



Herramienta HAID - IMPULSA

Proceso de Compra Pública
de Innovación (CPI)

Fase de Consulta Preliminar al Mercado
Documento compendio de aportaciones

Noviembre de 2017

Índice de Contenidos

Compra Pública de Innovación.....	3
El proceso de Consulta Preliminar al Mercado	3
Resultados de la Consulta	4
Estructura del análisis técnico	4
Cuestiones de Partida.....	6
Grado de Madurez Tecnológica	8
1. Aplicación concreta de desarrollos ad-hoc previos.....	8
2. Tareas de desarrollo ad-hoc y porcentajes.....	8
3. Conclusiones sobre grado de madurez tecnológica.....	9
Propuestas en el ámbito técnico	11
1. Aplicación concreta de productos / componentes de software libre.....	11
2. Aplicación concreta de productos / componentes de mercado.....	11
3. Propuestas para Arquitectura Big Data	11
4. Modelo (s) tecnológico (s) para el almacenamiento, acceso y distribución de información y datos.....	12
5. Modelo (s) tecnológico (s) para el análisis de datos, extracción de conocimiento y su presentación (visualización) a cada tipo de usuario final...	12
6. Herramientas visuales o productos con los que cuentan las compañías	13
7. Solución basada en la nube	13
8. Respuesta a aspectos Reglamento (UE) 2016/679	14
1. Metodologías de gestión y desarrollo del proyecto	15
2. Recursos humanos y materiales	16
Adecuación del presupuesto	17
1. Propuestas de mantenimiento de la herramienta y su coste	17
2. Adecuación del presupuesto previsto para satisfacer las necesidades de la herramienta	17

Compra Pública de Innovación

La contratación pública es un instrumento clave para la consecución de las tres prioridades de la Estrategia Europa 2020: Crecimiento inteligente, sostenible e integrador.

La Compra Pública de Innovación es una actuación administrativa a través de la cual se pretende satisfacer una necesidad no cubierta a partir de una solución innovadora. Tiene cuatro objetivos principalmente:

- Mejorar los servicios públicos mediante la adquisición de bienes y servicios innovadores.
- Impulsar la innovación desde el lado de la demanda.
- Estimular la Innovación en el tejido empresarial.
- Fomentar la internalización de la innovación a partir de la utilización del mercado público como plataforma de lanzamiento de bienes y servicios.

Desde "GIJÓN IMPULSA" se ha puesto en marcha una primera experiencia piloto en materia de Compra Pública de Innovación (CPI), para la adquisición de una "Herramienta de Análisis Inteligente de Datos" (HAID IMPULSA). Esta Herramienta de Análisis Inteligente de Datos deberá permitir hacer accesible información proveniente de diversas fuentes de datos en abierto para la ciudadanía y las empresas, con el fin de localizar las fortalezas de la ciudad, superar las debilidades y a partir de los datos obtenidos, las empresas puedan desarrollar su actividad de una forma más eficiente. De este modo, el conjunto de la ciudad dispondrá de una Herramienta que fortalecerá aquellos sectores que proporcionen mayor valor añadido y, por tanto, mejorarán las expectativas de crecimiento económico de la ciudad.

El proceso de Consulta Preliminar al Mercado

El lanzamiento de un procedimiento de Compra Pública de Innovación mediante el que se pretende implementar soluciones o procedimientos innovadores, requiere un cierto conocimiento del alcance de las soluciones que es posible plantear. La Directiva 2014/24/UE del Parlamento Europeo y del Consejo, de 26 de febrero de 2014, sobre contratación pública, reconoce de manera explícita en su artículo 40, la figura de la Consulta Preliminar al Mercado. Ésta permite a los poderes adjudicadores la realización de "consultas al mercado", con vistas a preparar la contratación e informar a los operadores económicos acerca de sus planes y de los requisitos de contratación, con anterioridad al inicio del procedimiento. Esta actuación aparece recogida de forma más concreta en la ley 9/2017 de 8 de noviembre en el artículo 115 (en vigor a partir del 9 de marzo de 2018).

Siguiendo el artículo 40 de la Directiva antes mencionada y, bajo las pautas de la iniciativa EAFIP de la Comisión Europea, se realiza la consulta preliminar al mercado con el fin de dar cumplimiento a las fases del proceso de una compra pública de innovación. Se elabora la documentación específica para solicitar al mercado el estado del mismo en la materia objeto de la consulta que no es otra que la existencia o no de una herramienta que reúna los requisitos y características especificados en la documentación que IMPULSA pone a disposición del mercado. Se trata de contar, a partir de éste, con la retroalimentación de actores interesados que permita verificar la

viabilidad, el alcance y características tecnológicas apropiadas para la realización e implementación de una Herramienta de Análisis Inteligente de Datos para "IMPULSA".

