las fechas y para unir fecha y hora en una misma columna, entre otras cosas. Se han eliminado algunos eventos/filas por tener errores o por no interesar para un corpus concreto. Siempre haciendo uso de los filtros para controlar qué datos se estaban modificando. Por último, cada vez que los datos estaban listos para introducirse en PALIA, se pasaba el archivo “.xlsx” a “.csv” para cargarlo en la herramienta de minería de procesos.
- **Matlab**: Se han creado varias funciones y scripts para llevar a cabo algunas de las transformaciones de los datos y corregir los errores detectados, entre otras cosas.

### *Modelado*

Una vez cargado el corpus de datos en la herramienta PALIA, es clave elegir los **filtros** que originen las **vistas** que aporten mayor información y que puedan responder a las preguntas planteadas por los expertos. Si PALIA alerta de algún error en los datos, como el solapamiento en el tiempo de eventos de un paciente, hay que volver atrás y preparar los datos de nuevo o asumir que esas muestras no se tendrán en cuenta en la inferencia, porque PALIA las elimina.

Una vez obtenida la vista deseada, se pasa a **analizar el flujo de pacientes** con las funcionalidades que ofrece PALIA o volviendo al corpus de datos para tener más detalle de muestras concretas. Se puede estudiar si se cumple el modelo del proceso diseñado por los gestores, cuántos lo cumplen y cuántos no, en qué servicios ocurre, qué relación existe entre las actividades del proceso, cómo es la interacción entre servicios que comparten pacientes, todos los aspectos del proceso que surjan a partir de la visualización del flujo de pacientes y que con el análisis en Excel no se podían percibir.

### *Evaluación clínica*

Después de la fase de modelado y habiendo extraído de la inferencia la máxima información posible, se pasa a evaluar si las **vistas** obtenidas han podido **responder a las preguntas planteadas** por los expertos. Además, se puede comprobar que los flujos de pacientes tienen sentido dentro del proceso y hacer estudios más exhaustivos de la causa de los caminos anormales. Por otra parte, analizar la utilidad de la herramienta, y su potencial, y ver si es del agrado de los expertos tanto a nivel visual, como práctico, y si podría satisfacer sus necesidades a la hora de gestionar las CEX.

### *Implantación de las soluciones*

Una vez llegados a esta fase, habiendo estudiado los datos y su calidad, y las vistas que son capaces de generar en la herramienta de PM, se pueden hacer **sugerencias** sobre varios aspectos: los modelos del proceso, los tipos y cantidad de datos y la adquisición de los mismos. Todas recomendaciones para aumentar la calidad de los datos y así poder obtener visualizaciones cada vez más útiles que permitan mejorar los procesos. Por último, se valorará la capacidad de la herramienta para participar en el análisis diario de las Consultas Externas y su incorporación en el futuro en el Departamento de Calidad del HUPLAFE para analizar ese y otros procesos hospitalarios.

Para cada fase del proyecto se ha contado con la colaboración de los gestores del HUPLAFE o los investigadores del SABIEN, siendo en algunos momentos necesaria la interacción entre ambos para clarificar objetivos y evaluar el progreso del proyecto. Esta **dualidad** y la necesidad de entender ambas partes, se manifiesta en algunas etapas del trabajo. En la fase de modelado, por ejemplo, era importante conocer las preguntas y las necesidades de los funcionarios del HUPLAFE y la parte más técnica del manejo de la herramienta que provenía del SABIEN.

## 5. RESULTADOS Y DISCUSIÓN

En este apartado se exponen los resultados obtenidos tras aplicar la metodología anterior. Para plasmar el orden cronológico en el que se obtuvieron, se mostrarán los resultados referenciando a qué etapa del método corresponden. Por otra parte, también se hará la discusión de los mismos.

### 5.1 Resultado del Análisis de los Datos

En la fase de análisis de los datos primero se caracteriza el corpus, es decir, se trata de familiarizarse con los datos y conocer el significado de cada campo y los valores que se pueden encontrar en ellos. Al hacer este análisis es importante identificar la calidad de los datos y detectar errores. Por último, se pueden hacer clasificaciones de pacientes, y de sus citas, que sirvan para determinar qué pacientes (o citas) se introducen en la herramienta de minería de procesos.

#### El corpus de datos

El corpus proporcionado por el HUPLAFE recoge, del sistema de información del hospital, todos los datos relativos a Consultas Externas del día 14 de noviembre de 2016. Para el resto del análisis hay que tener en cuenta las siguientes características del corpus:

- Hay un total de 5377 citas registradas (filas)
- Hay un total de 54 campos (columnas)
- Hay un total de 4498 pacientes
- Número de citas por paciente:
  - Pacientes con 1 cita: 3780 (84%)
  - Pacientes con 2 citas: 589 (13.10%)
  - Pacientes con 3 citas: 108 (2.43%)
  - Pacientes con 4 citas: 13 (0.30%)
  - Pacientes con 5 citas: 5 (0.10%)
  - Pacientes con 6 citas: 3 (0.07%)

La mayoría de pacientes que acude a Consultas Externas ese día (84%), únicamente tienen una cita, sea programada o no. Eso quiere decir que la mayoría de muestras que tendremos en el flujo de pacientes acudirán a un local y se irá. Por otra parte, el 16% de los pacientes se moverán por las consultas al tener más de una cita y podremos ver qué servicios están más relacionados o qué recorridos entre locales son los más comunes, entre otras cosas.

#### Los campos del corpus

A continuación, se muestran los campos del corpus (Tabla 3). Se somborean los campos que aportan más información del proceso de estudio y aquellos que se han utilizado en fases posteriores del trabajo o añaden características a la prestación.

CAMPO	TIPO	RANGO	DESCRIPCIÓN
<b>Ubicación</b>	Entero	0	El hospital
<b>NumeroRegistro</b>	Entero	p.e 1728567 Con celdas vacías	Código del registro
<b>ID nuevo</b>	Entero	1-4533	ID por paciente diferente al original
<b>FechaSolicitud</b>	Fecha dd-mm-aaaa	Desde noviembre de 2014 hasta noviembre de 2016	Se solicita la cita. Coincide con fechaRegistroCita
<b>FechaRegistro</b>	Fecha dd-mm-aaaa	Desde noviembre de 2014 hasta noviembre de 2016	La cita se registra. Coincide con fechaRegistroCita
<b>FechaAceptacion</b>	Fecha dd-mm-aaaa	Desde noviembre de 2014 hasta noviembre de 2016	La cita se acepta. Coincide con fechaRegistroCita
<b>MotivoRechazo</b>	-	Vacío	Porqué no se acepta la cita. No hay ningún caso
<b>idmotivorechazo</b>	-	Vacío	Identificador del motivo de rechazo
<b>Procedencia</b>	Entero	0-4 Con celdas vacías	Indica de dónde vienen los pacientes (si es de otro servicio del mismo centro, de otro centro, de urgencias, etc.)
<b>idprocedencia</b>	Entero	1-5 Con celdas vacías	Identificador de la procedencia
<b>Peticionario</b>	-	Vacío	Persona solicitante de la cita
<b>Prestacion</b>	Texto	p.e 2.1COA	Prestación realizada al paciente, lo que se le ha hecho. Ejemplos: ECG - electrocardiograma, ANALI - analítica, 2.1 - primera visita,

			2.2 - segundas
<b>Preferencia</b>	Texto	N, P	Si el paciente tiene preferencia con respecto a los demás, o no, es decir, si tiene que ser atendido antes que el resto
<b>idprioridad</b>	Entero	1, 2	Código con el que se registra la preferencia
<b>TipoTecnica</b>	-	Vacío	Tipo de técnica
<b>FechaCita</b>	Fecha dd-mm-aaaa	14-11-2016	Día en el que tiene lugar la visita/cita
<b>idcentro</b>	Entero	1	El hospital
<b>servicio</b>	Texto	p.e CCAI	Las consultas de cada servicio. Los gestores del HUPLAFE proporcionaron al proyecto un documento con la localización y el nombre completo de cada servicio. Ejemplos: CONC - Consulta Oncología Médica, CCAR - Consulta Cardiología
<b>idprestacion</b>	Entero	p.e 3473	Código con el que se registra la prestación
<b>fechaRegistroCita</b>	Fecha dd-mm-aaaa	Desde noviembre de 2014 hasta noviembre de 2016	Fecha en la que se le dio la cita al paciente. Si se le han hecho varias reprogramaciones (cambiarle la cita de día), se guarda en este campo cuándo se hizo la última de todas
<b>fechabusquedacita</b>	Fecha dd-mm-aaaa	Desde noviembre de 2015 hasta noviembre de 2016. Con celdas vacías	Los gestores del HUPLAFE desconocen el origen de este campo

<b>Citado</b>	Entero	0, 1	Si es 1, tenía cita previa, era programada
<b>FechaInicio</b>	Fecha dd-mm-aaaa	Desde noviembre de 2014 hasta noviembre de 2016	Cuando comienza el registro de la cita. Coincide con fechaRegistroCita
<b>FechaFin</b>	Fecha dd-mm-aaaa	14-11-2016	Día en el que tiene lugar la visita/cita
<b>TipoPrestacion</b>	Texto	ENF, PRI, SUC, TEC	Enfermería, Primera, Sucesivas y Técnicas (pruebas)
<b>Realizada</b>	Texto	N, S, U	Explicado al final de este subapartado
<b>ANYOCITA</b>	Entero	2016	Año en el que tiene lugar la visita/cita
<b>idSolicitud</b>	Entero	p.e 8960840 Con celdas vacías	Identificador de la solicitud
<b>AGENDA</b>	Texto	p.e ALE003	Asociada a un médico
<b>idAgenda</b>	Entero	p.e 4	Identificador de la agenda
<b>HORARIO</b>	Entero	p.e 5345 Con celdas vacías	Horario dentro de la agenda
<b>HorainicioProgramada</b>	Fecha hh:mm		Hora de la cita
<b>HoraFinProgramada</b>	Fecha hh:mm	Con celdas vacías	Hora a la que tiene que acabar teóricamente la cita
<b>HoraLlegada</b>	Fecha hh:mm	Con celdas vacías	Explicado en el apartado <i>Entorno de Trabajo</i>
<b>HoraLIAMADA</b>	Fecha hh:mm	Con celdas vacías	Explicado en el apartado <i>Entorno de Trabajo</i>
<b>HoraEntrada</b>	Fecha hh:mm	Con celdas vacías	Explicado en el apartado <i>Entorno de Trabajo</i>
<b>HoraSalida</b>	Fecha hh:mm	Con celdas vacías	Explicado en el apartado <i>Entorno de Trabajo</i>
<b>NOPRESENTADOS</b>	Entero	0, 1	El paciente no ha acudido a la cita
<b>PRESENTADOS</b>	Entero	0, 1	El paciente sí ha acudido a la cita

<b>NOCITADO</b>	Entero	0, 1	Si es 1, no tenía cita previa, no era programada
<b>TiempoEspera</b>	Entero (minutos)		Tiempo desde la hora de inicio programada hasta que entra al local. No es realmente el tiempo que espera el paciente, porque puede haber llegado antes o después de la hora a la que se le había citado.
<b>TIEMPOESPERA30</b>	Entero	0, 1	1: TiempoEspera > 30 minutos 0: TiempoEspera ≤ 30 minutos
<b>TIEMPOESPERA60</b>	Entero	0, 1	1: TiempoEspera > 60 minutos 0: TiempoEspera ≤ 60 minutos
<b>MESCITA</b>	Entero	11	Mes numérico en el que tiene lugar la visita
<b>idturno</b>	Entero	1, 2, 3	Indica el turno de mañana, tarde o noche
<b>idfranja</b>	Entero	p.e 2	Es la franja horaria de realización de la prestación
<b>TIEMPOESPERA15</b>	Entero	0, 1	1: TiempoEspera > 15 minutos 0: TiempoEspera ≤ 15 minutos
<b>keycitadia</b>	Texto	p.e 0_ALE003_2016-11-14	Combinación de la fecha de la cita y la agenda donde está registrada
<b>Duracion</b>	Entero (minutos)		HoraSalida – HoraEntrada. Duración de la prestación en función de las horas registradas
<b>DuracionPrevista</b>	Entero (minutos)	Con celdas vacías	HoraFinProgramada - HorarInicioProgramada. Cuánto tiempo está programado que dure la prestación
<b>NumReprogramaciones</b>	Entero	0, 1, 2, 3	Indica el número de veces que se le cambia la fecha de cita al paciente

Local	Texto	p.e B001	Explicado en el apartado <i>Entorno de Trabajo</i>
Keycitadialocal	Texto	p.e 0_B001_2016- 11-14	Combinación de la fecha de la cita y el local donde se atiende al paciente
idlocal	Entero	p.e 31	Identificador del local

Tabla 3. Campos del corpus proporcionado por el HUPLAFE.

### Campo “Realizada”

Gracias al análisis de los datos se puede ver la relación que existe entre algunos de los campos. En este caso, se presenta la relación entre los campos *Realizada*, *PRESENTADOS*, *NO PRESENTADOS* y *Citado*. Junto a ellos, el campo *Realizada* aporta información muy útil para clasificar a los pacientes y para conocer cómo ha sido la cita y, por tanto, el proceso en sí.

Sabiendo que *Realizada* se refiere a la **prestación**, aparecen las siguientes tres opciones:

- **Realizada = S** (la prestación se realiza). Todas las visitas dentro de este grupo tenían cita previa (*Citado* = 1) y el paciente se ha presentado en el local (*PRESENTADOS* = 1).
- **Realizada = N** (la prestación no se realiza). Todas las visitas dentro de este grupo tenían cita previa (*Citado* = 1), pero el paciente no se ha presentado (*PRESENTADOS* = 0).
- **Realizada = U** (la prestación se ha realizado de forma urgente/imprevista). Todas las visitas que aparecen como no citadas (*Citado* = 0) son urgentes y todas las urgentes son no citadas. En adelante, se tratarán como sinónimos. Por su parte, los campos *PRESENTADOS* y *NO PRESENTADOS* se encuentran a 0.

### Clasificación de las citas

El proceso de Consultas Externas, definido en el apartado *Entorno de trabajo*, se registra en el sistema de información a través de las **cuatro horas principales: Hora de Llegada, Hora de Llamada, Hora de Entrada y Hora de Salida**. En función de qué horas tenga guardadas cada cita, podemos establecer hipótesis del recorrido que hizo ese paciente o cómo fue su visita:

1. 3070 casos (57.1%): **Se tienen registradas las 4 horas**. Por lo tanto, son casos que siguen el proceso estándar, es decir, se ha guardado el momento en el que pasa el SIP por el Punto de Atención, cuándo su código aparece por pantalla, cuándo entra por la puerta del local correspondiente y cuándo sale del mismo.



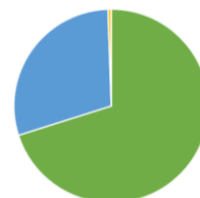
- 1.1. *Realizada* = S (la prestación se realiza). 3012 casos (98.1%). Estas son las citas que representan los casos ideales, dentro de las visitas con cita previa: el paciente ha acudido a la cita y se han guardado las cuatro horas.



