



UAECD

# Metodología para la Analítica de datos

—

Guía de buenas prácticas para analítica

Actualizado abril 2019

2018

[www.  
ideca.  
gov.  
co](http://www.ideca.gov.co)

Metodología para la Analítica de datos

Guía de buenas prácticas para analítica

---

Fecha de creación: Abril de 2019

Página web: [www.ideca.gov.co](http://www.ideca.gov.co)

Correo electrónico: [ideca@catastro.gov.co](mailto:ideca@catastro.gov.co)

Licencia: Attribution 4.0 International (CC BY 4.0)

Autores: Unidad Administrativa Especial de Catastro Distrital -  
Gerencia IDECA



Control De Versiones

Fecha	Autor/ Modificado por	Versión	Cambio efectuado
2019-04-30	Camila Torres James Holman Rivera Juan Carlos Méndez	1.0	Primera versión del documento. No hay cambios para registrar.



Tabla de Contenido

RESUMEN EJECUTIVO .....6

1.1. GLOSARIO .....6

1.2. ABREVIATURAS.....6

AUDIENCIA.....7

2.1. INTERNA .....7

2.2. EXTERNA.....7

INTRODUCCIÓN .....8

ANTECEDENTES .....9

4.1. CRISP-DM.....9

4.2. ASUM-DM.....10

4.3. VARIACIÓN DE ASUM-DM .....11

5.1. CICLO DE ELABORACIÓN .....15

5.1.1. Inicio .....15

5.1.2. Entendimiento del negocio.....15

5.1.3. Descubrimiento y Entendimiento de datos.....16

5.1.4. Preparación y gestión de datos .....19

5.1.5. Validación de datos.....21

5.1.6. Análisis.....22

5.1.7. Prototipo.....23

5.1.8. Validación y evaluación.....24

5.1.9. Nueva adquisición de datos .....25

5.2. CICLO DE PUBLICACIÓN .....26

5.2.1. Publicación.....26

5.2.2. Publicar Hallazgos .....27

5.2.3. Monitorear .....27

5.2.4. Operar.....28

5.2.5. Soporte.....29

5.2.6. Gestión de Infraestructura .....29

5.2.7. Integración de aplicaciones.....30

5.2.8. Gestión del ciclo de vida y mejora continua .....30

CONCLUSIONES .....32

REFERENCIAS .....33



## Listado de Gráficos

Ilustración 1 Metodología CRIP-DM..... 9

Ilustración 2 Adaptación de la metodología ASUM-DM..... 10

En Ilustración 3 Metodología ASUM-DM con validación posterior a la preparación de datos11

Ilustración 4 Variación de la Metodología ASUM-DM para analítica en IDECA .....13

Ilustración 5 Metodología para Analítica en IDECA. .... 14

## Listado de Tablas

Tabla 1 Ciclo de Elaboración, fase de Inicio.....15

Tabla 2 Ciclo de Elaboración, fase de Entendimiento del negocio .....15

Tabla 3 Ciclo de Elaboración, fase de Entendimiento de datos..... 16

Tabla 4 Ciclo de Elaboración, fase de Preparación y gestión de datos:..... 19

Tabla 5 Ciclo de Elaboración, fase de Preparación y gestión de datos: ..... 19

Tabla 6 . Ciclo de Elaboración, fase de Preparación y gestión de datos: Preparar datos.  
Fuente: Elaboración Propia ..... 20

Tabla 7 Ciclo de Elaboración, fase de Validación de datos.....21

Tabla 8 . Ciclo de Elaboración, fase de Análisis.....22

Tabla 9 Ciclo de Elaboración, fase de Prototipo.....23

Tabla 10 Ciclo de Elaboración, fase de Validación y evaluación. .... 24

Tabla 11 . Ciclo de Elaboración, fase de Nueva adquisición de datos.....25

Tabla 12 Ciclo de Publicación, fase de Publicación. .... 26

Tabla 13 . Ciclo de Publicación, fase de Publicar hallazgos. .... 27

Tabla 14 Ciclo de Publicación, fase de Monitorear. .... 27

Tabla 15 Ciclo de Publicación, fase de Monitorear. .... 28

Tabla 16 Ciclo de Publicación, fase de Soporte. .... 29

Tabla 17 Ciclo de Publicación, fase de Gestión de Infraestructura. .... 29

Tabla 18 Ciclo de Publicación, fase de Integración de aplicaciones. .... 30

Tabla 19 Ciclo de Publicación, fase de Gestión del ciclo de vida y mejora continua..... 30



---

## Resumen ejecutivo

### 1.1. Glosario

#### ASUM-DM

Por sus siglas en inglés, Analytics Solutions Unified Method for Data Mining/predictive analytics. Es una metodología para realizar implementaciones ágiles en temas de analítica de datos y minería de datos.