La consulta se realiza a partir de la publicación en la página web de IMPULSA de la documentación elaborada para la compra de la herramienta y se solicita al mercado que se cumplimente un cuestionario donde, a partir de las características publicadas sobre la herramienta, actores del mercado puedan contestar sobre la existencia o no de dicha herramienta y las posibilidades de llevarla a cabo.

La consulta al mercado, con un plazo de duración de dos meses, se realiza a partir de la publicación de la misma en el Suplemento de Diario Oficial de la Unión Europea (Ted - diario electrónico), Boletín Oficial del Estado (BOE) y en el Boletín Oficial del Principado de Asturias (BOPA) en el mes de mayo y la realización de una jornada de información pública sobre dicha consulta el día 20 de junio retransmitida en streaming con el fin de dar la difusión más amplia posible a dicha consulta.

Resultados de la Consulta

Tras el periodo abierto para la recepción de los cuestionarios cumplimentados sobre la consulta (desde el 13 de mayo al 14 de julio), se recibieron un total de ocho cuestionarios, correspondientes a las siguientes firmas:

- Respuesta conjunta de las empresas: Neosystems, S.L., Flame Analytics, S.L., Contein XXI, S.L., Sigea, Sistemas de protección de la Información, S.L.
- DELOITTE CONSULTING S.L.U.
- EVERIS SPAIN S.L.U.
- INFORMÁTICA DEL CORTE INGLÉS, S.A. (IECISA)
- PROYECTOS EN GESTION DEL CONOCIMIENTO, S.L.
- SUMINISTROS, IMPORTACIONES Y MANTENIMIENTOS ELECTRONICOS, S.A. (SERMICRO)
- TREELOGIC TELEMÁTICA Y LÓGICA RACIONAL PARA LA EMPRESA EUROPEA S.L.
- VITESIA MOBILE SOLUTIONS S.L.L.

Estructura del análisis técnico

En el presente documento se ha estructurado el análisis de las respuestas recibidas de manera que permita al lector tener una visión conjunta e integral de la información obtenida tras el periodo de consultas abierto en esta fase, al que han concurrido una serie de entidades que han manifestado su interés y aportado su visión sobre las cuestiones que se les han planteado.

Por tanto, se ha procurado sintetizar las aportaciones de cada organización, organizando los ítems en los grandes bloques objeto de la consulta al mercado para el desarrollo de la herramienta HAID IMPULSA.

Cuestiones de Partida

En el formulario de consulta distribuido, se plantean una serie de cuestiones de partida que suponen el núcleo de la información de la propuesta, puesto que hacen referencia a tres aspectos fundamentales:

- 1) ¿Dispone de una herramienta que pueda satisfacer las necesidades de esta consulta?**
- 2) En caso de que la respuesta sea afirmativa por favor haga una breve descripción de su herramienta y conteste las cuestiones de los puntos 3, 4,5 y 6 donde corresponda.**
- 3) En caso de que la respuesta sea negativa:**
 - a. Para la obtención de esta herramienta sería necesario un desarrollo de tecnología nueva o mejorada.**
 - b. Para la obtención de esta herramienta sería necesario un nuevo proceso de I+D**

Se expone a continuación un resumen sobre las aportaciones de las empresas a estas primeras cuestiones.

	1) ¿Dispone de una herramienta que pueda satisfacer las necesidades de esta consulta?	3) En caso de que la respuesta sea negativa:	
		a. Para la obtención de esta herramienta sería necesario un desarrollo de tecnología nueva o mejorada.	b. Para la obtención de esta herramienta sería necesario un nuevo proceso de I+D
SI	4 empresas	4 empresas	1 empresa
NO	4 empresas		4 empresas

De las ocho empresas, **cuatro** afirman **disponer** de una herramienta y **otras cuatro** por el contrario **no disponen** de dicha herramienta. Visto este escenario, podría considerarse de partida que el mercado dispone de herramientas comerciales que podrían asumir el alcance de la necesidad planteada o bien que para su obtención no sería necesario un proceso de innovación, pero nada más lejos de la realidad.