1.2. *Realizada* = U (la prestación se ha realizado de forma urgente). 55 casos (1.8%). A veces, no en muchos casos como se observa, se cita al paciente urgente en el momento que llega. De esta forma se registran las cuatro horas como si fuera un paciente con cita previa y los gestores tienen información más fiable de los tiempos de consulta.

1.3. *Realizada* = N (la prestación no se realiza). 3 casos (0.1%). Para poder el facultativo realizar la prestación al paciente, ha de pedir el primero la *Solicitud de prestación*. Los gestores apuntan a que estos casos, aunque muy pocos, pueden ser citas en las que se haya recibido al paciente en el local, pero no se ha podido guardar la prestación por no tener la solicitud comentada. Se originan por esta causa citas sin ninguna de las cuatro horas registradas y con la prestación realizada, porque se registra la prestación de ese paciente en otra cita con la solicitud, aunque no esté el paciente en ese momento.

2. 1127 casos (20.9%): **No hay hora de llegada ni de llamada, solo entrada y salida del local.** Si el paciente es “No citado” (o la prestación es realizada “U”) los facultativos anotan a posteriori las horas de entrada y salida únicamente. Concretamente, se pone de hora de entrada la hora a la que empezó la prestación y para la salida se suma la duración teórica. Se estudian a continuación los subgrupos presentes en estos casos:

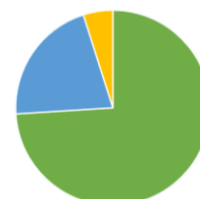


2.1. *Realizada* = U (la prestación se ha realizado de forma urgente). 790 casos (70.1%). Asociados a la explicación antes comentada.

2.2. *Realizada* = S (la prestación se realiza). 331 casos (29.4%). Los gestores no tienen una justificación clara de este fenómeno. Al tener solo estas dos horas, indica que el paciente no ha pasado el SIP por el Punto de Atención al Llegar y por lo tanto no ha aparecido su código en la pantalla y no lo ha podido llamar el médico. Lo que parece es que el paciente ha llegado a la zona de consultas o técnicas y ha pasado directamente a un local, se ha saltado los dos pasos anteriores o no funcionaban en ese momento.

2.3. *Realizada* = N (la prestación no se realiza). 6 casos (0.5%). Son casos muy extraños que pueden coincidir con la posibilidad comentada anteriormente: se ha recibido al paciente, pero no se ha podido registrar la prestación. Aun así, si realmente ha acudido el paciente al local, los dos primeros pasos del proceso se los ha saltado o no funcionaban en ese momento.

3. 773 casos (14.4%): **No se han registrado ninguna de las 4 horas.** Puede que sean pacientes “No presentados”, que no han acudido a la cita y, por tanto, no se registra ninguna hora relativa a la situación del paciente en el hospital. Se estudian a continuación los subgrupos presentes en estos casos:

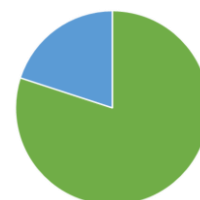


- 3.1. *Realizada* = N (la prestación no se realiza). 572 casos (74%). Sería la situación comentada como razón principal.
- 3.2. *Realizada* = S (la prestación se realiza). 163 casos (21.1%). Estas son las citas que se pueden generar si un paciente va a la consulta, pero no pueden almacenar su prestación y se guarda en otra cita sin registrar las horas del paciente en el proceso. Otra opción es que se haya colado antes, la prestación ya se ha realizado y en el registro de la cita que tenía se guarda la prestación. Otra opción es que es simplemente un error de registro y no se guardaron las horas o se borraron después.
- 3.3. *Realizada* = U (la prestación se ha realizado de forma urgente). 38 casos (4.9%). De nuevo otro caso anormal. Si es urgente, se supone que al menos se guardan las horas de entrada y salida, pero en estas citas no se han registrado ninguna de las cuatro. Se desconoce la causa, no obstante, puede tratarse de un error de registro y no se guardaron las horas o se borraron después.
4. 372 casos (6.9%): **No hay hora de llamada, pero sí de llegada, entrada y salida.** Se debe a que el facultativo no llama al paciente a través del sistema y entonces no se guarda la hora de llamada, pero tampoco la pone después para que quede constancia. Un motivo de que esto ocurra puede ser porque no funcione en esos momentos la pantalla de la sala de espera.



Pese a fallar la hora de llamada, son casos que siguen los pasos del protocolo de Consultas Externas al tener las otras horas registradas: llega, espera en la sala, entra al local y sale de él.

- 4.1. *Realizada* = S (la prestación se realiza). 335 casos (90.1%).
- 4.2. *Realizada* = U (la prestación se ha realizado de forma urgente). 37 casos (9.9%).
5. 35 casos (0.7%): **No hay hora de llegada, pero sí de llamada, entrada y salida.** Lo que reflejan los datos es que los pacientes involucrados no pasan su SIP por el Punto de Atención y, como consecuencia, no aparece en el monitor del médico que ha llegado el paciente. Sin embargo, sí que han sido llamados a través del sistema (código en la pantalla de la sala de espera) o los facultativos han puesto la hora de llamada a posteriori. La razón de que el paciente haya sido llamado pese a no constar la llegada en la pantalla del médico, puede ser porque han visto físicamente que estaba allí. Otra opción es que sea un error del sistema y realmente sí que pasaron el SIP, pero no se ha registrado la hora o se ha borrado más tarde.



- 5.1. *Realizada* = S (la prestación se realiza). 28 casos (80%).
- 5.2. *Realizada* = U (la prestación se ha realizado de forma urgente). 7 casos (20%).

En la figura 16, el 57% de las citas del día 14 de noviembre del 2016 se registraron completas, es decir, el proceso de consultas externas se siguió y registró correctamente. El 21%, que corresponde a citas con constancia únicamente de la entrada y salida del paciente del local, en su mayoría (70%) son aquellos que fueron de forma urgente e imprevista a la consulta o la técnica y los gestores son conocedores de que en estos casos la cita solo reflejará la entrada y salida del local. Por otra parte, el 14% de las citas son las que no tienen guardadas ninguna de las cuatro horas principales, los gestores atribuyen estos casos a que el paciente no se ha presentado a su cita, y es cierto, pero en un 74% de las citas. Por último, el resto de citas (8%) se consideran con errores. Tal y como se han descrito al desarrollar cada tipo de cita, los errores pueden ser causados por los facultativos, el paciente que se salta algún paso o el propio sistema informático tanto software, programa de llamada de pacientes, por ejemplo, o hardware, la pantalla de la sala de espera o el Punto de Atención.

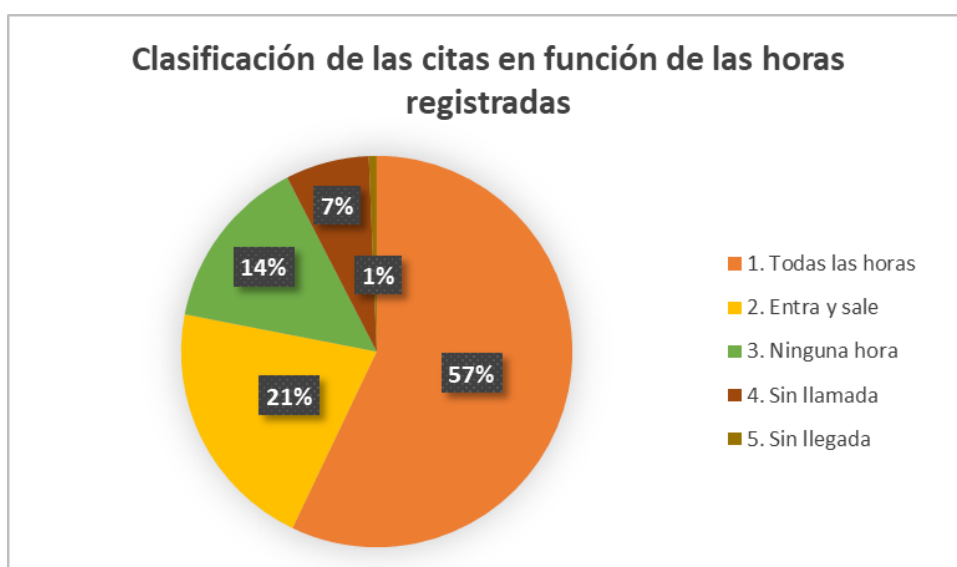


Figura 16. Porcentajes de cada tipo de cita en el corpus.

Calidad de los datos y errores

#### *Fiabilidad de las cuatro horas principales*

Se han estado utilizando estas cuatro horas para conocer cómo habían sido las citas de los pacientes. Dado que reflejan el flujo de los pacientes por el proceso, es importante estudiar su fiabilidad para tener en cuenta cuáles de ellas podemos utilizar más adelante en el modelado.

**Hora de llegada:** La hora va del Punto de Atención, cuando el paciente pasa el SIP, al programa de llamada de pacientes del médico, dónde se registra. La hora de llegada se asocia a la cita correspondiente porque la elige el paciente en el Punto de Atención. Por lo tanto, si el paciente está pasando el SIP en cuanto llega y antes de pasar a la sala de espera, sí que representa este campo con exactitud el momento de llegada del paciente. No obstante, como se ha visto anteriormente, no todas las citas tienen hora de llegada.

**Hora de llamada:** También se considera fiable porque el facultativo llama al paciente a través del sistema y se guarda en ese momento la hora. No obstante, como se ha visto anteriormente, no todas las citas tienen hora de llamada.

Analizando la **relación entre la hora de llamada y la de entrada** se obtiene que:

- Hay 2624 citas (el 49% de 5377) en las que **coinciden** la hora de llamada y la de entrada. Esta similitud tiene sentido en el proceso, ya que desde que llaman al paciente hasta que entra al local no tiene por qué pasar ni un minuto.
- Hay 1 única cita en la que la **hora de entrada es menor a la hora de llamada**, es decir, parece que ha entrado y luego lo han llamado. No tiene sentido. Para no tener problemas luego con esta cita, se ha puesto de hora de llamada la de entrada en función de lo observado al estudiar el resto de horas de la cita.
- Hay 480 citas (9% de 5377) en las que la **hora de entrada es mayor a la hora de llamada**. El orden de las horas tiene sentido, primero lo llaman y luego entra, pero se quiso estudiar cuánto tiempo pasaba desde la llamada a la entrada:
  - Menos de 20 minutos en 437 citas (91% de estas 480). En la mayoría de las citas en las que el paciente tarda en entrar al local, lo hace dentro de los 20 minutos. Se suponen las horas como correctas y que realmente puede haber tardado ese tiempo al estar ausente el paciente en el momento de la llamada y no enterarse, pero finalmente entrar.
  - En las 43 citas restantes (9%) la diferencia es superior a 20 minutos, sobrepasando en 8 casos la hora de diferencia. Sin duda los gestores consideran anómalos estos intervalos de tiempo entre que el paciente es llamado y entra. Por ese motivo la fiabilidad de la hora de llamada se pone en duda (la de entrada en estos casos se considera la correcta).
- El resto de citas, 2272 (42% de 5377), son las que **no tienen hora de llamada** registrada.

**Hora de entrada:** Es considerada por los expertos la más fiable de las cuatro. Todas las citas registradas tienen esta hora guardada (excepto las que no tienen ninguna de las cuatro). El médico la registra en el sistema cuando el paciente entra por la puerta y, además, en el caso de que no lo haga, se copia la de llamada como hora de entrada.

**Hora de salida:** Esta hora es la menos fiable de todas. Como se explicaba en el apartado *Entorno de trabajo*, hay tres opciones para registrar la hora de salida al acabar la prestación en el local: marcar como finalizada la prestación, enviar a citación al paciente y firmar el informe. El problema reside en que no se sabe la hora registrada de cuál de las tres opciones proviene. De forma que, por ejemplo, si el médico despacha al paciente, pero no ha finalizado ni lo ha enviado a citación y se deja el informe para acabarlo más adelante, cuando firme el informe habrá pasado mucho tiempo desde que realmente salió el paciente del local.

### *Problemas generados por la hora de salida*

La poca fiabilidad de la hora de salida causa que no se pueda saber cuánto tiempo dura la prestación, cuánto tiempo pasa el paciente en local. Habrá casos (no se sabe cuántos) en los que **la hora de salida será mucho mayor a la real**. Este fenómeno provoca que, si el paciente tiene varias citas ese día, acuda a la segunda estando registrado todavía que permanece en el local de la primera. Del mismo modo ocurre si lo vemos desde el punto de vista de los locales y su ocupación, pueden entrar los siguientes pacientes, pero, como aún consta que está dentro el primero, se registra que hay más de un paciente en el local.

PALIA, la herramienta de minería de procesos que se usará más adelante, es **sensible a los solapes en el tiempo**, es decir, dentro de un mismo paciente (ID/muestra) los eventos que se siguen en el tiempo no pueden solaparse. Cada evento debe empezar después del anterior y acabar antes del siguiente. Por este motivo, las citas de un mismo paciente que se solapen en el tiempo por tener mal la hora de salida harán saltar una alarma en PALIA. La herramienta tiene esta funcionalidad para avisar que en los datos hay una incoherencia, un paciente no puede estar en dos eventos a la vez.

La solución que presentan los gestores ante este problema es hacer una **depuración de la hora de salida**. Consiste en ver las citas ordenadas por local, cambiar la hora de salida de las citas que duran más allá de la siguiente cita y poner como hora de salida la hora de entrada del siguiente que entra en el local (figura 17). De esta forma se van corrigiendo las horas de salida y se ponen los pacientes en cada local uno detrás de otro en el tiempo. Las horas que se obtienen siguiendo este método no tienen porqué ser un reflejo exacto de la realidad, pero sí son mejores que las anteriores al evitar que haya dos pacientes en el mismo local a la vez.