#### BICC

Por sus siglas en inglés: Business Intelligence Competency Center, es un Centro de Competencias de Inteligencia de Negocios es un equipo multidisciplinar y multifuncional de la organización cuya misión es respaldar y promover el uso adecuado de la BI/BA.

#### Big data

Es un término usado para describir la gran cantidad de datos digitales o que pueden ser digitalizados y que pueden ser almacenados, analizados u usados para un fin en particular.

#### CRISP-DM

Por sus siglas en inglés, Cross Industry Standard Process for Data Mining, es una metodología que presenta un proceso jerárquico en donde se tienen unas tareas para hacer minería de datos de una manera ágil y con mejores resultados.

### 1.2. Abreviaturas

<b>ARA:</b>	Arquitectura de Referencia para Analítica
<b>ATR:</b>	Arquitectura Tecnológica de Referencia para la UAECD
<b>ATS:</b>	Arquitectura Tecnológica de Solución
<b>IDECA:</b>	Infraestructura de Datos Espaciales de Bogotá
<b>UAECD:</b>	Unidad Administrativa Especial Catastro Distrital



**UAECD**

---

## Audiencia

### 2.1. Interna

- Científicos de datos, Ciudadanos científicos de datos
- Arquitectos de negocio, empresarial y solución de la entidad
- Demás dependencias de la dirección y otras gerencias que tengan propuestas e implementen iniciativas orientadas a la construcción de soluciones relativas a la explotación y aprovechamiento de los datos en la entidad (inteligencia de negocios, analítica de negocios, big data, etc).

### 2.2. Externa

- Proveedores de software y hardware alineados con la propuesta de valor de la entidad
- Demás socios estratégicos (entidades de la administración distrital y otros).



---

## Introducción

La Arquitectura de Referencia para la Analítica de datos (ARA) establece el marco para el aprovechamiento de datos con fines estratégicos para el negocio. De esta arquitectura es posible generar arquitecturas de solución que pueden resultar en ejercicios como Analítica de datos, Big Data o Aprendizaje de Máquina.

Los proyectos de aprovechamiento de la información con la que cuenta la organización, proyectos tales como Analítica de datos, debe ayudar al negocio en varios propósitos, como los pronósticos, las proyecciones, el descubrimiento de patrones, comportamientos, tendencias, oportunidades, que no son identificables fácilmente y que pueden usarse para mejorar o transformar el negocio.

El uso de una metodología para soluciones de analítica de datos asegura en un alto porcentaje el éxito del proyecto, debido a que los elementos involucrados en el ciclo de vida de los ejercicios de analítica mantienen unas directrices, reglas y lineamientos establecidos para cumplir con un propósito en particular, la metodología de analítica permite establecer un plan de trabajo adecuado.

La metodología de analítica permite:

- Establecer una forma probada para llevar un problema a una solución de analítica
- Alinear la analítica con la intención estratégica de la organización
- Lograr eficiencia en algún aspecto del negocio

Este documento constituye una propuesta que aborda aspectos como guía de buenas prácticas para ejercicios de analítica, recomendaciones y demás relacionados con la integración de la información geográfica gestionada por IDECA con enfoque hacia técnicas de analítica de datos.