Esto es así porque las cuatro empresas que afirman disponer de la herramienta en cuestión, evidencian a lo largo del resto del formulario, que requieren de un grado de desarrollo (entendiendo por desarrollo la ejecución de trabajos de personalización de herramientas de las que ya disponen) que supera ampliamente el presupuesto disponible, incluso una de esas soluciones es la más costosa de las cuantificadas. El hecho de que una personalización requiera más esfuerzo que todo un desarrollo completo de una herramienta, evidencia que realmente su afirmación inicial no es del todo correcta.

Por otro lado, todas las empresas que afirman no disponer de una herramienta con las funcionalidades y alcance requerido, coinciden en que no sería necesario un proceso de I+D para su obtención, tan sólo una fase de desarrollo, salvo en el caso de Gestión del Conocimiento.

Una de las razones que podrían argumentarse para que las empresas considerasen que el proceso no requeriría un proceso de Innovación y Desarrollo (I+D) es que las empresas que no disponen de la herramienta no consideren que la creación de la misma se enmarque dentro de un proceso "puro" de I+D+i. Esta cuestión ya había sido aclarada en el proceso abierto de Consulta Preliminar, mediante la publicación de una serie de Preguntas Frecuentes (fecha de 29 de junio de 2017), que se transcribe a continuación:

Con tecnología NUEVA o MEJORADA, el cuestionario se refiere a si la tecnología empleada no se ha empleado NUNCA (por tanto sería NUEVA) en este ámbito o para este fin o bien, se trata de una tecnología que ya ha sido empleada en este campo pero requiere de una mejora, actualización, revisión u otro proceso para poder desarrollar las funcionalidades que se le requieren.

En el caso de optar por la opción del proceso de I+D, el cuestionario se refiere a la necesidad de que se realice un nuevo desarrollo completo de la tecnología desde un estadio embrionario, es decir, con un TRL bajo o muy bajo.

Por tanto, se deduce del formulario que para llegar a desarrollar una herramienta con las funcionalidades requeridas y que permita obtener el alcance propuesto, las empresas requieren de un proceso que se enmarca dentro del campo de la INNOVACIÓN (TRLs medios-altos) y no dentro de la INVESTIGACIÓN o el DESARROLLO BÁSICO.

En concreto, tal y como se observa a lo largo de la información aportada por las empresas, que la construcción de la herramienta aún para aquellas que parten de un menor grado avance, requeriría de un proceso de adición de "piezas tecnológicas" ya disponibles que serían conveniente personalizadas y adaptadas tanto a las fuentes de información como a su tratamiento y a la visualización/presentación de los resultados.

Grado de Madurez Tecnológica

Para conocer el grado de madurez tecnológica del sector en el desarrollo de este tipo de herramientas, se ha intentado conocer el porcentaje de desarrollo específico que habrían de desplegar las empresas o si bien disponen de avances o experiencias previas que les permitan disminuir la carga en esta fase y centrarse en la personalización de dichos desarrollos previos.

1. Aplicación concreta de desarrollos ad-hoc previos

El término “ad hoc” en este contexto, se refiere a una solución o pieza tecnológica que había sido específicamente elaborada con anterioridad para dar solución a un problema o fin preciso, pero que podría reutilizarse en el marco del presente proyecto.

Dos compañías manifiestan que disponen de desarrollos ad-hoc de otros proyectos, los cuales podrían ser utilizados de forma directa en la herramienta HAID IMPULSA.

En concreto, se trata de desarrollos para la captura de datos desde dispositivos IoT a través de dataloggers propios, el análisis semántico de los comentarios en redes sociales o foros a través de APIs específicamente diseñadas para ello o herramientas de extracción de información sobre formatos PDF.

El resto de compañías no hacen mención a desarrollos ad-hoc en modelos anteriores, por lo que se limitan a enumerar los posibles desarrollos necesarios en el presente proyecto que se concretan en el apartado siguiente.

2. Tareas de desarrollo ad-hoc y porcentajes.

El desarrollo ad hoc en el marco del proyecto al que se refiere este apartado, se entienden tanto a procesos de innovación o investigación, como a todas aquellas tareas específicas presentes en el diseño y ejecución de cualquier herramienta informática, como los modelos de datos, sistemas de almacenamiento, gestión de la información, visualización, etc.

En este apartado se trata de comprender los esfuerzos y recursos necesarios que deberán desplegar las empresas para dar cumplimientos a los requisitos y funcionalidades de la herramienta HAID IMPULSA, intentando escindir los propios de un proceso de investigación, desarrollo e innovación (I+D+i) de aquellos que son comunes a la implantación de una aplicación.