#### **Depurar\_h\_salida.m:**

```
[datos,txt]=xlsread('Datos_depurar_h_salida.xlsx','Hojal');
tamano_tabla=size(datos);
num_citas=tamano_tabla(1);

%Leemos los campos
ID=datos(:,1);
in_h_entrada=datos(:,2);
in_h_salida=datos(:,3);
local=txt(2:end,4);

%Constantes
un_seg=0.0000115740740740741;

%Leemos la cabecera
c_ID=strjoin(txt(1,1));
c_h_entrada=strjoin(txt(1,2));
c_h_salida=strjoin(txt(1,3));
c_local=strjoin(txt(1,4));

out_h_salida=[];
```

```
for i=1:num_citas-1 %cada fila menos la última
    if strjoin(local(i))==strjoin(local(i+1)) %no importa la h_salida
del último paciente del local
        if in_h_entrada(i) >= in_h_entrada(i+1) %si coinciden las
horas de entrada o es menor la siguiente
            in_h_salida(i)=in_h_entrada(i)+un_seg; %añado un segundo
            in_h_entrada(i+1)=in_h_salida(i)+un_seg; %añado un segundo
            if in_h_entrada(i+1) > in_h_salida(i+1) %se igualan
                in_h_salida(i+1)=in_h_entrada(i+1);
            end
        elseif in_h_entrada(i)==in_h_salida(i) %se corrige
            in_h_salida(i)=in_h_entrada(i)+un_seg;
        end
        if in_h_salida(i) >= in_h_entrada(i+1)
            out_h_salida(i)=in_h_entrada(i+1)-un_seg; %detrás
        else
            out_h_salida(i)=in_h_salida(i);
        end
    else
        out_h_salida(i)=in_h_salida(i);
    end
end

%última cita/fila
out_h_salida(num_citas)=in_h_salida(num_citas);

cabecera_salida={c_ID,c_h_entrada,c_h_salida,c_local,'out_h_salida'};

xlswrite('depurado_h_salida_0.xlsx',ID,['A2:A' num2str(num_citas+1)])
xlswrite('depurado_h_salida_0.xlsx',in_h_entrada,['B2:B'
num2str(num_citas+1)])
xlswrite('depurado_h_salida_0.xlsx',in_h_salida,['C2:C'
num2str(num_citas+1)])
xlswrite('depurado_h_salida_0.xlsx',local,['D2:D'
num2str(num_citas+1)])
xlswrite('depurado_h_salida_0.xlsx',out_h_salida,['E2:E'
num2str(num_citas+1)])

xlswrite('depurado_h_salida_0.xlsx',cabecera_salida,'A1:E1')
```

Figura 17. Script de Matlab para depurar la hora de salida

### Errores

En minería de procesos es **importante corregir los errores** para obtener modelos más fáciles de entender por los expertos y evitar el llamado *Spaghetti Effect* (Fernandez-Llatas y otros, 2015). A parte de los errores que han ido apareciendo en el análisis de los datos, se han detectado los siguientes:

- Dos citas que no tienen registrado a qué hora está programado que acabe la prestación. Campo HoraFinProgramada vacío. No se ha intentado corregir el error porque la hora de fin programada no se utiliza en pasos posteriores y la duración teórica que se podría calcular con ella no interesa incorporarla en el estudio.
- Con HoraFinProgramada errónea, valor no válido. No se han corregido por el motivo anterior:
  - En 3 citas la hora de fin programada es 31:30:00.
  - En 1 cita la hora de fin programada es 32:30:00.
- Una cita en la que la hora de llamada es menor a la de llegada. Una cita en la que la hora de entrada es menor a la hora de llamada. Cuatro citas en las que la hora de salida es menor a la de entrada. Estos casos se han estudiado uno a uno viendo si se podía cambiar algún valor y quedarse la cita en el corpus o por el contrario se eliminaban.
- Un tipo de error que provoca solapes en el tiempo y, por tanto, error en PALIA, son las **citas duplicadas**. Una vez depurada la hora de salida y sin tener en cuenta las citas que no tenían registradas ninguna de las cuatro horas y las de los pacientes que no se presentaron, se detectó con Matlab que había 39 citas duplicadas.

Hay pacientes, de los que acuden a varias citas, que tienen citas duplicadas. Se han considerado que son aquellas citas consecutivas que tienen lugar en el mismo local y el tiempo que transcurre desde que ha estado en una hasta que entra en la otra es menor de un minuto. Si no fuera así, se puede entender que ha entrado dos veces al mismo local sin tratarse de un error, pero si están registradas justo una detrás de otra se consideran errores y basta con que quede constancia que ha acudido a ese local una vez. A continuación, se muestra un **ejemplo** de un paciente con varias citas duplicadas:

El **paciente con ID 2920** tiene **6 citas registradas**, pero 4 de ellas son en el mismo local y, como cumplen las características comentadas, 3 de ellas las consideramos duplicadas. Las queremos detectar automáticamente y eliminar del corpus. En la Tabla 4, la columna *duplicada* proviene de un script de Matlab diseñado (figura 18) y marca con un 1 las citas duplicadas. En la tabla 5, se han eliminado las citas marcadas en la gráfica anterior como duplicadas. El problema que nos encontramos ahora al tener solo estas tres citas es que hay un solape entre las dos últimas. Con otro script en Matlab (figura 19) corregimos este error poniendo una delante de la otra en el tiempo (tabla 6).

ID nuevo	local	HoraEntrada	HoraSalida	<i>duplicada</i>
2920	D209	9:06:00	9:16:00	0
2920	D206	10:41:00	10:55:59	0
2920	D208	10:53:00	10:53:01	1
2920	D208	10:53:02	10:53:03	1
2920	D208	10:53:04	10:53:05	1
2920	D208	10:53:06	10:53:59	0

Tabla 4. Gráfica que muestra las citas del paciente ID 2920 y tiene marcadas qué citas son las duplicadas.

HoraEntrada	HoraSalida
9:06:00	9:16:00
10:41:00	10:55:59
10:53:06	10:53:59

Tabla 5. Citas del paciente ID 2920 una vez eliminadas las duplicadas.

HoraEntrada	HoraSalida
9:06:00	9:16:00
10:41:00	10:55:59
10:56:01	10:56:54

Tabla 6. Citas corregidas del paciente ID 2920.

```

Detecta_citas_duplicadas.m:

[datos,txt]=xlsread('datos_limpio_h_salida_depurada.xlsx','Hojal');
tamano_tabla=size(datos);
num_citas=tamano_tabla(1);
%las citas están ordenadas por ID

%Leemos los campos
ID=datos(:,1);
in_h_entrada=datos(:,34);
in_h_salida=datos(:,35);
local=txt(2:end,50);
    
```



```
%Constantes
un_seg=0.0000115740740740741;
un_min=0.0006944444444444444;

%Leemos la cabecera
c_ID=strjoin(txt(1,1));
c_h_entrada=strjoin(txt(1,34));
c_h_salida='HoraSalida';
c_local=strjoin(txt(1,50));

duplicada=zeros(num_citas,1);

%Si acaba de estar en ese local, hace menos de un minuto (de forma
%consecutiva), se elimina la cita anterior porque se considera que
está duplicada
%y basta con que quede constancia una vez de que ha acudido a ese
local.

%Primero se ha de cambiar las horas de salida de aquellas citas en las
%que coincida la hora de entrada con la de salida
n=0;
for i=1:num_citas %cada fila
if in_h_entrada(i) == in_h_salida(i)
    in_h_salida(i)=in_h_entrada(i)+un_seg;
    n=n+1;
end
end

for i=2:num_citas %cada fila menos la primera
    if ID(i) == ID(i-1) %cogemos todas las citas del paciente menos la
primera
        if strjoin(local(i)) == strjoin(local(i-1)) %el local actual
coincide con el anterior
            if in_h_entrada(i) - in_h_salida(i-1) < un_min
                duplicada(i-1)=1;
            end
        end
    end
end
end

cabecera_salida={c_ID,c_h_entrada,c_h_salida,c_local,'duplicada'};

xlswrite('elimina_citas_duplicadas.xlsx',ID,['A2:A'
num2str(num_citas+1)])
xlswrite('elimina_citas_duplicadas.xlsx',in_h_entrada,['B2:B'
num2str(num_citas+1)])
xlswrite('elimina_citas_duplicadas.xlsx',in_h_salida,['C2:C'
num2str(num_citas+1)])
xlswrite('elimina_citas_duplicadas.xlsx',local,['D2:D'
num2str(num_citas+1)])
xlswrite('elimina_citas_duplicadas.xlsx',duplicada,['E2:E'
num2str(num_citas+1)])
```

```
xlswrite('elimina_citas_duplicadas.xlsx',cabecera_salida,'A1:E1')
```

Figura 18. Script de Matlab que detecta las citas duplicadas.

```
Corrige_solapes.m:  
  
[datos,txt]=xlsread('datos_limpio_h_salida_depurada_sin_duplicadas.xlsx','Hojal');  
tamano_tabla=size(datos);  
num_citas=tamano_tabla(1);  
%las citas están ordenadas por ID  
  
%Leemos los campos  
ID=datos(:,1);  
in_h_entrada=datos(:,34);  
in_h_salida=datos(:,35);  
  
%Constantes  
un_seg=0.0000115740740740741;  
un_min=0.0006944444444444444;  
  
%Leemos la cabecera  
c_ID=strjoin(txt(1,1));  
c_h_entrada=strjoin(txt(1,34));  
c_h_salida=strjoin(txt(1,35));  
  
solapado=zeros(num_citas,1);  
out_h_entrada=in_h_entrada;  
out_h_salida=in_h_salida;  
  
for i=2:num_citas %cada fila menos la primera  
    if ID(i) == ID(i-1) %cogemos todas las citas del paciente menos la  
        primera  
            s=[out_h_entrada(i) out_h_salida(i)]; %local actual  
            duracion=in_h_salida(i)-in_h_entrada(i);  
            t=[out_h_entrada(i-1) out_h_salida(i-1)]; %local anterior  
            if s(2)>=t(1) && s(1)<=t(2) %si están solapados  
                solapado(i)=1;  
                out_h_entrada(i)=out_h_salida(i-1)+2*un_seg;  
                out_h_salida(i)=out_h_entrada(i)+duracion;  
            end  
        end  
    end  
end  
  
cabecera_salida={c_ID,c_h_entrada,c_h_salida,'out_h_entrada','out_h_salida','solapado'};  
  
xlswrite('solapes.xlsx',ID,['A2:A' num2str(num_citas+1)])  
xlswrite('solapes.xlsx',in_h_entrada,['B2:B' num2str(num_citas+1)])
```

```
xlswrite('solapes.xlsx',in_h_salida,['C2:C' num2str(num_citas+1)])  
xlswrite('solapes.xlsx',out_h_entrada,['D2:D' num2str(num_citas+1)])  
xlswrite('solapes.xlsx',out_h_salida,['E2:E' num2str(num_citas+1)])  
xlswrite('solapes.xlsx',solapado,['F2:F' num2str(num_citas+1)])  
  
xlswrite('solapes.xlsx',cabecera_salida,'A1:F1')
```

Figura 19. Script de Matlab que detecta los solapes de las citas de un mismo paciente y las corrige.

## 5.2 Resultado de la Preparación de los Datos y el Modelado

Dentro de este apartado se presentan las modificaciones que se hicieron al corpus de datos para **construir los registros de eventos** que necesita PALIA. Se añadieron campos, se transformaron algunos de ellos y se hizo una selección para convertir las citas en eventos. También se presentan los **flujos de pacientes** que muestra PALIA en función del corpus cargado y que se obtuvieron en la fase de modelado.

### De cita a evento

Ahora mismo las filas del corpus son las citas de los pacientes. Cada una tiene los campos para cada hora de la cita:

HorainicioProgramada	HoraFinProgramada	HoraLlegada	HoraLLAMADA	HoraEntrada	HoraSalida
----------------------	-------------------	-------------	-------------	-------------	------------

Los eventos que se introducen en PALIA tienen que tener un inicio (elemento **START**) y un fin (elemento **END**), son intervalos de tiempo. Por este motivo se ha de transformar el corpus de datos en función del evento que queramos definir.

Si interesara tener las horas anteriores en forma de intervalos para que cada evento fuera una etapa del proceso asistencial, **una cita estaría formada por tres o cuatro eventos**. Pasaría de tener todos los datos en una fila a estar repartida en tres o cuatro filas. A continuación, se muestra un ejemplo:

————— —————  
**HoraLlegada → HorainicioProgramada**

El paciente llega antes de la hora a la que tiene programada la cita.

**HorainicioProgramada → HoraLLAMADA**

Tiempo de retraso de la cita. Es el tiempo que espera el paciente a causa del hospital.

ó

**HoraLlegada → HoraLLAMADA**

Tiempo de espera. Tiempo que pasa el paciente esperando en la sala de espera.  
————— —————

**HoraLLAMADA → HoraEntrada**

Tiempo de transición desde que el paciente va de la sala de espera al local.

**HoraEntrada → HoraSalida**

Duración de la prestación. Tiempo que pasa el paciente en el local.

Otra opción es que interese que **cada cita sea un único evento**. En este caso, se elegiría cuál de todos los intervalos anteriores es el más adecuado para representar a la cita. Variaría en función de lo que se quisiera estudiar: el tiempo de espera, el tiempo de retraso, la duración de la prestación, etc.

Como se ha comentado antes, PALIA es sensible a los **solapes** en el tiempo de los eventos de un mismo paciente y los considera errores. Por lo tanto, es importante que cuando se creen estos intervalos se tenga en cuenta este hecho. Las horas de la cita están en formato hh:mm, por lo que si el tiempo de fin de un evento coincide con el de inicio de otro (también se consideran solapados) podemos **diferenciarlos con los segundos**. Por ejemplo, 15:30:00-15:30:01, el intervalo de tiempo entre un evento y otro sería de un segundo únicamente y evitaríamos el solapamiento. Por otra parte, también pueden coincidir START y END de un evento y, como ha de ser un intervalo, se haría lo mismo, diferenciarlos un segundo. De esta forma se habría creado un evento de un segundo de duración.

#### Niveles de localización

En el corpus de datos hay un campo que corresponde al **local** donde se realiza la prestación. Ya se comentó en el apartado *Entorno de trabajo* cómo funcionaba la numeración de los locales. El código C703, por ejemplo, incluye la información de dónde está el local dentro del hospital: Torre C, Planta 7, Consulta 03. Por lo tanto, para incorporar esa información en el corpus y tener los diferentes **niveles de localización de los pacientes** (Torre, Planta, Tipo 2), se añaden tres campos más como en la figura 20.

local	TORRE	PLANTA	TIPO 2
B101	B	1	CONSULTA
C215	C	2	CONSULTA
C703	C	7	CONSULTA
C519	C	5	TECNICAS

Figura 20. Ejemplos de locales desglosados en tres columnas para tener los niveles de localización.

## Consultas Externas en PALIA Web

Una vez conocidos los datos y preparados para poder cargarse en la herramienta de minería de procesos, los expertos pueden hacer sus preguntas sobre el proceso para visualizar a través de PALIA los flujos de pacientes que les permitan responder a las mismas. Esta **metodología de minería de procesos basada en preguntas** es común en la mayoría de proyectos de aplicación de técnicas de minería de procesos y fue analizada por Rojas y otros (2016). En este proyecto, las preguntas de los gestores del HUPLAFE son:

- ¿Entre qué dos plantas se mueven más pacientes? Los gestores quieren saber **cómo se mueven los pacientes por Consultas Externas**. ¿Qué desplazamientos son los que más se repiten? ¿Entre qué servicios? Toda esta información les puede servir para estudiar el número de pacientes que va a cada planta de cada torre y adaptar los destinos de los ascensores o reorganizar la distribución de los servicios por el hospital.
- ¿En qué servicios los tiempos de retraso son mayores? ¿Y en cuáles menores? Es decir, ¿en cuáles los pacientes esperan más o menos después de cumplirse la hora de la cita? Los gestores quieren conocer **los tiempos de retraso en cada servicio**. También se preguntan: ¿Los pacientes que más tiempo de retraso tienen acumulado son los que tienen varias consultas/técnicas en el día?