## Antecedentes

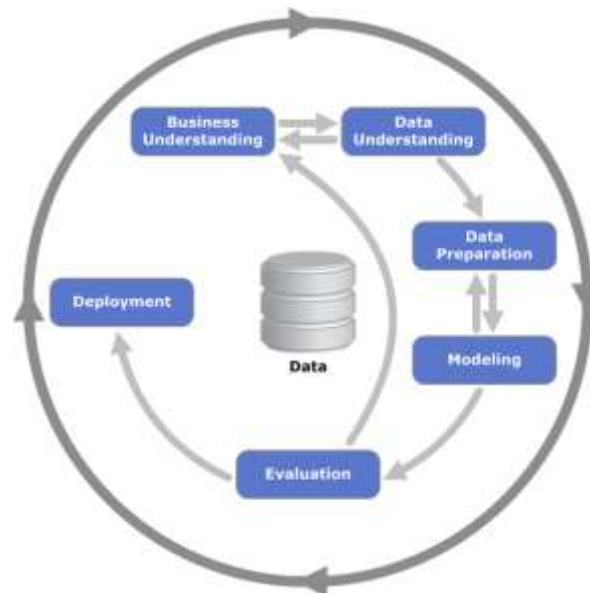
### 4.1. CRISP-DM

La metodología CRISP-DM (*Cross Industry Standard Process for Data Mining*) es un modelo estándar abierto propuesto en 1999 por IBM para proyectos relacionados con minería de datos.

CRISP-DM divide el proceso de minería de datos en seis fases principales según se muestra en la Ilustración 1 Metodología CRIP-DM<sup>1</sup>: Entendimiento de negocio, entendimiento de datos, preparación de datos, modelamiento, evaluación y publicación.

G1

Ilustración 1 Metodología CRIP-DM  
Fuente: Citación 1



Posteriormente en el año 2015 IBM propuso una versión extendida de CRISP-DM denominada ASUM-DM (*Analytics Solutions Unified Method*) la cual contiene los mismos pasos de CRISP-DM bajo el nombre de “Ciclo de Elaboración” e introduce un nuevo ciclo denominado “Operacional”.

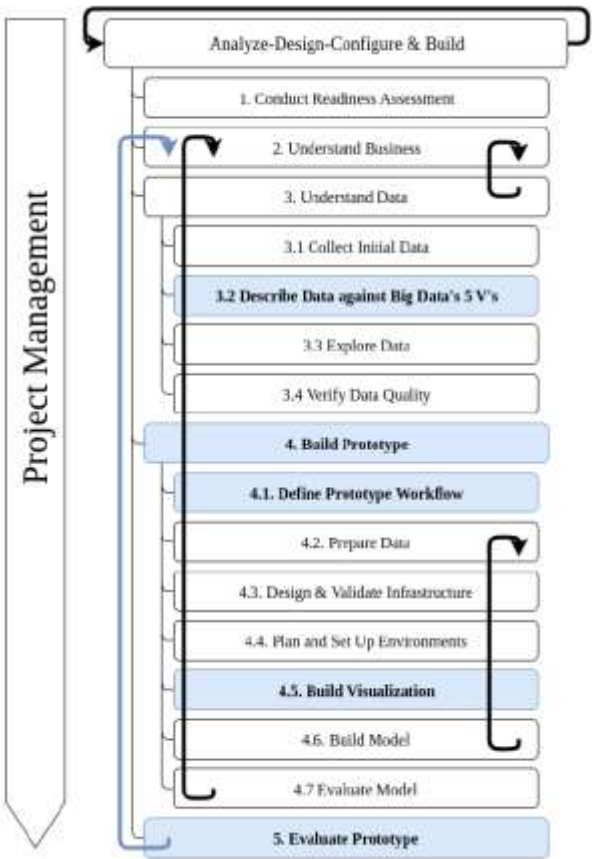
<sup>1</sup> [https://commons.wikimedia.org/wiki/File:CRISP-DM\\_Process\\_Diagram.png](https://commons.wikimedia.org/wiki/File:CRISP-DM_Process_Diagram.png)



4.2. ASUM-DM

ASUM-DM de IBM se basa en la metodología CRISP-DM y hace énfasis en las nuevas prácticas en la ciencia de datos como el uso de volúmenes de datos muy grandes, la incorporación de análisis de texto, en el modelado predictivo y la automatización de algunos procesos<sup>2</sup>.

G2 Ilustración 2 Adaptación de la metodología ASUM-DM  
Fuente: Citación 1



<sup>2</sup> CAOBA. Perfil Alianza Caoba, Reporte técnico. Universidad de Los Andes. Julio 2017



IBM define ASUM como una guía paso a paso para efectuar una implementación completa del ciclo de vida de las soluciones de analítica de datos. Fue creado para acelerar el tiempo para la generación de valor y disminución del riesgo mediante el establecimiento de enfoques y procesos coherentes que aumentan la eficiencia de su implementación. Contiene pasos estructurados, actividades de desarrollo, roles y responsabilidades, plantillas y directrices<sup>3</sup>.