En general, el porcentaje de desarrollo que manifiestan las empresas, no parece corresponderse con las funcionalidades que ofrecen las tecnologías Big Data propuestas, la ingesta y procesamiento de los datos es una tarea más orientada a la integración de diversas fuentes de datos que al desarrollo de software propiamente dicho. Además, la mayoría de tecnologías propuestas ya ofrecen herramientas funcionales basadas en configuración y con un amplio catálogo de

conectores, lo que permite conectar automáticamente gran variedad de fuentes de datos.

La mayoría de compañías sitúan el desarrollo ad-hoc en torno al 60%-70%, con dos excepciones, que lo sitúan en un 30% y un 80% respectivamente y otra que no lo especifica, pues considera que la herramienta a integrar es una herramienta visual que no precisa de desarrollo (no obstante, aunque no lo contabiliza como desarrollo, plantea la necesidad de definir los procesos de integración y puesta en marcha de la herramienta).

Muchas de las compañías destacan que el porcentaje de desarrollo ad-hoc depende del alcance del mismo por lo que, a pesar de dar un porcentaje concreto, no justifican el porqué de dicho porcentaje de trabajos de personalización.

3. Conclusiones sobre grado de madurez tecnológica

En función de las respuestas de estos dos apartados, se pueden obtener las siguientes conclusiones.

- Tan sólo dos empresas exponen abiertamente que disponen de desarrollos anteriores de otros proyectos y que podrían reutilizarse para la herramienta HAID IMPULSA.
- El resto de empresas, incluso aquellas que en el primer bloque de preguntas afirmaban disponer de la herramienta necesaria, enumeran una serie de posibles desarrollos específicos que deberían llevarse a cabo para poner en funcionamiento la herramienta.

Por tanto, se puede deducir que independientemente de si se dispone de una solución previa o de piezas tecnológicas concretas, será necesario un desarrollo importante de trabajos concretos localizados en segmentos concretos.

En este sentido, resulta complejo ubicar los segmentos en los que mayoritariamente se encuentran los procesos de innovación, pero de las respuestas aportadas, se deduce que serían los siguientes:

- **Recolección de información.**
 - o En el proceso de recolección de información se utilizan conectores, entendiéndose como tales los módulos software que recogen, pre-procesan y normalizan la información de un tipo concreto de fuente de eventos, estructurada o no.
- **Almacenamiento de la información.**
 - o Almacenamiento, búsqueda y procesamiento eficiente de grandes volúmenes de información; utilizando técnicas de indexación que permitan realizar investigaciones, análisis y estudios en profundidad sobre los datos almacenados
- **Procesamiento de la información.**
 - o Incluye los subprocesos de ingestión de los datos (en función de las fuentes de datos), el procesamiento de los mismos (en función

del tipo de tratamiento de información que se desee realizar) y la generación de algoritmos predictivos y de machine learning.

- **Interfaz (visualización).**
 - o Interfaz de administración y cuadros de mandos intuitivos desde la que configurar todas las opciones disponibles en la herramienta. La visualización dinámica de información y su segmentación personalizada.

- **Generación de informes.**
 - o Generación de distintos tipos de informes según el perfil al que vayan destinados. Con indicadores (KPIs) adaptados a las necesidades y con fuentes alternativas de información, etc.

La mayoría de las empresas coinciden en que el grueso de las tareas de innovación y el desarrollo de soluciones ad hoc, se centrará en la **recolección de información y el procesamiento de la misma** (algoritmos predictivos y machine learning). Hay que decir, no obstante, que las tecnologías seleccionadas para realizar el machine learning ya incluyen los algoritmos predictivos, y la tarea de desarrollo debería centrarse en el entrenamiento de los algoritmos y su integración en la herramienta final.

Por otro lado, no parece que la **visualización** o la **generación de informes** vayan a precisar unos esfuerzos relevantes según las respuestas de las empresas. Sin embargo, la generación de informes y sobre todo la obtención de KPIs adecuados, con una proactividad adecuada para buscar fuentes enriquecedoras o alternativas (en caso de caída de fuentes de información) resulta vital para el proyecto, ya que es su objetivo fundamental.

Además de todo lo anterior, existe un punto crítico al que las empresas tan sólo se refieren de forma puntual y es el desarrollo de una contingencia que permita obtener los diferentes KPIs y la información asociada aun cuando las fuentes de datos no se mantengan en el tiempo o cuando su modelo de datos cambie.