Los **datos de los que se parte** para responder a ambos grupos de preguntas son datos limpios: se han corregido los errores detectados en el análisis de los datos, la hora de salida está depurada y no hay solapes en el tiempo. El siguiente paso es hacer una selección de las citas que se quieren incluir en el corpus para cada propósito. En ambos casos se eliminarán todas aquellas citas cuyos pacientes no se hayan presentado porque quiere decir que no han acudido y que no se ha realizado la prestación. No vamos a tener en cuenta el desplazamiento al lugar de esa cita ni su tiempo de retraso. Eliminaremos también las citas que no tengan registradas ninguna de las cuatro horas principales por falta de datos.

### *Desplazamiento de los pacientes por CEX*

Para **estudiar el movimiento de los pacientes por el hospital** se seleccionan aquellos pacientes que hayan acudido a alguno de los locales de Consultas Externas, ya sea porque tenían cita previa o necesitaban ser atendidos de forma urgente. Si en el corpus inicial había registradas 5377 citas, después se ha limpiado el corpus quitando las citas con errores (3 en total) y las citas duplicadas (39), y se han eliminado los “No presentados” (3+6 citas) y las citas con las cuatro horas principales vacías (773 citas), quedan en el corpus 4553 citas y 3951 pacientes. El **número de pacientes** en el corpus se puede obtener con una sencilla función de Matlab:

```
[entrada,txt]=xlsread('corpus_mov_CEX.xlsx','Hoja1');

%ID en la primera columna
ID=entrada(:,1);

pacientes=unique(ID);

num_pacientes=length(pacientes)
```

No obstante, en relación a las **torres**, se ha visto que, en el corpus a parte de la Torre B y la Torre C (Torres específicas de CEX) hay citas que tienen lugar en las torres D, E y G. Esto se debe a que, pese a estar destinados estos edificios a Hospitalización, hay locales que sirven para pasar consulta. Son muy pocos y los gestores no los están estudiando todavía porque no hay personal que dependa del área ambulatoria (Consultas Externas). Por esos motivos, se eliminan también del corpus y estudiamos los flujos de pacientes entre las torres B y C. Ahora quedan 3996 citas y 3516 pacientes.

Cada cita se convierte en un evento para que indique únicamente en qué lugar se realizó la prestación. El intervalo elegido es **HoraEntrada** → **HoraSalida** que ya tenemos preparado para PALIA porque después de depurar la hora de salida, se eliminaron las citas que tenían duplicados estos intervalos y se corrigieron los solapes (como en el ejemplo del ID 2920) para no tener ningún error en la herramienta de minería de procesos.

Hay que indicarle a PALIA qué tipo de elementos hay en la **cabecera** del archivo que se carga (Tabla 7). En este caso, ID serán los ID's de los pacientes, START será la hora de entrada, END será la hora de salida y habrá dos campos tipo ACTIVITY, la torre y la planta, para situar a los pacientes y sus desplazamientos por las plantas de las torres.

ID	START	END	ACTIVITY	ACTIVITY
ID nuevo	h_entrada	h_salida	TORRE	PLANTA
1	14/11/2016 12:30:00	14/11/2016 12:32:59	B	4
2	14/11/2016 9:14:00	14/11/2016 9:22:00	B	1
⋮	⋮	⋮	⋮	⋮

Tabla 7. Cabeceras del corpus para estudiar los desplazamientos.

Una vez cargado el archivo “.csv” con el corpus de datos descrito y la cabecera indicada, podemos jugar con los **filtros** y con el **mapa de frecuencia** de PALIA para obtener las vistas que permitan a los gestores del HUPLAFE responder a las preguntas que se planteaban acerca del desplazamiento de los pacientes por CEX.

#### Vista 1

En la vista (figura 21) se encuentran los nodos que representan las **localizaciones de los pacientes de la forma Torre\_Planta**. A la izquierda se han colocado de abajo arriba las plantas de la torre B y a la derecha lo mismo con la torre C.

Dejamos de lado los tiempos de los **nodos** porque corresponden a la duración de las prestaciones en esas localizaciones (HoraEntrada → HoraSalida) de forma aproximada y no interesan para responder a las preguntas de los gestores.

El color de las **flechas** representa el número de ocurrencias de la transición, es decir, cuántas veces han pasado por ahí los pacientes, pudiendo haber sido más de una vez el mismo. El color

rojo y el anaranjado que diferencia algunas flechas del resto indica que son las que ocurren con mayor frecuencia. Todas las **transiciones rojas** que se ven en la imagen son saltos dentro de un mismo nodo, es decir, los pacientes van a varios locales de esa Torre\_Planta de forma consecutiva, al menos a dos. Si se pasa el ratón por las flechas, PALIA muestra el número de ocurrencias. Al hacerlo se confirma lo que se ve en la imagen, pocos son los pacientes que cambian de planta y dentro de los que lo hacen hay bastante variabilidad. Las **flechas verdes** más claro únicamente representan el paso de un paciente y las anaranjadas que cruzan de planta son las transiciones de 5-7 pacientes cada una, únicamente.

Se estudian a continuación las **transiciones más comunes**, pero hay que tener en cuenta dos aspectos: solo son los flujos de pacientes de un día y las flechas entre plantas diferentes representan muy pocos pacientes. Pese ambas limitaciones, la herramienta PALIA ha permitido visualizar rápidamente los caminos más comunes y lo podría hacer en estudios con más días y más transiciones entre plantas.

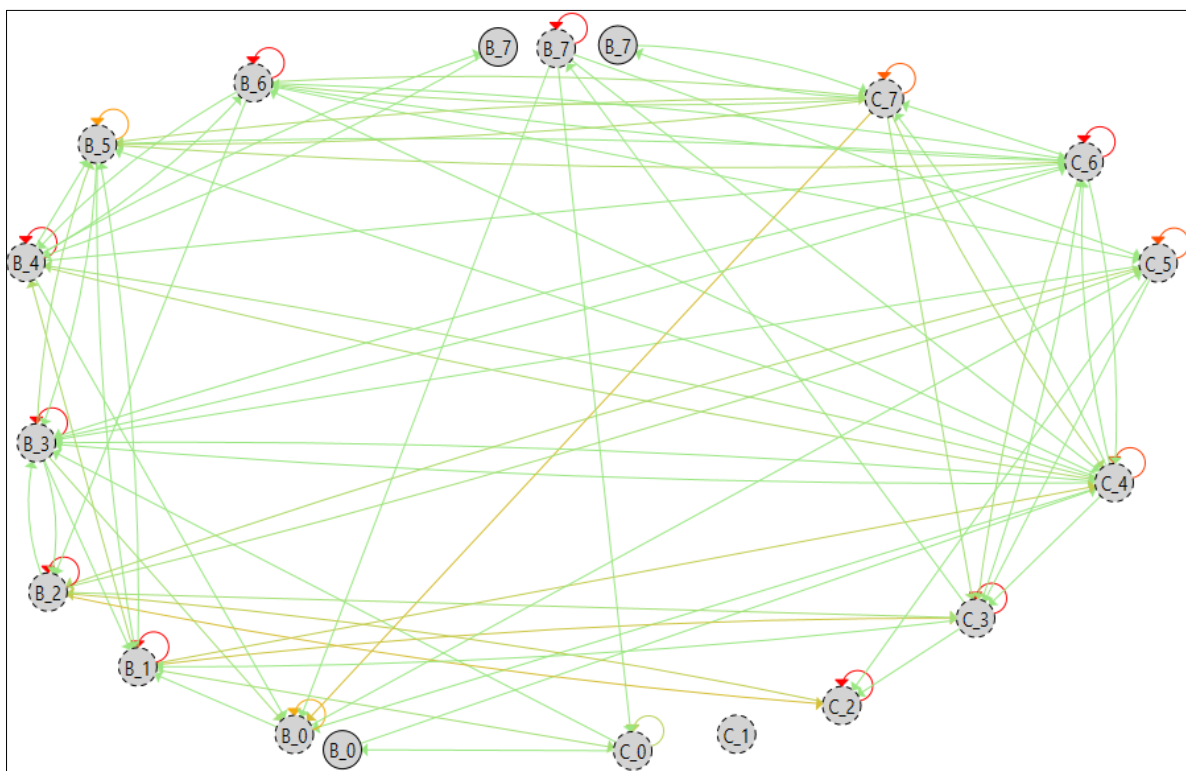


Figura 21. Vista 1

La transición **por la que pasan más veces** los pacientes (50 veces) es la planta 4 de la torre B, dicho de otra forma, la mayoría de pacientes que van a más de una consulta/técnica lo hacen allí (Figura 22). Si se quiere ver **qué servicios** hay en la planta 4 de la torre B, se puede consultar el documento proporcionado por el hospital que contiene la localización de los servicios en las torres de Consultas Externas (Tabla 8).

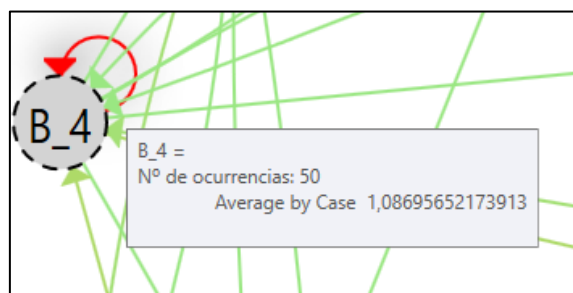


Figura 22. Detalle de la vista 1.

Torre B	Planta 4	Consulta
CCCV. CIR CARDIOVASCULAR		
CCMX. CIR MAXILOFACIAL		
CCVA. ANGIOLOGIA Y CIR VASCULAR		
CUAC. U ARRITMIAS CARDIACAS		
CEPI. MEDICINA PREVENTIVA		
Torre B	Planta 4	Técnica
CCAR. CARDIOLOGIA		

Tabla 8. Servicios situados en la planta 4 de la torre B.

Por otra parte, esta vista sirve para responder a la pregunta que se hacían los gestores de **entre qué dos plantas se mueven más pacientes**. Para ello hay que mirar las transiciones entre nodos diferentes. Se señalan en la figura 23 las flechas más anaranjadas que representan el movimiento más común entre plantas. Son 5 y corresponden a las transiciones siguientes:

1. C7 → B0 (7 veces)
2. B2 → C2 (7 veces)
3. B1 → C3 (6 veces)
4. B1 → C4 (5 veces)
5. C2 → B2 (5 veces)

Gracias a haber visualizado los flujos de pacientes llama la atención que los desplazamientos más comunes entre plantas sean entre plantas de dos torres diferentes, en vez de ser en una misma torre. Entre las 5 transiciones resaltan **C2 → B2 y B2 → C2**, lo que indica que el flujo de pacientes entre las dos torres por la planta 2 puede ser habitual, tanto en un sentido como en el otro. Esta relación tiene sentido puesto que los servicios involucrados son todos los **infantiles**. Para conocer de forma más concreta qué servicios son los que se relacionan en estas transiciones, PALIA tiene una funcionalidad que muestra las **estadísticas de servicios** en el flujo inferido. Lo aplicaremos como ejemplo en el caso de la transición **C7 → B0** (primera de la lista).



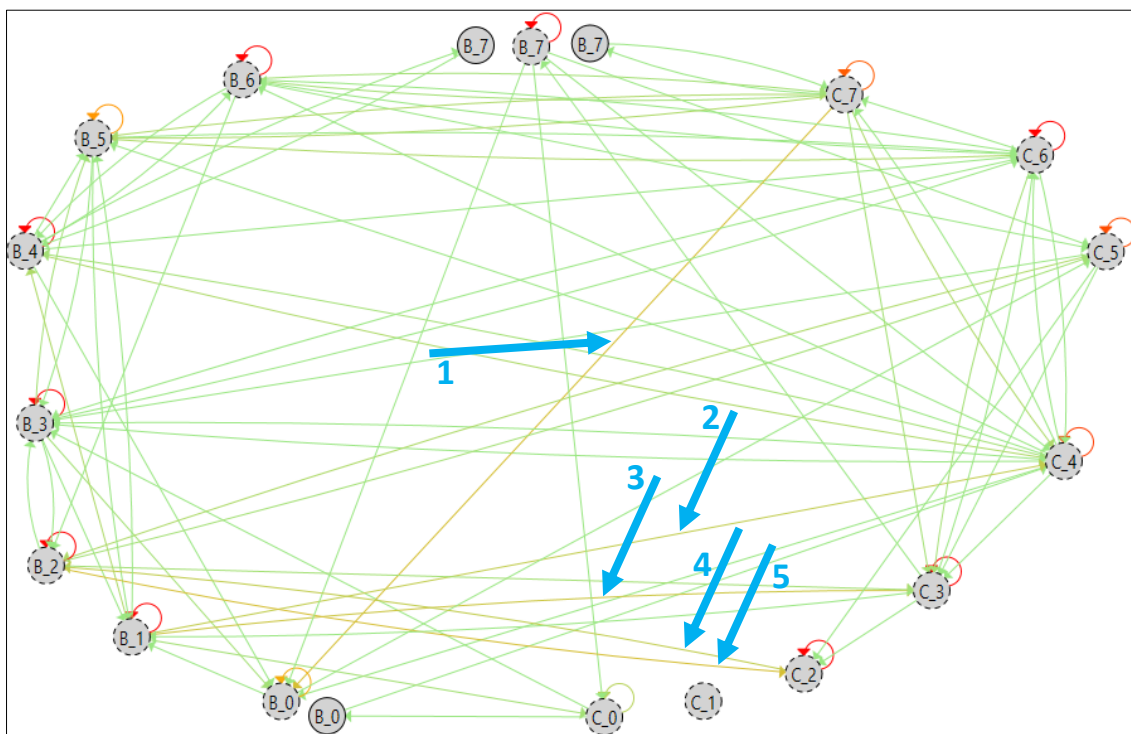


Figura 23. Vista 1 con las transiciones anaranjadas señaladas.

La figura 24 muestra el flujo de los 7 pacientes/muestras que va de C7 a B0. Hemos **seleccionado esta transición** haciendo uso del filtro *Selección de tramos*. Para ver los servicios hemos incorporado este campo como SAMPLE en el corpus y así podemos visualizar, junto al flujo de pacientes, **qué porcentaje de pacientes hay en cada servicio** dentro del camino visualizado. El servicio CRTE (Oncología Radioterápica) se encuentra en B0 y CONC (Oncología médica) y CHEM (Hematología y hemoterapia) en la C7, lo que indica que los pacientes pasan de CONC o CHEM en C7 a CRTE en B0. Vemos que son **pacientes oncológicos** que van en su mayoría de la consulta de oncología a radioterapia. Si ya se quisiera entrar en cada caso concreto, PALIA lista en la parte inferior izquierda los ID's de los pacientes que se muestran en el flujo, de esta forma, para más detalle se podrían buscar en el corpus de datos.

La cantidad de formas de analizar los flujos y de inferir conocimiento se demuestra extensa. Partiendo de los estudios de transiciones concretas que se acaban de realizar, **se da pie a que los gestores del HUPLAFE hagan análisis posteriores** de los movimientos de los pacientes entre plantas. No obstante, la pregunta planteada, ¿Entre qué dos plantas se mueven más pacientes?, ha quedado respondida.