En la Ilustración 2 Adaptación de la metodología ASUM-DM tomada de<sup>4</sup>. Se plantea una modificación al modelo ASUM-DM original en donde se plantean nuevos elementos o fases orientados a proyectos que específicamente involucren grandes volúmenes de información ("Big Data") las cuales se encuentran sombreadas con color azul claro (numerales 3.2, 4, 4.1, 4.5, 5).

### 4.3. Variación de ASUM-DM

En Ilustración 3 Metodología ASUM-DM con validación posterior a la preparación de datos la basada en <sup>5</sup> y <sup>6</sup> se muestran los dos ciclos fundamentales de la metodología propuesta: *Ciclo de Elaboración* y *Ciclo de Publicación*.

En la Ilustración 4, se presenta una nueva gráfica, modificación de la Ilustración 3, en la cual se incorpora la evaluación del modelo, dentro de la fase de Validación del modelo. Así mismo, se presentan algunos ajustes al ciclo de publicación de datos, cuya finalidad es dar mayor alcance a los ejercicios de analítica. Cada una de las fases son descritas en el siguiente capítulo (Metodología).

---

<sup>3</sup> IBM Analytics. Analytics Solutions Unified Method. Implementations with agile principles. IBM Corporation. Estados Unidos, marzo de 2016. En internet: <ftp://ftp.software.ibm.com/software/data/sw-library/services/ASUM.pdf>

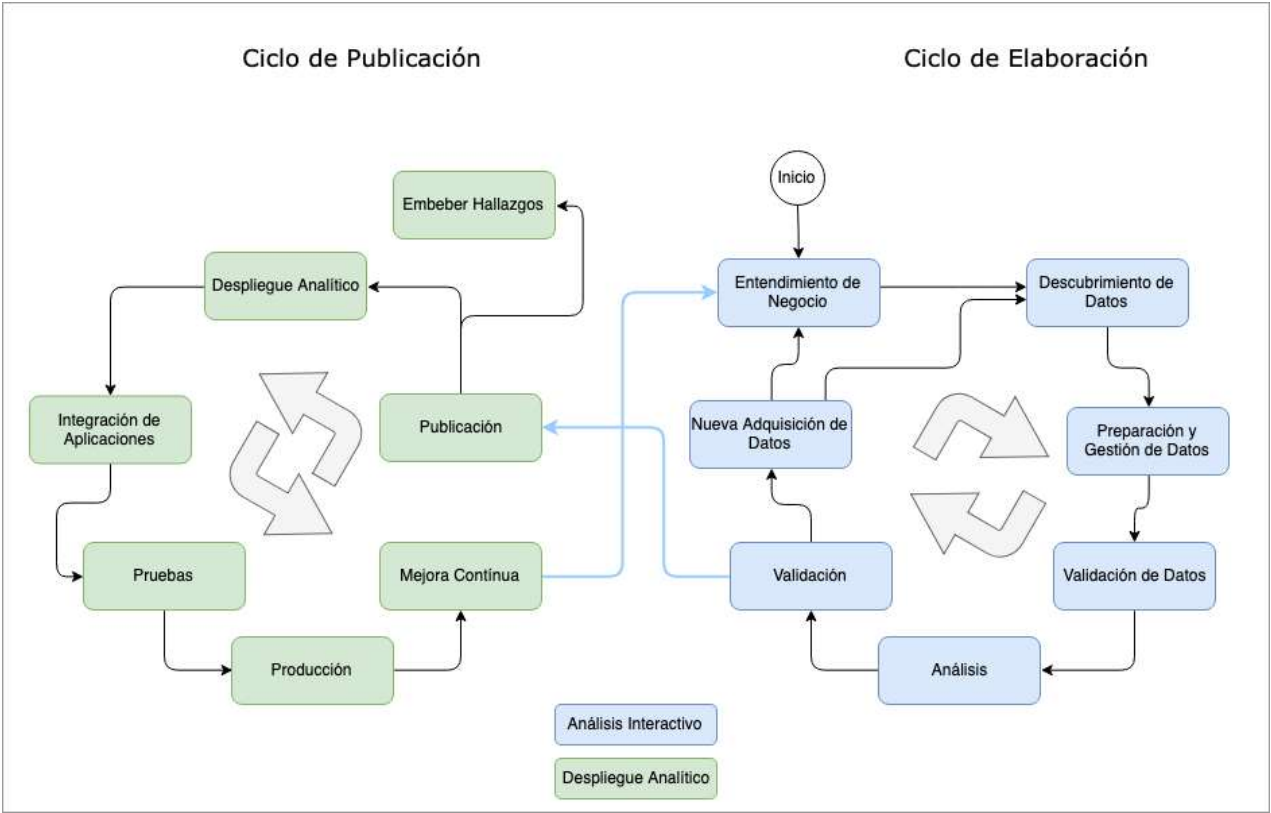
<sup>4</sup> Angée, S., Lozano-Argel, S. I., Montoya-Munera, E. N., Ospina-Arango, J.-D., & Tabares-Betancur, M. S. (2018). Towards an Improved ASUM-DM Process Methodology for Cross-Disciplinary Multi-organization Big Data & Analytics Projects. In Communications in Computer and Information Science (pp. 613–624). Springer International Publishing. [https://doi.org/10.1007/978-3-319-95204-8\\_51](https://doi.org/10.1007/978-3-319-95204-8_51)