Un ejemplo de variaciones de este tipo que han de ser tenidas en cuenta y han de estar respaldadas por un equipo experto, ocurrió en su día con la información meteorológica suministrada por la AEMET, de la que se nutrían multitud de aplicaciones y que dejó de ser una fuente a la que poder conectarse, con lo que los servicios dependientes de la misma se cayeron inmediatamente sin que hubiese una planificación alternativa.

Propuestas en el ámbito técnico

1. Aplicación concreta de productos / componentes de software libre

La mayoría de las compañías se inclina por la aplicación de software libre, indicando además los componentes de software libre relacionados con el almacenamiento y el procesamiento de datos, así como la visualización y la captura de datos, detallando el uso que se le puede dar a cada uno.

Se describen con gran profusión de detalles algunos de los componentes tecnológicos que se aplicarían, que ya se encuentran probados en proyectos anteriores y de los que se garantiza su robustez, escalabilidad y posibilidades de adaptación a distintos escenarios, por lo que su utilización en el marco del presente proyecto resulta adecuada.

Tan sólo una de las empresas no propone componentes de software libre y se limita a mencionar librerías concretas de machine learning o las herramientas de análisis en lenguajes concretos, sin especificar cuál sería su rol dentro de la herramienta final.

2. Aplicación concreta de productos / componentes de mercado

Entre las compañías participantes, cuatro de ellas no plantean el uso de componentes de mercado.

En concreto, no les parece coherente utilizar componentes o productos que no sean *open source*, pues consideran que hay numerosas soluciones y productos *open source* solventes y tecnológicamente avanzadas que pueden responder perfectamente a las necesidades del proyecto. Además, existen tecnologías de código abierto, las cuales son estándares dentro de las soluciones Big Data, por lo que se podría prescindir completamente soluciones de mercado.

Por el contrario, las compañías restantes sí plantean el uso de componentes de mercado. La elección de los mismos se basa en su papel como partner de diferentes firmas de software especializada o por haber desarrollado su propio ecosistema Big Data, pero su elección no parece aportar una ventaja adicional ni por funcionalidades, ni por robustez ni por coste, tal y como se verá en los apartados siguientes.

3. Propuestas para Arquitectura Big Data

La mayoría de compañías especifican de forma detallada los componentes de la arquitectura propuesta, indicando cuáles son los roles de cada tecnología Big Data.

Las empresas participantes han propuesto gran variedad de tecnologías basadas en análisis Big Data. Algunas compañías sugieren tecnologías que sí son adecuadas para el desarrollo del proyecto, pero sin definir qué partes de cada

una de ellas se utilizarían, así como la justificación de su elección (funcionalidades, costes, facilidad de uso y desarrollo, etc.).

Como rasgos comunes a las tecnologías Big Data propuestas, se pueden citar los siguientes:

- La mayoría de soluciones tecnológicas propuestas poseen un catálogo extensible de conectores para una gran diversidad de orígenes de datos, tanto estructuradas como desestructuradas, lo que las hace configurables a la hora de añadir nuevas fuentes de información.
- Se plantean varias tecnologías que ofrecen un amplio rango de opciones con los que abordar cualquier problema de modelado predictivo. Es importante destacar que algunas compañías también plantean la combinación de varios lenguajes para la creación de modelos analíticos aplicados a grandes cantidades de datos, los cuales también pueden ser integrados a su vez sobre plataformas de procesado Big Data, tanto comerciales como *open source*.

Adicionalmente, un reducido número de empresas proponen otras tecnologías para cubrir otras necesidades que, aunque no son expresamente tecnologías basadas en Big Data, merece la pena mencionar en este apartado, como la indexación de términos a través de un buscador basado en lenguaje natural o soluciones para la securización de la plataforma, que resulta uno de los aspectos críticos que se deberán tener en cuenta.

4. Modelo (s) tecnológico (s) para el almacenamiento, acceso y distribución de información y datos.

Las compañías detallan, en mayor o menor medida, sus modelos para el almacenamiento, acceso y distribución de la información y los datos. La mayoría de compañías basan sus modelos en una combinación de bases de datos (tanto relacionales como no relacionales) y sistemas de archivos de tipo Big Data.

La mayoría fundamenta el modelo tecnológico en el uso de bases de datos relacionales y no relacionales, especificando en algunos casos tecnologías concretas.