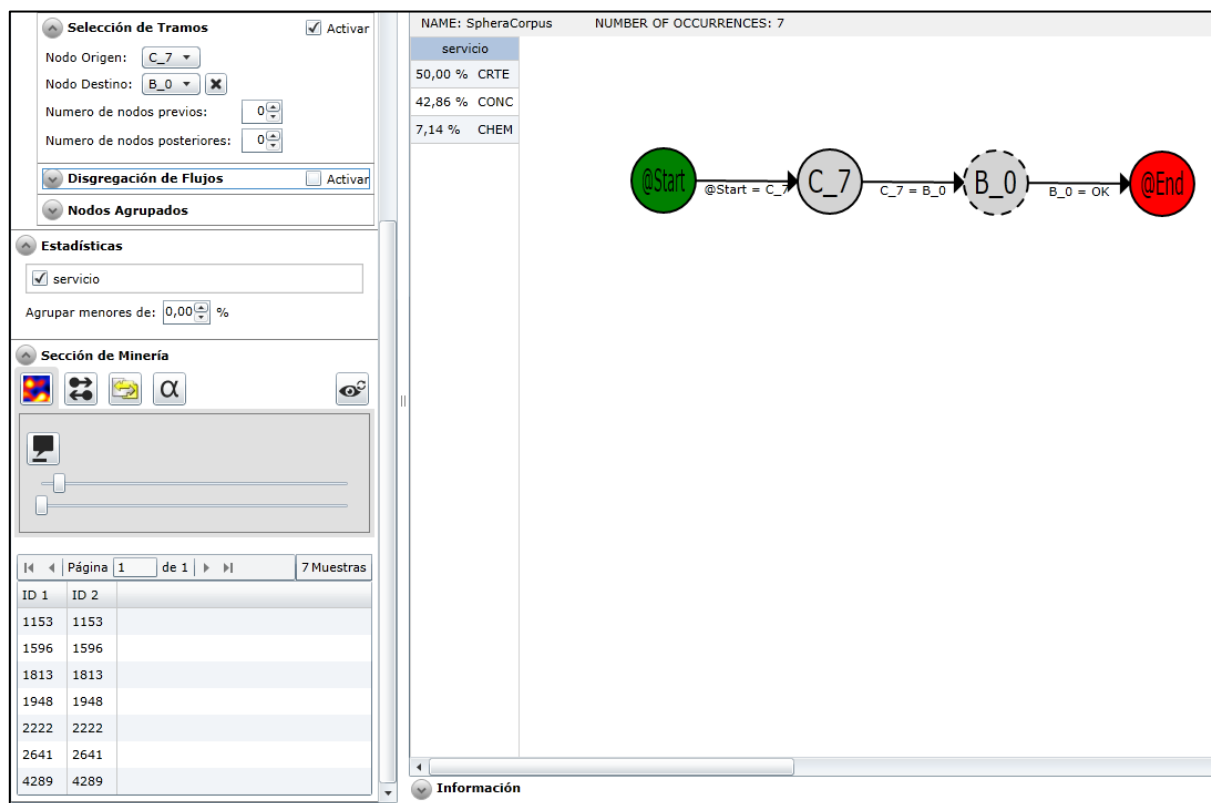


Figura 24. Captura de pantalla de PALIA con uno de los dos tramos más transitados.

## Vista 2

Sobre el mismo corpus con el que se ha trabajado en la vista anterior (3996 citas y 3516 pacientes), se mantiene la hora de entrada como START y se coloca como END la hora de entrada más un segundo. Los elementos de la **cabecera** son los mismos (Tabla 9).

ID	START	END	ACTIVITY	ACTIVITY
ID nuevo	h_entrada	h_salida_modificada	TORRE	PLANTA
1	14/11/2016 12:30:00	14/11/2016 12:30:01	B	4
2	14/11/2016 9:14:00	14/11/2016 9:14:01	B	1
⋮	⋮	⋮	⋮	⋮

Tabla 9. Cabeceras de la vista 2

Este cambio se hace para poder ver a través de los nodos **cuántos pacientes hay en cada Torre\_Planta**, porque el tiempo de los nodos será de un segundo por cada paciente que haya estado allí. Esta información puede ser útil a los gestores del hospital a la hora de determinar dónde para cada ascensor. El **servicio de ascensores** se satura con rapidez y las colas que se forman crean caos entre los pacientes. Por este motivo, la distribución de los ascensores en las plantas está en constante cambio siempre buscando solucionar este problema.

En la figura 25 se pueden visualizar las **plantas de ambas torres** (B izquierda, C derecha). Las flechas de esta vista no nos interesan ya que las estudiamos en la vista anterior. Los nodos, por su parte, se colorean en función del número de pacientes que han acudido ese día a esa Torre\_Planta. PALIA permite conocer el número de muestras/pacientes simplemente pasando el ratón por el nodo. De esta forma se obtiene la siguiente información:

- Los **verdes más claros** son aquellos a los que han acudido menos pacientes, como es el caso de C1 (3 pacientes), C0 (41 pacientes), B0 (77 pacientes) y B1 (89 pacientes).
- En un **color más marrón** podemos destacar B3 (197 pacientes) y B7 (218 pacientes).
- Unos **colores más anaranjados** tienen B4 (313 pacientes) y C5 (320 pacientes).
- Con un **color más rojizo** están C6 (358 pacientes) y C3 (362 pacientes).

No obstante, la ventaja de herramientas de minería de procesos como PALIA es que **visualmente** se puede ver dónde hay más pacientes sin necesidad de acudir a los datos. A continuación, se analizará cada torre por separado.

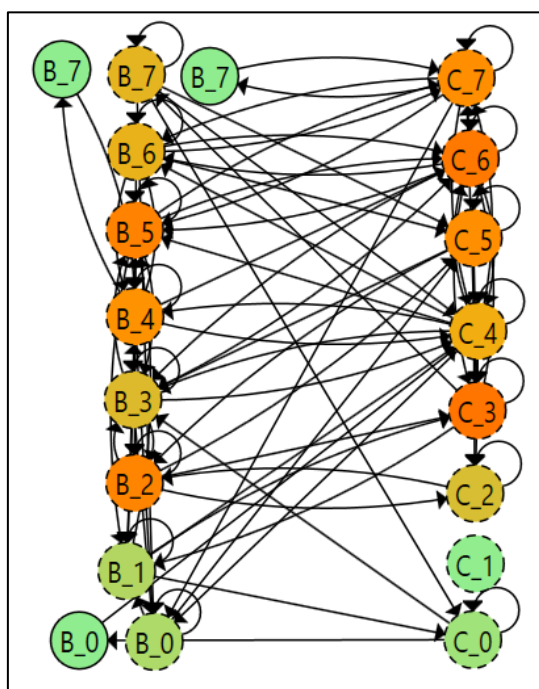


Figura 25. Vista 2 con las dos torres.

**Torre B.** En la figura 26 se han recolocado los nodos de forma que aquellos con mayor número de pacientes se encuentran arriba y a los que han acudido menos pacientes se quedan en la parte inferior. Por lo tanto, teniendo en cuenta que contamos con los datos de un día, se podría gestionar la distribución de los ascensores de forma que las plantas más accesibles fueran la **B5**, la **B2** y la **B4**.

**Torre C.** En la figura 27, de forma análoga a la torre B, se muestran los nodos ordenados por el número de pacientes. Esta vez la distribución de ascensores se debería centrar en las plantas **C3** y **C6**, y en **segundo lugar** en las plantas **C5** y **C7**.

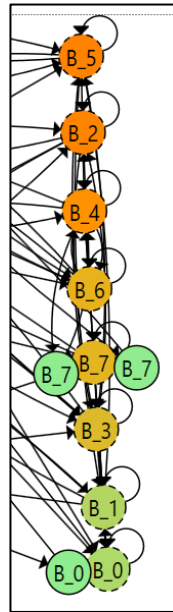


Figura 26: Torre B

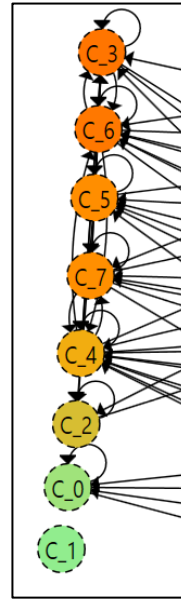


Figura 27: Torre C

### Vista 3

Al ver los gestores la vista anterior, **surgen nuevas preguntas**. Siendo **C3** la Torre\_Planta que más pacientes recibe, ¿qué relación hay entre ella y el resto de plantas y servicios? ¿O los pacientes vienen de la calle y después ya no van a más consultas/técnicas?

Con el mismo corpus que la vista anterior, utilizamos los filtros *Selección de Tramos* y *Disgregación de Flujos* para estudiar las relaciones de C3 con el resto de plantas de CEX.

En la figura 28 encontramos el flujo de pacientes tras pasar el corpus el filtro *Selección de Tramos*. Concretamente hemos seleccionado el nodo C3 para que muestre todos los tramos que tienen como nodo origen esta Torre\_Planta, y represente también todos los nodos anteriores y posteriores, es decir, se visualizará el flujo completo de los **pacientes que pasen en algún momento por C3** (362 pacientes).

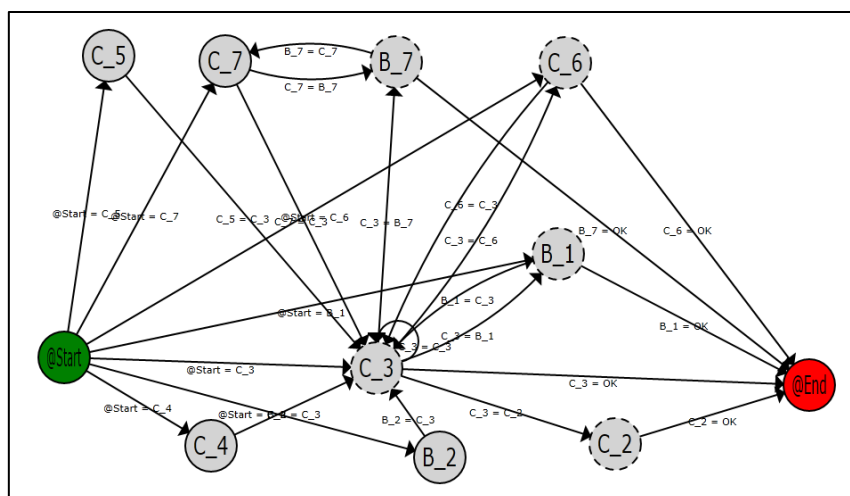


Figura 28: Relación de C3 con el resto de nodos.

Para ver de forma más clara los recorridos de los pacientes que han pasado por C3, se utiliza, junto con el filtro *Selección de Tramos*, el filtro **Disgregación de Flujos** con 0% de similitud para que cada grupo inferido represente un único camino posible (agrupe los estrictamente iguales) y una los grupos menores del 1% para que junte los caminos que sigue únicamente una persona (figura 29).

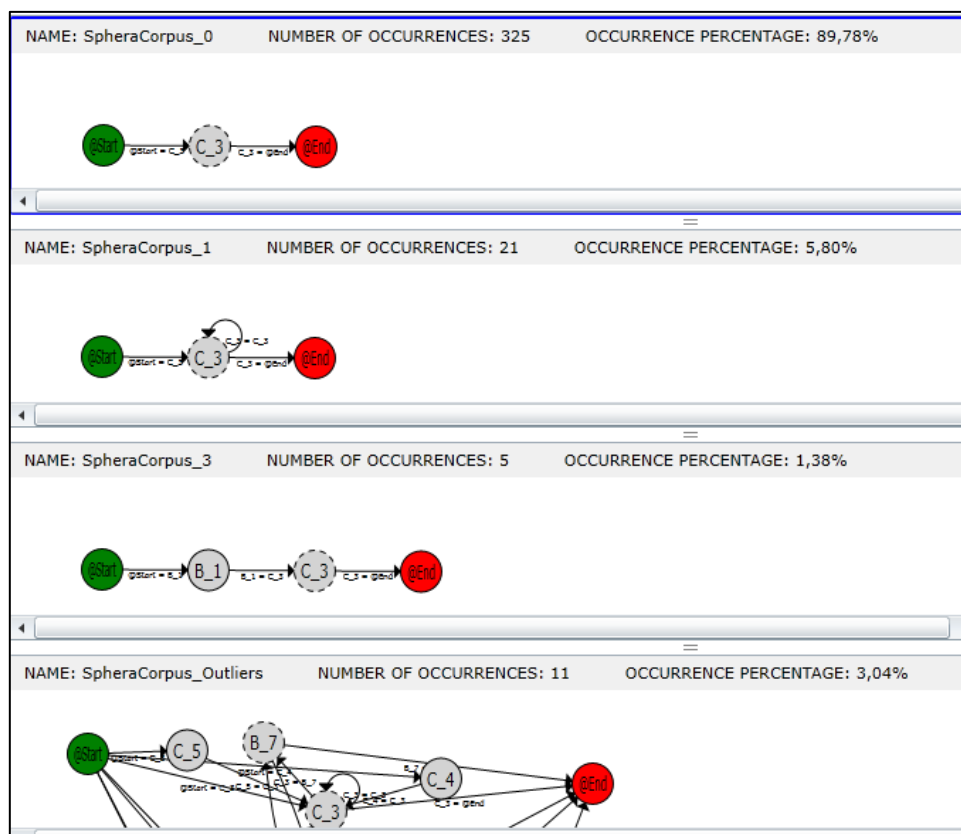


Figura 29. Vista de PALIA con los flujos de los pacientes que pasan por C3.

Como ya muestra PALIA en la parte superior de cada grupo, se obtiene que:

- El **89.78%** de los pacientes que pasan por C3, únicamente acuden a esa cita en el día y se van. Antes y después de C3 no hay ningún local más y en C3 solo tienen una cita.
- El **5.80%** de los pacientes que pasan por C3, únicamente acuden a esa Torre\_Planta, pero lo hacen varias veces, es decir, van a dos o tres citas, ya sean programadas o urgentes, en la misma planta de la misma torre.
- El **1.38%** de los pacientes que pasan por C3, primero acudieron a la planta primera de la torre B (B1) y luego pasaron a un local de C3, una sola vez, para después marcharse. Esta relación también la vimos reflejada en el estudio de las transiciones en la *vista 1*. Para obtener más información de esta transición **B1 → C3**, repetiremos el mismo proceso que con la transición **C7 → B0**, en la figura 30. Los gestores del hospital podrán seguir con este análisis y profundizar en él, si lo quisieran. La herramienta les ha mostrado por dónde empezar a mirar.
- Por último, se visualiza en un único grupo el conjunto de **caminos** seguidos únicamente por **un paciente**.