<sup>5</sup> <https://www.kdnuggets.com/2017/02/analytics-grease-monkeys.html>

<sup>6</sup> <https://sharing.luminis.eu/blog/the-forgotten-step-in-crisp-dm-and-asum-dm-methodologies/>

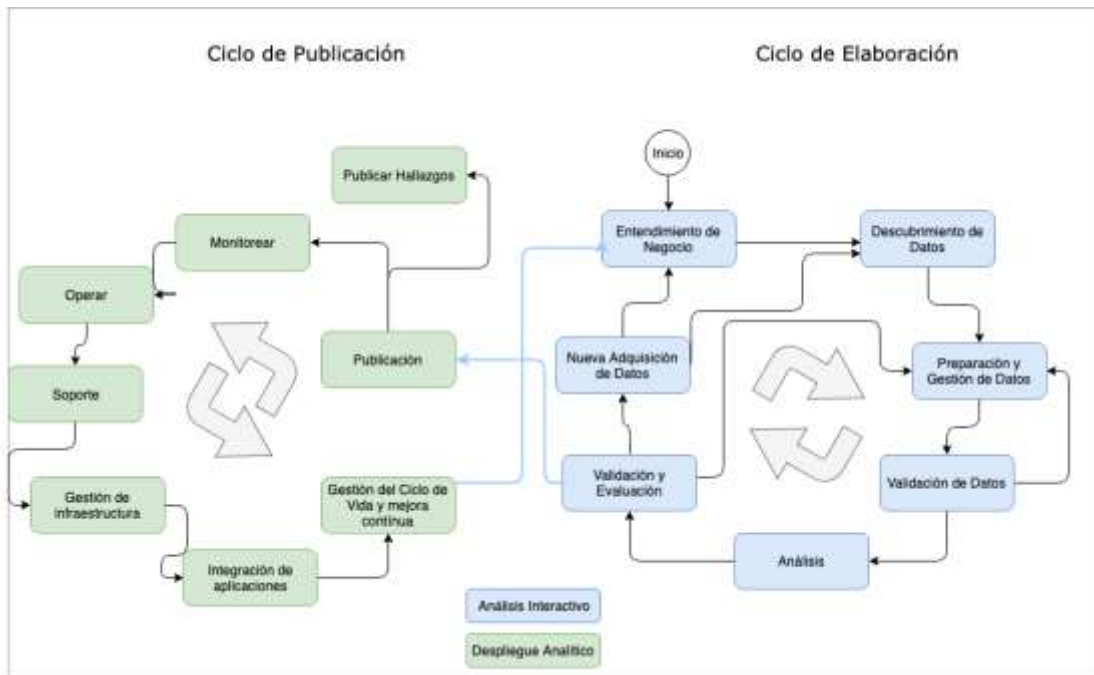


G3 Ilustración 3 Metodología ASUM-DM con validación posterior a la preparación de datos  
Fuente: Citación 5 y 6



G4

Ilustración 4 Variación de la Metodología ASUM-DM para analítica en IDECA  
Fuente: Citación 5 y 6.