Las plataformas se plantean con varios niveles o capas para la recolección de los datos, su tratamiento y la activación de la información, que parece adecuada a los objetivos de HAID IMPULSA.

5. Modelo (s) tecnológico (s) para el análisis de datos, extracción de conocimiento y su presentación (visualización) a cada tipo de usuario final.

La mayoría de compañías detallan el modelo tecnológico para el análisis de datos, la extracción de conocimiento y su presentación a cada tipo de usuario final.

En concreto, algunas propuestas se basan en soluciones escalables basadas en herramientas *open source* como Pentaho, dado que, si el volumen de datos no es muy elevado, resulta posible hacer analítica directamente desde la fuente en cambio, si el volumen es muy grande (o se trata de datos no estructurados) será necesario disponer de un entorno Big Data complementario a éste, para realizar el almacenamiento y el procesamiento de los datos. Por otro lado, para realizar la implementación de los algoritmos de machine learning se plantea el uso de herramientas como Flink o Spark, o los lenguajes de programación *open source* R y Python. La propuesta es coherente y plantea dos alternativas según sea el tipo y tamaño de los datos.

El resto de compañías propone el uso de herramientas comerciales sobre las que actuarían componentes específicos complementarios *open source* y la utilización mixta de lenguajes de programación tanto libres como propietarios.

6. Herramientas visuales o productos con los que cuentan las compañías

Varias compañías cuentan con herramientas visuales o productos propios para el propósito del proyecto, por lo que podrían adaptarse a la herramienta HAID IMPULSA. Las propuestas constan principalmente de aplicaciones web responsivas que permiten el acceso desde cualquier dispositivo y la construcción ágil de cuadros de mando, permitiendo el desarrollo autónomo de nuevas visualizaciones de forma independiente y transparente para el usuario. También se menciona una librería *open source* que permite visualizar datos masivos, desarrollada en el marco de un proyecto H2020.

Finalmente, y aunque no son herramientas visuales de desarrollo propio, algunas compañías proponen el uso de herramientas desarrolladas por terceros (de mercado) como herramientas visuales con las que es posible realizar cuadros de mando (tanto generalistas como a medida con diseño específico), reportes, cubos, y gráficos diversos; así como embeber cualquier visualización en aplicaciones existentes.

7. Solución basada en la nube

La mayor parte de las propuestas sugieren el despliegue de la solución en la nube, ya que la naturaleza del proyecto se alinea con los beneficios aportados por la virtualización y distribución de los recursos. Ya que el volumen de datos y capacidad de procesamiento puede aumentar de manera incremental, los recursos de almacenaje y computación podrán ser ajustados de forma elástica. La gestión de estos recursos en la nube se realiza con mayor eficiencia, comodidad, accesibilidad, rapidez, seguridad, escalabilidad, tolerancia fallos, y mayor personalización en la configuración.

Además, la mayoría de los elementos de software mencionados por la mayoría de las propuestas tienen soporte para la virtualización y despliegue de forma distribuida en la nube.

No obstante lo anterior, aunque la herramienta sea una solución basada en la nube, los costes de alojamiento deberán ser fijos durante un periodo que se habrá de precisar en la fase de una hipotética contratación.

8. Respuesta a aspectos Reglamento (UE) 2016/679

La mayor parte de las propuestas aseguran el cumplimiento de la nueva legislación en materia de protección de los datos (GDPR), centrada en la protección de los datos personales gestionados por la herramienta. En general, no se especifican los protocolos o controles que se implementarán, ya que dependerán tanto de las fuentes de información como del grado de anonimización que se requiera por parte de sus equipos (aludiendo además a la obligatoria cooperación por parte de los proveedores de datos).

Sólo en un caso, se indica que no será necesario la aplicación de procesos o protocolos en este sentido puesto que no se tratarán datos personales. Este enfoque contrasta con el resto de propuestas analizadas, puesto que sí se espera que este aspecto sea crucial, ya que las fuentes de información a las que se accederán (muchas de ellas desestructuradas o bien no anonimizadas), requerirán de un proceso de trabajo para precisamente cumplir los requisitos legales.

Propuestas en la gestión y la explotación

1. Metodologías de gestión y desarrollo del proyecto

Scrum como metodología de desarrollo

La totalidad de las propuestas apuestan por metodologías tipo **Scrum** (o Kanban) para la gestión del proyecto durante su fase de desarrollo. Scrum permitirá definir y llevar a cabo una estrategia ágil e incremental, que permitirá plasmar resultados tangibles desde las primeras etapas del desarrollo, permitiendo adecuar los posibles cambios en los requisitos y desviaciones que se produzcan en la fase de desarrollo de la plataforma.