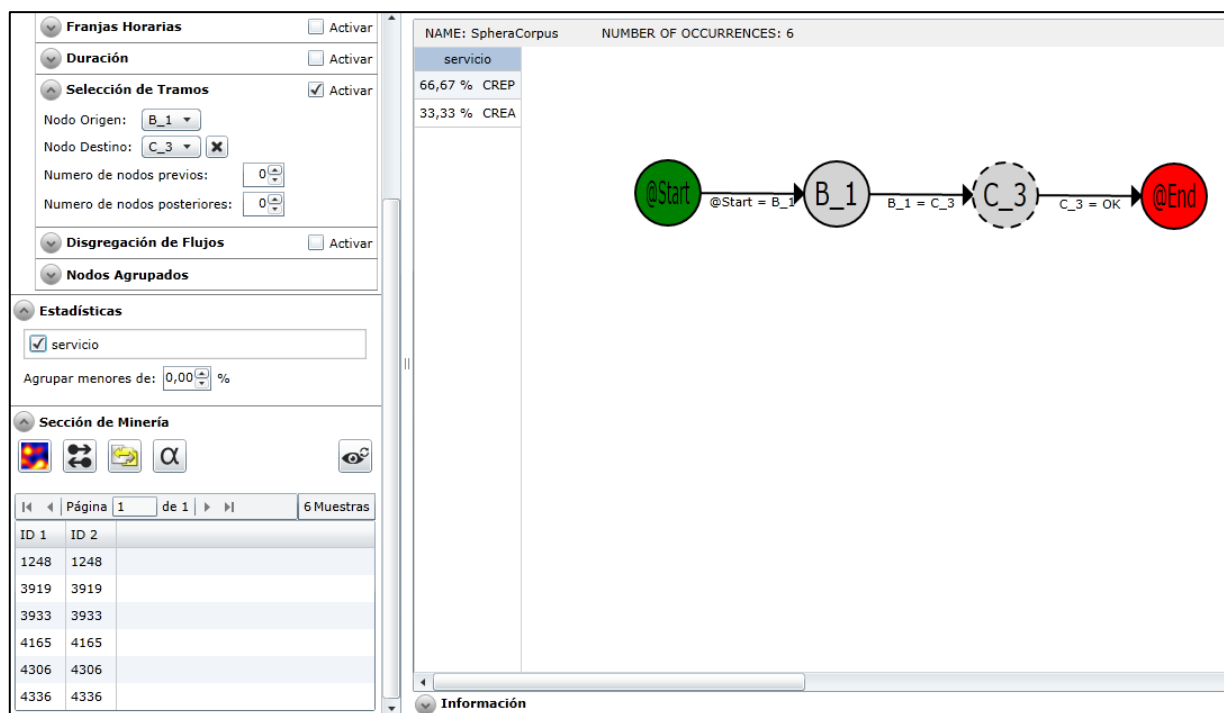


Figura 30. Vista en PALIA del recorrido B1 → C3.

### Los tiempos de retraso en cada servicio

Para estudiar los tiempos de retraso en los servicios de las torres B y C, primero se han de **seleccionar las citas que queremos que estén dentro del análisis**, es decir, en este caso no interesan los pacientes que acuden de forma urgente porque no tenían cita programada y se les ha atendido de forma imprevista. Su hora de inicio programada simplemente es para que tengan una cita registrada y les puedan realizar la prestación. Por lo tanto, el tiempo de retraso de este tipo de citas no es indicativo del tiempo que se retrasa la asistencia en un servicio. No obstante, el conjunto de citas urgentes que tiene registradas las cuatro horas principales, y las que tienen todas menos la de llamada, sí que se incluirán porque son pacientes que se han citado al llegar y tienen una hora de inicio programada real. Por otra parte, los pacientes “No presentados” y las citas que no tienen ninguna de las cuatro horas registradas también se eliminarán del corpus sin errores. El corpus para este análisis se queda con 3413 citas y 3069 pacientes.

Para este estudio **cada cita será un único evento** que represente el tiempo que se ha retrasado la cita. Definíamos anteriormente:

HoraInicioProgramada → HoraLLAMADA

Tiempo de retraso de la cita. Es el tiempo que espera el paciente a causa del hospital.

Sin embargo, tras analizar la fiabilidad de las cuatro horas principales, se vio que la **hora de entrada** era más robusta y representativa del momento en el que empieza la prestación. La escogeremos en lugar de la hora de llamada, pero sigue indicando lo mismo, cuándo llaman al paciente y este entra en el local para iniciar la consulta/técnica.

Este parámetro les es de interés a los gestores porque es un indicador de la eficiencia del servicio y del estado de satisfacción con el hospital que tienen los pacientes que esperan a ser atendidos. El **tiempo de retraso** se define ahora de la siguiente manera:

#### **HorainicioProgramada → HoraEntrada**

Tiempo de retraso de la cita. Es el tiempo que espera el paciente a causa del hospital.

Para definir el comienzo y el final del intervalo del evento, en vez de elegir HorainicioProgramada como START y HoraEntrada como END, que sería lo más lógico, se escoge la siguiente configuración.

Siendo:

$$\text{Tiempo de retraso (TR)} = \text{HoraEntrada} - \text{HorainicioProgramada}$$

Se describe:

$$\text{START} == \text{HoraEntrada}$$

$$\text{END} == \text{HoraEntrada} + \text{Tiempo de retraso (TR)}$$

De esta forma, la duración del evento/nodo sigue representando el TR, pero la diferencia es que las citas se ordenan en el tiempo de acuerdo a la hora de entrada. Así **se mantiene la secuencia** real de pacientes que entró en cada local y la de las citas de un mismo paciente.

Las citas en las que el paciente haya entrado antes de ser la hora de inicio programada (**HoraEntrada < HorainicioProgramada**), es decir, no se ha retrasado la cita, incluso se ha adelantado, el TR se considerará de solamente un segundo para que no sume al tiempo de retraso total del servicio. Si se hiciera la media del TR de un servicio, el TR de estas citas haría que bajara la media porque ha habido citas que se han adelantado y no tienen retraso. Por otra parte, a las citas puntuales en las que el paciente ha entrado a la hora de inicio programada (**HoraEntrada = HorainicioProgramada**) se les hace lo mismo, TR de un segundo. No suman al retraso total del servicio y disminuyen la media porque no se han retrasado esas citas.

Por último, antes de cargar el corpus en PALIA, **se corrigen los solapes** que se hayan formado entre los eventos de un paciente. La figura 31 muestra en script de Matlab diseñado para crear el START y el END descritos arriba y para colocar los eventos solapados uno detrás de otro, manteniendo la duración (tiempo de retraso) de cada uno.

```
Crea_corrige_TR.m:

[entrada,txt]=xlsread('datos para retraso (matlab).xlsx','Hoja1');
tamano_tabla=size(entrada);
num_citas=tamano_tabla(1);

%Leemos los campos
ID=entrada(:,1);
TR=entrada(:,2); %Tiempo de retraso
h_ini_p=entrada(:,3); %HoraInicioProgramada
h_entr=entrada(:,4); %HoraEntrada
```

```
%Leemos la cabecera
c_ID=strjoin(txt(1,1));
c_TR=strjoin(txt(1,2));
c_h_ini_p=strjoin(txt(1,3));
c_h_entr=strjoin(txt(1,4));

%Constantes
un_seg=0.0000115740740740741;

%Iniciamos
START=[];
END=[];

for i=1:num_citas %cada fila

    START(i)=h_entr(i);
    END(i)=START(i)+TR(i);

    if length(find(ID==ID(i))) > 1 %el paciente tiene más de una cita

        citas_paciente=find(ID==ID(i));

        if i>citas_paciente(1) %no es la primera cita del paciente

            for j=1:length(citas_paciente)-1 %para comparar todas las
citas

                actual=[START(i),END(i)];
                anterior=[START(i-j),END(i-j)];

                if actual(2)>=anterior(1) && actual(1)<=anterior(2)

%solape

                    START(i)=END(i-1)+un_seg;
                    END(i)=START(i)+TR(i);

                end
            end
        end
    end

end

cabecera_salida={c_ID,c_h_ini_p,c_h_entr,c_TR,'START','END'};

xlswrite('retrasos_tras_matlab.xlsx',ID,['A2:A' num2str(num_citas+1)])
xlswrite('retrasos_tras_matlab.xlsx',h_ini_p,['B2:B'
num2str(num_citas+1)])
xlswrite('retrasos_tras_matlab.xlsx',h_entr,['C2:C'
num2str(num_citas+1)])
xlswrite('retrasos_tras_matlab.xlsx',TR,['D2:D' num2str(num_citas+1)])
```



```
xlswrite('retrasos_tras_matlab.xlsx', START', ['E2:E'  
num2str(num_citas+1)])  
xlswrite('retrasos_tras_matlab.xlsx', END', ['F2:F'  
num2str(num_citas+1)])  
  
xlswrite('retrasos_tras_matlab.xlsx', cabecera_salida, 'A1:F1')
```

Figura 31. Script de Matlab que prepara los eventos del tiempo de retraso.

El corpus cargado en PALIA tiene la **estructura** de la tabla 10 para poder visualizar los tiempos de retraso en cada servicio de Consultas Externas.

ID	START	END	ACTIVITY
ID nuevo	h_entrada	h_entrada + TR	servicio
1	14/11/2016 12:30:00	14/11/2016 12:30:01	CCAR
2	14/11/2016 9:14:00	14/11/2016 9:43:00	CREA
⋮	⋮	⋮	⋮

Tabla 10. Cabeceras de la vista 4.

#### Vista 4

Al visualizar los registros de eventos en PALIA se obtiene la figura 32 que representa el **retraso que hay en cada servicio** de las CEX del HUPLAFE. Concretamente se muestra en cada nodo el retraso medio por cita que hay en ese servicio, es decir, cuánto retraso tiene de media una cita en ese servicio.

En **naranja más intenso** se pueden ver los servicios donde este tiempo es mayor y, al pasar el ratón por ellos, se obtienen los valores de la tabla 11.

Servicio	Retraso medio por cita
CNFI (NEFROLOGIA INFANTIL)	2:27:52
CCPI (CIR PLASTICA INFANTIL)	2:33:00
CREI (REUMATOLOGIA INFANTIL)	2:17:11
CONI (ONCOLOGIA INFANTIL)	2:05:28
CQUE (QUEMADOS)	2:07:40
CPIN (PEDIATRIA INFECCIOSA)	2:01:00
⋮	⋮

Tabla 11. Servicios con mayor retraso medio por cita.

Llama la atención que los servicios con mayor tiempo de retraso por cita sean los **infantiles**. Queda de manifiesto que esta área es más complicada de gestionar.

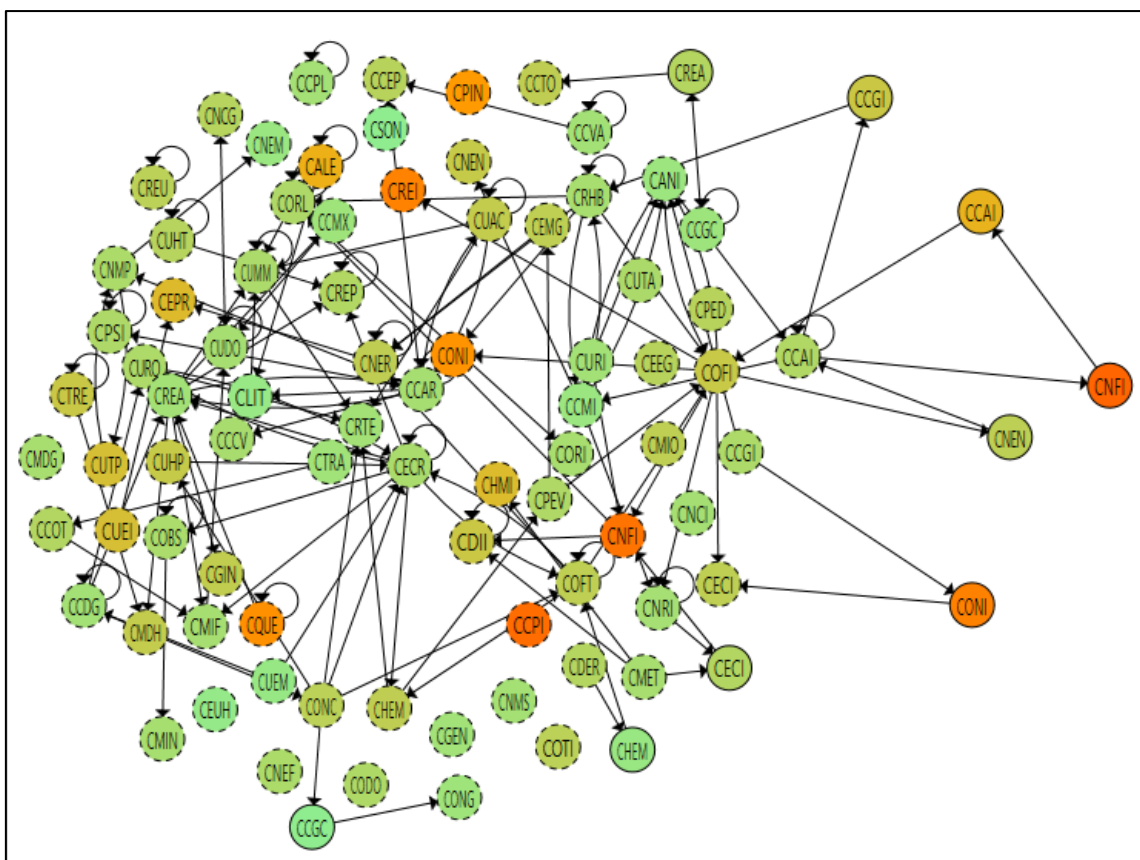


Figura 32. Vista 4 resaltando los nodos naranjas.

PALIA tiene una funcionalidad que permite con unas barras deslizantes cambiar los valores inferiores y superiores del mapa de calor, de forma que se pueden ir **modificando los colores** de los nodos para ver con más detalle la diferencia entre ellos o resaltar algunos por encima de otros. Por ejemplo, en la figura 33 los nodos con mayor retraso de media se han saturado al rojo, de forma que ahora **destacan en verde** los que servicios que menor retraso tienen (Tabla 12).

Servicio	Retraso medio por cita
CSON (SUEÑO)	0:00:01
CEUH (U ENFERMERA ULCERAS Y HERIDAS)	0:06:20
CUEM (CIR ENDOCRINOMETABOLICA)	0:08:00
CLIT (LITOTRICA)	0:08:14
CCMI (CIR MAXILOFACIAL INFANTIL)	0:11:43
CCGC (CONSEJO GENETICO CANCER)	0:14:39
⋮	⋮

Tabla 12. Servicios con menor retraso medio por cita.