Este tipo de metodologías permitirá disponer de resultados (herramientas, módulos y prototipos funcionales) desde las primeras semanas de vida del proyecto, ofreciendo la posibilidad de ajustar los diseños, alcances y otras desviaciones desde un inicio, sin que afecte críticamente al proyecto.

Se considera un tipo de metodología de desarrollo adecuado para el proyecto y está avalada por las principales empresas productoras de tecnología. Gracias a esta tecnología se podrán adecuar los requisitos inestables (por ejemplo, posibilidad de incorporar nuevos conjuntos de datos o tipos de análisis) de una forma flexible y rápida. Se asegura el cumplimiento de las expectativas del cliente ya que se centra en obtener resultados anticipados alineados con los requisitos del proyecto. Scrum es útil para mitigar riesgos en la gestión del proyecto, teniendo en cuenta la productividad del equipo y la calidad del producto y servicio.

Scrum está basado en la organización autónoma de equipos multidisciplinares, fomentando la comunicación constante entre todos los miembros involucrados en el proyecto. Para llevar a cabo esta metodología, el equipo del proyecto debe disponer de roles específicos que garanticen la productividad y efectividad de las iteraciones periódicas o ciclos en los que se compone el proyecto. El rol de *ScrumMaster*, el facilitador de los procesos Scrum, es crucial para un correcto avance del proyecto. Para la correcta consecución del proyecto, se recomienda incluir a un representante de IMPULSA en las reuniones de revisión de cada *sprint*.

En la planificación cabe destacar el concepto de 'sprint', Un *sprint* es cada uno de los periodos temporales en los que realiza el trabajo en sí. Estos *sprints* son ajustables en función de las necesidades del proyecto y suelen tener una duración de entre 2 y 4 semanas. Al final de cada *sprint* el equipo deberá presentar los avances logrados y éstos suelen producir un producto o servicio funcional, cuyas características aumentan y mejoran tras cada interacción.

En contraposición del uso de metodologías ágiles, una única empresa se apoya el uso de la metodología incremental en cascada **MÉTRICA 3**, promovida por el Ministerio de Hacienda y Función Pública del Gobierno de España para la sistematización de actividades del ciclo de vida de los proyectos software en el ámbito de las administraciones públicas.

Metodología de administración del proyecto

Únicamente dos propuestas sugieren el uso de metodologías de gestión y administración del proyecto a alto nivel. Una de ellas propone el uso de metodologías estándar de administración y gestión de proyectos **Prince2 e ITIL**, para las que dispone de equipo certificado. Ambas metodologías se basan en la definición y planificación de procesos para asegurar la correcta gestión del proyecto, desde la concepción del producto hasta los servicios de mantenimiento y actualización.

Asimismo, la otra empresa propone una administración de proyecto basada en una metodología propia basada en **Modelo de Madurez de la Capacidad Integrado**.

Para el diseño, desarrollo e implementación de las soluciones de minería de datos e inteligencia de negocio, otras empresas sugieren el uso de **CRISP-DM** (Cross Industry Standard Process for Data Mining), un estándar de facto utilizado en la ejecución de proyectos de referencia en el ámbito analítico y de minería de datos. CRISP-DM se basa en un proceso iterativo dividido en fases, y con una serie de actividades enfocadas en conseguir unos objetivos concretos.

2. Recursos humanos y materiales

La mayoría de empresas mencionan los distintos perfiles de personal que serán necesarios para ejecutar la solución. Casi todas las compañías coinciden en el tipo de perfiles (data scientists, economistas, analistas, programadores, diseñadores, jefes de proyecto, consultores funcionales, consultores de Business Intelligence y Big Data, analistas programadores y arquitectos Big Data).

En este punto es importante resaltar el importante papel que deberían tener los expertos dedicados a garantizar la emisión de los distintos KPIs y el mantenimiento de su funcionamiento y reemplazo en caso de pérdida de conexión. No hay que olvidar que el objetivo final de la herramienta es la obtención de dichos indicadores, bajo cualquier circunstancia.

El resto de empresas sí especifican los recursos materiales, especificando en algunos casos de forma detallada los recursos materiales necesarios para llevar a cabo su propuesta, pero en general no se especifican los costes o riesgos.

Algunas empresas no especifican los recursos materiales, por el grado de madurez en la definición de los requisitos de la herramienta.

Adecuación del presupuesto

1. Propuestas de mantenimiento de la herramienta y su coste

Existen compañías que plantean su propuesta de mantenimiento, pero no detallan los costes asociados, ya que afirman no disponer de suficiente información al respecto.

En concreto, demandan información de aspectos que habrán de concretarse en una hipotética fase de adquisición de la herramienta, siguiendo algunos de ellos los siguientes:

- Si la aplicación va a estar alojada en sistemas propios de Impulsa o en terceros.
- Si la gestión de usuarios va a estar integrada en la herramienta o se va a gestionar desde la plataforma del ayuntamiento.
- La futura evolución de las fuentes, sobre todo las externas, tanto en variedad como en volumen y su correspondiente integración en la aplicación.
- Los requerimientos de soporte al usuario (nivel de servicio, horarios...).
- Que se incluya en el planteamiento final del proyecto información relativa a los costes de mantenimiento de los tres primeros años.

Otras compañías que ofrecen soluciones corporativas de mercado, afirman que el mantenimiento, al estar toda la herramienta sobre una plataforma comercial, lo ofrece el propio fabricante. En este caso, las empresas afirman que el soporte lo aporta el propio fabricante con un coste anual que depende del tamaño del entorno, ya que se licencia por los núcleos del mismo. Por tanto, es preciso realizar un trabajo previo de dimensionamiento para determinar el alcance y los costes previsibles. No obstante, plantean la posibilidad de partir de un presupuesto básico que se vaya incrementando conforme aumenten las necesidades. También se plantea un coste anual para el soporte de la infraestructura pero sin especificar su alcance.

Finalmente, hay dos compañías que sí dan una estimación detallada del mantenimiento y su coste asociado sin necesidad de disponer de información adicional. En este caso, su coste resulta variable en función del mantenimiento de la propia infraestructura cloud y los reajustes en caso de que haya cambios en las fuentes de datos. Con respecto a los costes de los reajustes, afirman que no pueden cuantificarse al depender de terceras partes, no obstante, estos costes deberían ser muy bajos al defender la compañía que la plataforma permite integrar futuras fuentes de datos gracias a su diseño modular.

2. Adecuación del presupuesto previsto para satisfacer las necesidades de la herramienta

A la hora de valorar la adecuación del presupuesto previsto para satisfacer las necesidades de la herramienta, se podría decir que existen tres posiciones bien diferenciadas:

- 1) *Compañías que afirman no disponer de suficientes elementos de juicio para determinar si el presupuesto es adecuado.*

Estas compañías, independientemente de las soluciones que propongan implantar, plantean una serie de indeterminaciones en la definición de las necesidades y alcance de la herramienta, por lo que se debería calcular los costes de personal, hardware y licencias necesarias para la realización del porcentaje de desarrollo ad-hoc estimado y poder así determinar si el presupuesto es adecuado.

- 2) *Compañías que consideran adecuado el presupuesto si sólo se considera la realización de alguna de las partes de la herramienta, existiendo además dos supuestos concretos en ese sentido, que se citan a continuación.*

Alguna compañía considera que el presupuesto es adecuado si sólo se contempla la plataforma analítica y los servicios de integración de datos y visualización, prescindiendo del ecosistema Big Data completo (la infraestructura, su tamaño y su implantación) por lo que llega a proponer un trabajo de consultoría previo para dimensionar el tamaño necesario de dicha infraestructura. No obstante, dada la cantidad de fuentes de datos que intervienen en el proyecto, así como la creación de los modelos predictivos y el desarrollo de la herramienta, considera que el presupuesto propuesto para el desarrollo global no es adecuado.

Otra compañía plantea una primera fase de consultoría e implantación de una herramienta comercial para todo el proceso, pero que dicha herramienta excluiría las altas capacidades de almacenamiento y tratamiento de información que se requieren en el marco del proyecto, por lo que no se adecúa a las necesidades planteadas.

- 3) *Una compañía que sí lo considera adecuado.*

Finalmente, una compañía considera que el presupuesto es adecuado para cubrir los objetivos, y detalla las partidas de gasto a las que va a ir dedicado: diseño y desarrollo de los componentes software, las licencias (aunque plantea usar las distribuciones gratuitas manteniendo la posibilidad de adquirir las licencias "enterprise" con soporte y garantía), y los gastos de mantenimiento del servicio cloud (sin especificar su coste).