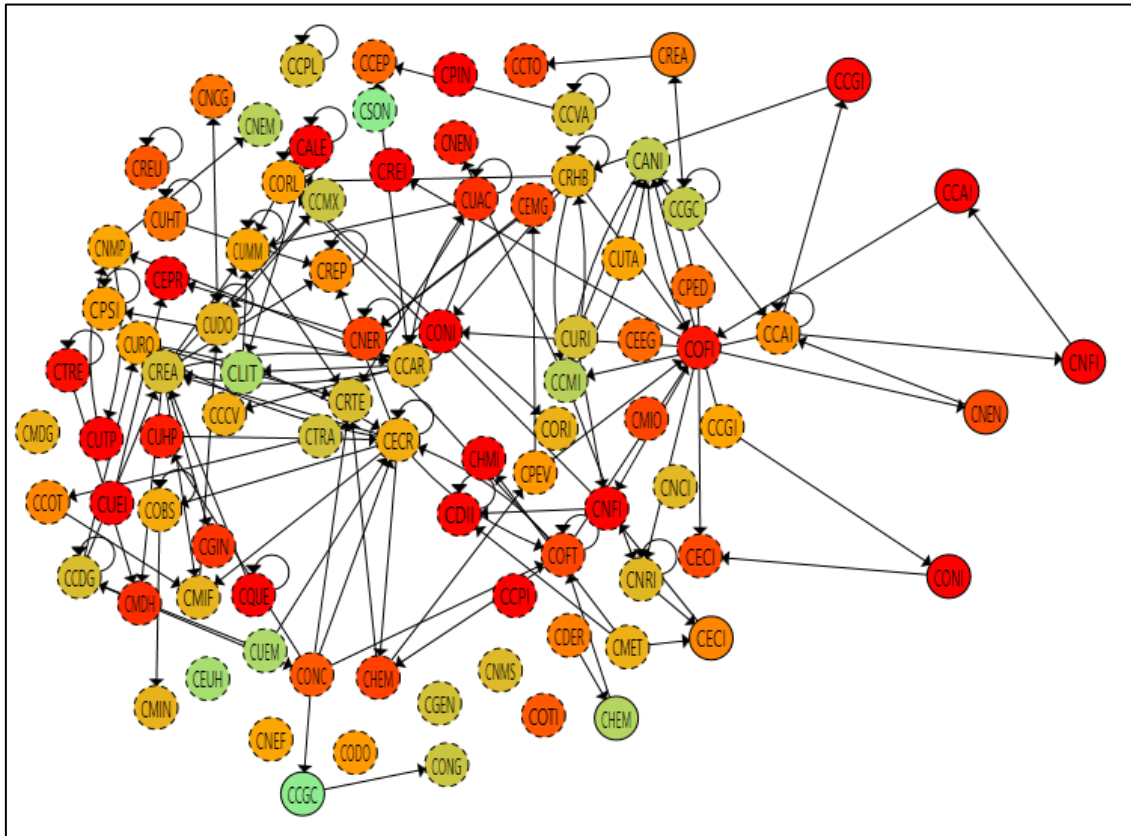


Figura 33. Vista 4 resaltando los nodos verdes.

#### Vista 5

Se visualiza a continuación la **afirmación**: Los pacientes que más tiempo de retraso tienen acumulado son los que tienen varias consultas/técnicas en el día.

Para ello, se agruparán los **pacientes en función del número de citas** que hayan tenido el día de estudio (locales visitados) y los nodos representarán el **tiempo de retraso medio por paciente**. La estructura del corpus es la de la tabla 13.

ID	ACTIVITY	START	END
ID nuevo	num_citas/paciente	h_entrada	h_entrada + TR
1	1 local	14/11/2016 12:30:00	14/11/2016 12:30:01
2	1 local	14/11/2016 9:14:00	14/11/2016 9:43:00
⋮	⋮	⋮	⋮

Tabla 13. Cabeceras de la vista 5.

La vista que se obtiene es la de la figura 34. Antes de analizar los tiempos de retraso medio, pasamos el ratón por los cuatro nodos para saber **cuántos pacientes** hay en cada uno. Los **tiempos de retraso medio por paciente** son los de la tabla 14.

	1 local	2 locales	3 locales	4 locales
Pacientes	2753	291	22	3
TR medio	0:38:22	1:14:23	2:04:39	5:08:42

Tabla 14. Número de pacientes por nodo y tiempos de retraso medio

La razón de obtener el número de pacientes por nodo es que se puede ver que algunos nodos **no tienen suficientes pacientes** para considerar su tiempo de retraso medio significativo. El caso más extremo es el del nodo donde se encuentran los pacientes que acuden a cuatro locales (3 pacientes únicamente). Sin embargo, visualizando los tiempos del resto de nodos, **la afirmación de partida parece cierta**. Los pacientes que acuden a más consultas/técnicas tienen un retraso acumulado mayor, es decir, la suma de todo el tiempo que ha estado esperando porque se ha retrasado la cita es mayor.

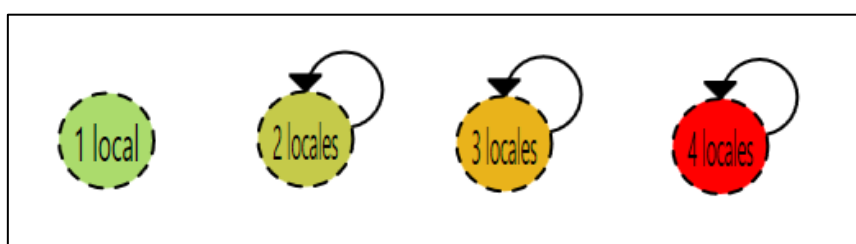


Figura 34. Vista 5

### 5.3 Discusión

Las preguntas iniciales de los gestores del HUPLAFE han sido el centro de todas las fases del trabajo. Entorno a ellas se han elegido los datos y se han visualizado en PALIA para poder responder a las preguntas. Las vistas, originariamente diseñadas para responder a esas preguntas, han dado pie a nuevas cuestiones que antes no se planteaban siquiera, porque al **visualizarse los flujos de pacientes**, han llamado la atención de un vistazo un nodo más rojo, una flecha más verde, reflejando un problema que antes era invisible en la marea de datos.

Así es, las imágenes obtenidas en PALIA les han permitido a los gestores responder a las preguntas y recoger información de cómo están funcionando las Consultas externas en aspectos como los desplazamientos entre locales y los tiempos de retraso de las consultas/técnicas, pero hay una limitación principal, el corpus con el que se ha trabajado **únicamente recoge datos de un día**. De esta forma, a la hora de sacar conclusiones de las vistas mostradas, hay que tener en cuenta este hecho. Ya se manifestó en el estudio de las transiciones de pacientes entre plantas. Eran muy pocos los pacientes que se estaban visualizando en ese caso para declararlos como algo habitual en el hospital.

Al ser este proyecto una prueba de uso de PALIA en el análisis de procesos sanitarios reales, se asumió esta limitación y se desarrollaron todas las fases de la metodología para dar a conocer la herramienta y sus capacidades. Así se ha plasmado en los resultados. Se han analizado las vistas obtenidas para sacar información que pudiera **dar solución a problemas existentes**, como la distribución de ascensores en las plantas, y que mostrara nuevos, como que los cambios de planta más frecuentes eran entre torres. Se ha querido mostrar también que estas primeras sentencias sacadas de las imágenes marcaban el comienzo de un camino por el que empezar un estudio más exhaustivo de la realidad y un foco para la gestión de las CEX.

No obstante, hay una **segunda limitación** que disminuye la fiabilidad de las vistas, la calidad de los datos. A medida que se avanzaba en la caracterización y el estudio de los datos se iba manifestando una heterogeneidad de los mismos muy superior a la que se podía imaginar en un principio. En la fase de análisis de los datos se ha querido plasmar la dificultad a la hora de limpiar los datos y prepararlos para PALIA, eran numerosas las citas anómalas, ya fuera por carecer de datos como por tener una relación entre sus campos incorrecta. También había problemas entre las citas de un mismo local o un mismo paciente, sobre todo de solapamiento, ya que había registros que daban a entender que un paciente estaba esperando a ser atendido y se encontraba a la vez en la consulta de otra cita, por ejemplo.

Los problemas en los datos se atribuyen a la poca fiabilidad de algunos de ellos y, a pesar de que se ha tratado de compensar este y otros aspectos, arreglando o quitando las muestras con errores y ajustando todos los tiempos de las citas, entre otras cosas, todavía puede haber fallos en los datos que hagan que la calidad no sea suficiente para obtener buenos resultados. Por ese motivo, **se sugiere** al equipo informático, que la hora de entrada al local y de salida del mismo, se registre de una forma similar a la de llegada, pero más sencillo. El paciente podría cruzar por la puerta pasando su SIP sin tener que indicar nada y el médico simplemente tendría que asociar la entrada a la cita que tiene como siguiente. A la salida el paciente volvería a hacer lo mismo y se guardaría automáticamente la hora con su entrada en el mismo local.

Dejando la segunda limitación de lado, se presentan a continuación más sugerencias y nuevos planteamientos extraídos de las vistas obtenidas que podrían aplicarse para mejorar la calidad de los procesos asistenciales en el HUPLAFE.

Las vistas 1 y 3 demuestran la utilidad y la posibilidad de visualizar todos los movimientos de los pacientes por Consultas Externas para descubrir aquellos caminos poco transitados y anómalos que pueden ser alguna desviación indicativa de fallo en el proceso asistencial, y los que más pacientes realizan, en los que se podría **profundizar dentro de los niveles de localización** pasando por el servicio y llegando hasta el local concreto.

La vista 2 que mostraba la cantidad de pacientes en cada Torre\_Planta del hospital ha abierto la puerta a la mejora del servicio de ascensores de CEX. Si dispusiéramos de más días en la base de datos se podría obtener esta misma vista, pero para cada día de la semana y de esta forma hacer una **distribución de plantas en los ascensores dinámica**, que pudiera variar en función del día de la semana sabiendo de antemano a qué pisos querrán acudir los pacientes en su mayoría. Del mismo modo, podría cambiar dentro de un mismo día por horas, ya que PALIA permite ver los flujos dentro de un rango de horas determinado.

La vista 4 en la que se puede ver el tiempo de retraso en cada servicio es una vista interesante para llevar un control del funcionamiento de los servicios. Se podría **seguir la evolución** de los servicios, si a medida que pasan los días, semanas o meses van mejorando los tiempos de retraso de la prestación. También podría evaluarse el antes y el después tras haber aplicado un cambio en el servicio y quisieran ver qué efectos ha ocasionado sobre los tiempos de retraso. Un paso más allá sería poder **ver la vista en directo**, es decir, hacer un seguimiento al momento de dónde los pacientes llevan mucho tiempo esperando, para poder tomar medidas en ese momento o conocer cuál es la causa antes de que se pase y ya sea más difícil de averiguar.

## 6. CONCLUSIONES Y TRABAJO FUTURO

En este trabajo ha quedado demostrada la capacidad de las tecnologías de minería de procesos para analizar, mejorar y estandarizar los procesos hospitalarios apoyando a la gestión a **hacer frente a la variabilidad clínica** y favoreciendo la toma de decisiones que redunden en una mejora de la calidad sanitaria, así como en una reducción de sus costes. Como dijimos al principio es el paradigma dominante bajo asedio, por lo que creo que este trabajo abre la puerta a nuevas líneas de investigación que ayuden a mantener la relevancia, eficacia y legitimidad de la sanidad pública.

Se aplicó una de las herramientas de minería de procesos (PALIA) en el Hospital Universitario la Fe de Valencia para probar su uso y su potencial, concretamente sobre el área de Consultas Externas dada la necesidad de sus gestores por **conocer la ejecución real** de estos servicios y mejorar la calidad asistencial de la misma. Para ello se necesitó entender y conocer su funcionamiento, todas las etapas y los actores de cada uno de los procesos que se siguen. Así se hizo, requería descubrir el hospital desde su estructura hasta los sistemas informáticos para comprender los datos que genera y que se iban a analizar.

La **caracterización del corpus** de datos ha sido costosa dada la heterogeneidad de los mismos, pero necesaria para poder aplicar la herramienta de minería de procesos y dar luz a los flujos de pacientes en el proceso asistencial. Se han corregido de diversas formas y se han seleccionado minuciosamente para poder ofrecer a los gestores vistas en PALIA que permitieran de un solo vistazo comprender lo ocurrido en la realidad y crear nuevo conocimiento diferente a lo que percibían en un principio. Siempre teniendo en cuenta la limitación de que solo se disponía de los registros de un día y que la calidad de los datos dejaba mucho que desear.

Para poder entender qué vistas eran más útiles en la gestión de las CEX, se ha trabajado de la mano de los gestores del hospital para hacerles ver las posibilidades que ofrecía la herramienta y las preguntas que podía responder a partir de los datos recogidos. De esta forma se han visualizado diversos flujos de pacientes y características del proceso asistencial que **han dado respuesta a las cuestiones planteadas** y han aportado nuevo conocimiento para continuar el estudio en ese y otros caminos que se han abierto en consecuencia.

Por otro lado, se buscaba **mejorar la aplicación PALIA** y reconocer funcionalidades nuevas que se pudieran incorporar a las ya existentes con el fin de acercar su aplicación a otros procesos sanitarios en el futuro. Ante las preguntas de los expertos y la necesidad de obtener vistas que las respondieran han surgido nuevas formas de representación de los flujos. No obstante, queda pendiente en próximos estudios evaluar el grado de satisfacción con la herramienta que tienen los gestores y así poder incluir toda sugerencia en futuras versiones.

Otro objetivo marcado al principio del proyecto fue favorecer la **comunicación y el intercambio de conocimiento** entre los investigadores del SABIEN-ITACA-UPV y los gestores del HUPLAFE. Durante el desarrollo del proyecto se aplicaron las técnicas y métodos de los investigadores para analizar los datos y obtener las mejores vistas en PALIA y, del mismo modo, la información de los gestores del hospital volvía a la UPV para entender los problemas en los procesos y ofrecer mejoras en la herramienta que facilitarían su visualización.

El último objetivo secundario hacía referencia a saber reconocer **otras herramientas**, a parte de la de minería de procesos, que pudieran ser útiles en alguna etapa del proyecto. Excel ha sido el principal programa para visualizar y manejar los datos, y se ha necesitado Matlab en momentos concretos para transformar los datos y construir los registros de eventos.

## 6.1 Trabajo Futuro

Los resultados obtenidos han sido los propios para poder responder a las preguntas iniciales de los gestores partiendo de una base de datos de únicamente un día, de forma que en otros TFG's o en el TFM se podría responder con mayor certeza a las dudas de los gestores si se dispusiera de datos de **más días** y, por otra parte, de **mejor calidad**. Este segundo punto es clave para obtener información fiable que se pueda traducir en propuestas de mejora para el hospital, por ello, un proyecto futuro en el HUPLAFE tendría como objetivo mejorar la calidad de los datos.

Partiendo de la base de datos mejorada comentada se podría dar respuesta a nuevas preguntas de los gestores, generando vistas diferentes y únicas. Ya no solo dentro de Consultas Externas, sino en **áreas del hospital** dónde PALIA pudiera reflejar la realidad de los procesos detrás de los datos, por ejemplo, en Urgencias. También se podría aplicar la herramienta en **otros hospitales** interesados, y no solo PALIA, **otras técnicas** de minería de procesos como las nombradas en este trabajo, entre ellas Disco, serían opciones diferentes que probar sobre los datos de estos hospitales. En definitiva, poco a poco **expandir el rango de aplicación** de las tecnologías de minería de procesos en la gestión de los procesos sanitarios con el fin de mejorar la calidad asistencial.



## 7. REFERENCIAS

Bose, R., & van der Aalst, W. (2011). Analysis of patient treatment procedures: The BPI Challenge case study. *Bpmcenter.Org*, 26.

Cabo Salvador, J. (2017). 9. Gestión por procesos | Gestion Sanitaria - Grupos Relacionados por el Diagnóstico (GRD). *gestion-sanitaria.com*. Retrieved 10 May 2017, from <http://www.gestion-sanitaria.com/9-gestion-procesos.html>

Dumas, M., Aalst, v., & Hofstede, t. (2005). *Process-aware information systems: bridging people and software through process technology*. Hoboken: Wiley-Interscience.

Fernández-Llatas, C., & García-Gómez, J. (2015). *Data mining in clinical medicine*. Totowa, N.J.: Humana Press.

Fernandez-Llatas, C., Lizondo, A., Monton, E., Benedi, J., & Traver, V. (2015). Process Mining Methodology for Health Process Tracking Using Real-Time Indoor Location Systems. *Sensors*, 15(12), 29821-29840. <http://dx.doi.org/10.3390/s151229769>

Fernandez-Llatas, C., Martinez-Millana, A., Martinez-Romero, A., Benedi, J., & Traver, V. (2015). Diabetes care related process modelling using Process Mining techniques. Lessons learned in the application of Interactive Pattern Recognition: coping with the Spaghetti Effect. *2015 37th Annual International Conference Of The IEEE Engineering In Medicine And Biology Society (EMBC)*. <http://dx.doi.org/10.1109/embc.2015.7318809>

Fernandez-Llatas, C., F. Pileggi, S., Traver, V., & Benedi, J. (2011). Timed Parallel Automaton: A Mathematical Tool for Defining Highly Expressive Formal Workflows. *Fifth Asia Modelling Symposium*.

Grol, R., & Grimshaw, J. (1999). Evidence-based implementation of evidence-based medicine. *Jt Comm J Qual Improv.*, 10, 503-13.

Gunther, C., & Rozinat, A. (2012). Disco: Discover Your Processes. *BPM 2012 Demonstration Track*, 46.

Hernández Nariño, A., Nogueira Rivera, D., Medina León, A., & Marqués León, M. (2012). Inserción de la gestión por procesos en instituciones hospitalarias. Concepción metodológica y práctica. *Rev. Adm. (São Paulo)*.

Kaymak, U., Mans, R., van de Steeg, T., & Dierks, M. (2012). On process mining in health care, in *Proceedings. IEEE International Conference On Systems, Man, And Cybernetics, SMC*, 1859–1864.

*La Fe – Hospital Universitari i Politècnic*. (2017). *Lafe.san.gva.es*. Retrieved 4 March 2017, from <http://www.lafe.san.gva.es/el-hospital-en-cifras>

Mans, R., Schonenberg, M., Song, M., van der Aalst, W., & Bakker, P. (2008). Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital. *Biomedical Engineering Systems And Technologies*, 425-438.

Mans, R., van der Aalst, W., & Vanwersch, R. (2015). *Process Mining in Healthcare*. (1st ed.). Springer International Publishing.

Perimal-Lewis, L., Qin, S., Thompson, C., & Hakendorf, P. (2012). Gaining insight from patient journey data using a process-oriented analysis approach. *Proceedings Of The Fifth Australasian Workshop On Health Informatics And Knowledge Management, Volume 129*(2012), Pages 59-66.

Pino, E., & Lara, M. (2013). *Los Estados De Bienestar En La Encrucijada*. Madrid Tecnos.

Radnor, Z., Holweg, M., & Waring, J. (2012). Lean in healthcare: The unfilled promise? *Social Science & Medicine*, 74(3), 364-371. <http://dx.doi.org/10.1016/j.socscimed.2011.02.011>

Rebuge, Á., & Ferreira, D. (2012). Business process analysis in healthcare environments: A methodology based on process mining. *Information Systems*, 37(2), 99-116. <http://dx.doi.org/10.1016/j.is.2011.01.003>

Rojas, E., Muñoz-Gama, J., Sepúlveda, M., & Capurro, D. (2016). Process mining in healthcare: A literature review. *Journal Of Biomedical Informatics*, 61, 224-236. <http://dx.doi.org/10.1016/j.jbi.2016.04.007>

Sharma, G., Boeckmann, J., & Wong, B. (2016). Evidence-Based Medicine. *Facial Plastic And Reconstructive Surgery*, 39-46

*Sistemas de ayuda a la decisión basados en datos biomédicos para clínica asistencial y salud pública*. (2017). Apuntes de la asignatura Sistemas de la Información y la Comunicación II.

van der Aalst, W. (2012). Process Mining: Overview and Opportunities. *ACM Transactions On Management Information Systems*, 3(2), 1-17. <http://dx.doi.org/10.1145/2229156.2229157>

van der Aalst, W. (2014). *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Berlin: Springer Berlin.

van der Aalst, W., van Dongen, B., Herbst, J., Maruster, L., Schimm, G., & Weijters, A. (2003). Workflow mining: A survey of issues and approaches. *Data & Knowledge Engineering*, 47(2), 237-267. [http://dx.doi.org/10.1016/s0169-023x\(03\)00066-1](http://dx.doi.org/10.1016/s0169-023x(03)00066-1)

van der Aalst, W., & Weijters, A. (2004). Process mining: a research agenda. *Computers In Industry*, 53(3), 231-244. <http://dx.doi.org/10.1016/j.compind.2003.10.001>

van Dongen, B., de Medeiros, A., Verbeek, M., Weijters, A., & van der Aalst, W. (2005). *The ProM Framework: A New Era in Process Mining Tool Support*.

Zhou, J. (2009). *Process mining: acquiring objective process information for healthcare process management with the CRISP-DM framework*.

# PRESUPUESTO

APLICACIÓN EN UN HOSPITAL DEL ANÁLISIS DE PROCESOS SANITARIOS  
HACIENDO USO DE TECNOLOGÍAS DE MINERÍA DE PROCESOS

Marta Mares García



## Índice del Presupuesto

1. Cuadro de precios de mano de obra .....	4
2. Cuadro de precios de maquinaria y software .....	4
3. Cuadro de precios unitarios .....	4
4. Cuadro de precios descompuestos .....	5
5. Presupuestos parciales.....	7
6. Presupuesto total de ejecución por contrata .....	8



A la hora de exponer el presupuesto de este proyecto se ha optado por considerar cuatro unidades de obra:

1. Definición del proyecto y obtención de los datos
2. Análisis del área de Consultas Externas
3. Preparación de los datos y estudio de la herramienta
4. Carga de datos en la herramienta y obtención de vistas del proceso

Las cuatro representan las fases para realizar una prueba de uso de la herramienta de minería de procesos en un entorno sanitario. Consta de un primer acercamiento con el Hospital Universitario y Politécnico de la Fe para identificar los problemas existentes en la gestión de procesos, concretamente Consultas Externas, y aplicación de la herramienta de minería de procesos del grupo SABIEN-ITACA-UPV una vez analizados los datos proporcionados por el hospital.

A continuación, se detallan las unidades de obra junto con la mano de obra, los equipo y el software necesario para realizar el proyecto. Hay que tener en cuenta que los costes por hora de la mano de obra incluyen las tasas y la seguridad social.

<b>CUADRO DE PRECIOS DE MANO DE OBRA</b>						
Nº recurso	Código	Descripción del recurso	Unidad	Precio unitario (€)	Cantidad	Total (€)
1	MO01	Investigador (Tutor)	€/hora	35	40	1400
2	MO02	Investigador (Colaborador)	€/hora	35	40	1400
3	MO03	Gestor del hospital (Colaborador)	€/hora	30	30	900
4	MO04	Ingeniero biomédico (Alumno)	€/hora	25	305	7625
					<b>Total</b>	<b>11325</b>

<b>CUADRO DE PRECIOS DE MAQUINARIA Y SOFTWARE</b>						
Nº recurso	Código	Descripción del recurso	Unidad	Precio unitario (€)	Cantidad	Total (€)
1	MAQ01	Ordenador portátil	€	500	4	2000
2	MAQ02	Microsoft Office 365	€/año	69	4	276
3	MAQ03	Mathworks Matlab	€/año	2000	2	4000
4	MAQ04	PALIA ILS Suite Web Tool	€/año	4000	2	8000
					<b>Total</b>	<b>14276</b>

<b>CUADRO DE PRECIOS UNITARIOS</b>			
Nº Actividad	Código	Descripción de las unidades de obra	Precio (€)
1	UO1	Definición del proyecto y obtención de los datos: identificar los objetivos del proyecto, determinar el área del hospital a analizar y conseguir los datos del mismo. DOS MIL OCHOCIENTOS SESENTA	2860
2	UO2	Análisis del área de Consultas Externas: Conocer esta área hospitalaria, todos los datos registrados y la calidad de los mismos. TRES MIL VEINTICINCO	3025



3	UO3	Preparación de los datos y estudio de la herramienta: Preprocesado de los datos para poder introducirlos en la herramienta y el estudio de ésta para aprovechar todas sus funcionalidades. SIETE MIL SETECIENTOS SESENTA	7760
4	UO4	Carga de datos en la herramienta y obtención de vistas del proceso: Manejo de la herramienta para obtener vistas del proceso que puedan responder a las preguntas de los gestores. DOCE MIL TRENTA Y UNO	12031
<b>Total</b>			<b>25676,00</b>

<b>CUADRO DE PRECIOS DESCOMPUESTOS</b>						
<b>1.</b>	<b>Definición del proyecto y obtención de los datos</b>					
<b>Nº</b>	<b>Capítulo</b>					
1.1	Reunión inicial de todos los participantes					
	<b>Código recurso</b>	<b>Descripción del recurso</b>	<b>Precio unitario (€)</b>	<b>Unidad</b>	<b>Cantidad</b>	<b>Precio (€)</b>
	MO01	Investigador (Tutor)	35	h	4	140
	MO02	Investigador (Colaborador)	35	h	4	140
	MO03	Gestor del hospital (Colaborador)	30	h	4	120
	MO04	Ingeniero biomédico (Alumno)	25	h	4	100
	MAQ01	Ordenador portátil	500	h	4	2000
					<b>Subtotal</b>	<b>2500</b>
1.2	Traspaso de datos entre el hospital y el grupo investigador					
	MO01	Investigador (Tutor)	35	h	4	140
	MO03	Gestor del hospital (Colaborador)	30	h	4	120
	MO04	Ingeniero biomédico (Alumno)	25	h	4	100
					<b>Subtotal</b>	<b>360</b>
					<b>Total</b>	<b>2860</b>
<b>2.</b>	<b>Análisis del área de Consultas Externas</b>					
2.1	Reuniones del gestor y el alumno					

	MO03	Gestor del hospital (Colaborador)	30	h	3	90
	MO04	Ingeniero biomédico (Alumno)	25	h	3	75
					<b>Subtotal</b>	<b>165</b>
2.2	Caracterización del corpus de datos					
	MO02	Investigador (Colaborador)	35	h	14	490
	MO03	Gestor del hospital (Colaborador)	30	h	4	120
	MO04	Ingeniero biomédico (Alumno)	25	h	90	2250
					<b>Subtotal</b>	<b>2860</b>
					<b>Total</b>	<b>3025</b>
<b>3.</b>	<b>Preparación de los datos y estudio de la herramienta</b>					
3.1	Preprocesado de datos					
	MO02	Investigador (Colaborador)	35	h	10	350
	MO04	Ingeniero biomédico (Alumno)	25	h	90	2250
	MAQ03	Mathworks Matlab	2000	h	2	4000
					<b>Subtotal</b>	<b>6600</b>
3.2	Estudio de las funcionalidades de la herramienta					
	MO02	Investigador (Colaborador)	35	h	6	210
	MO04	Ingeniero biomédico (Alumno)	25	h	38	950
					<b>Subtotal</b>	<b>1160</b>
					<b>Total</b>	<b>7760</b>
<b>4.</b>	<b>Carga de datos en la herramienta y obtención de vistas del proceso</b>					
4.1	Reuniones del gestor y el alumno					
	MO03	Gestor del hospital (Colaborador)	30	h	3	90
	MO04	Ingeniero biomédico (Alumno)	25	h	3	75
					<b>Subtotal</b>	<b>165</b>
4.2	Manejo de la herramienta para obtener vistas deseadas					
	MO01	Investigador (Tutor)	35	h	26	910

	MO03	Gestor del hospital (Colaborador)	30	h	6	180
	MO04	Ingeniero biomédico (Alumno)	25	h	70	1750
	MAQ04	PALIA ILS Suite Web Tool	4000	h	2	8000
					<b>Subtotal</b>	<b>10840</b>
4.3	Reunión de todos los participantes					
	MO01	Investigador (Tutor)	35	h	6	210
	MO02	Investigador (Colaborador)	35	h	6	210
	MO03	Gestor del hospital (Colaborador)	30	h	6	180
	MO04	Ingeniero biomédico (Alumno)	25	h	6	150
	MAQ02	Microsoft Office 365	69	h	4	276
					<b>Subtotal</b>	<b>1026</b>
					<b>Total</b>	<b>12031</b>
			<b>TOTAL PRECIOS DESCOMPUESTOS</b>			<b>25676,00</b>

<b>CUADRO DE PRESUPUESTOS PARCIALES</b>					
<b>1.</b>	<b>Definición del proyecto y obtención de los datos</b>				
Nº	Unidad	Capítulo	Cantidad	Precio (€)	Importe (€)
1.1	h	Reunión inicial de todos los participantes	20	125	2500
1.2	h	Traspaso de datos entre el hospital y el grupo investigador	12	30	360
				<b>Subtotal</b>	<b>2860</b>
<b>2.</b>	<b>Análisis del área de Consultas Externas</b>				
2.1	h	Reuniones del gestor y el alumno	6	27,50	165
2.2	h	Caracterización del corpus de datos	108	26,48	2860
				<b>Subtotal</b>	<b>3025</b>
<b>3.</b>	<b>Preparación de los datos y estudio de la herramienta</b>				

3.1	h	Preprocesado de datos	102	64,71	6600
3.2	h	Estudio de las funcionalidades de la herramienta	44	26,36	1160
				<b>Subtotal</b>	<b>7760</b>
<b>4.</b>	<b>Carga de datos en la herramienta y obtención de vistas del proceso</b>				
4.1	h	Reuniones del gestor y el alumno	6	27,5	165
4.2	h	Manejo de la herramienta para obtener vistas deseadas	104	104,23	10840
4.3	h	Reunión de todos los participantes	28	36,64	1026
				<b>Subtotal</b>	<b>12031</b>
		<b>TOTAL PRESUPUESTOS PARCIALES</b>			<b>25676,00</b>

<b>PRESUPUESTO TOTAL DE EJECUCIÓN POR CONTRATA</b>		
<b>Código</b>	<b>Descripción de las unidades de obra</b>	<b>Importe (€)</b>
U01	Definición del proyecto y obtención de los datos	2860
U02	Análisis del área de Consultas Externas	3025
U03	Preparación de los datos y estudio de la herramienta	7760
U04	Carga de datos en la herramienta y obtención de vistas del proceso	12031
Presupuesto de Ejecución Material		25676,00
Gastos generales (20%)		5135,20
Beneficio industrial (6%)		1540,56
Suma		32351,76
IVA (21%)		6793,87
<b>Presupuesto de Ejecución por Contrata</b>		<b>39145,63</b>