## Metodología de Analítica para IDECA

Para la realización de analítica en IDECA se plantea una metodología basada en ASUM-DM detallada en la sección

Metodología en donde se introducen algunas variaciones relacionadas con elementos adicionales a tener en cuenta respecto a la validación de datos e implementación de proyectos de Big Data.

Para la realización de analítica en IDECA se plantea una metodología basada en ASUM-DM, incorporando una validación posterior a la preparación de datos y un ciclo de vida para la publicación de datos.

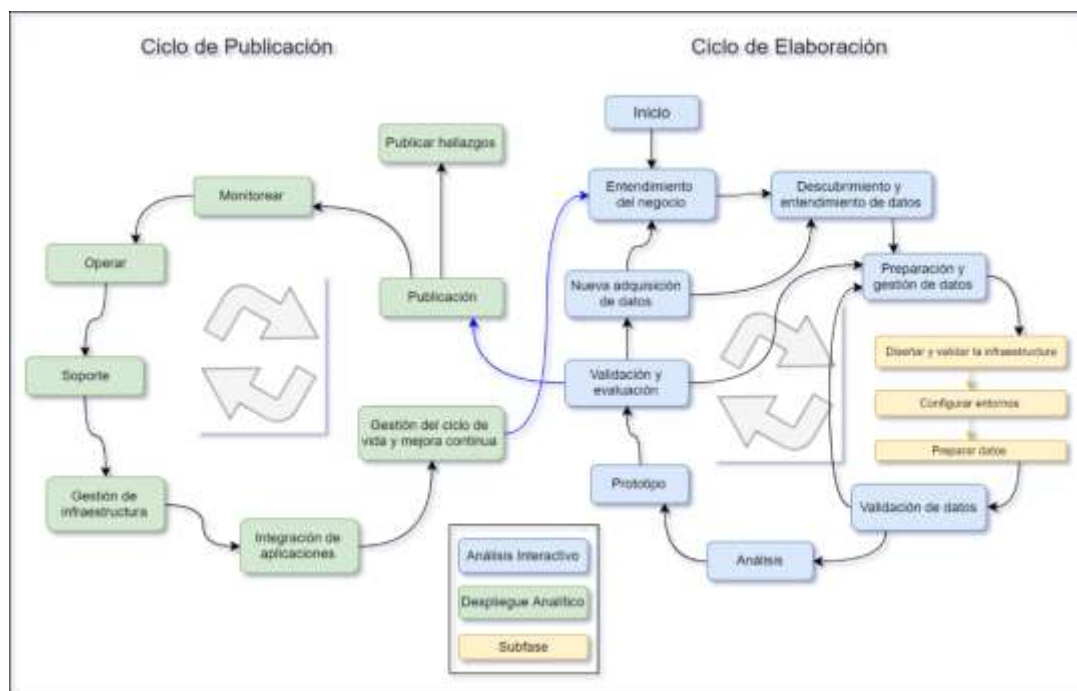


UAECI

www.ideca.gov.co | (57) 1 2347600 | ideca@catastrobogota.gov.co

La metodología propuesta para proyectos y ejercicios de analítica en IDECA y UAECD se presenta en la **¡Error! No se encuentra el origen de la referencia..**

G5 Ilustración 5 Metodología para Analítica en IDECA.  
Fuente: Elaboración propia



En esta metodología se presenta el resumen de las actividades a realizar durante la ejecución de proyectos de analítica, que se divide en dos ciclos: Ciclo de elaboración y ciclo de publicación. A continuación, se describen cada una de las fases.



## 5.1. Ciclo de Elaboración

### 5.1.1. Inicio

T1

Tabla 1 Ciclo de Elaboración, fase de Inicio.  
Fuente: Elaboración Propia

Tarea	Evaluar la preparación de la entidad para la implementación. Es decir, evalúa que tan lista está la organización/entidad para iniciar el proyecto de analítica.	
Descripción	Se evalúa los diferentes factores de implementación / administrativos, factores de construcción y ubicación, factores de tecnología, fuentes de datos y factores de volúmenes que podrían afectar la implementación del proyecto e identifique soluciones para mitigar el impacto de esos factores.	
Actividades	Producir plan del proyecto	
Roles	<b>RA01</b> Patrocinador del Proyecto <b>RA13</b> Científico de Datos <b>RA19</b> Arquitecto Empresarial <b>RA04</b> Gerente de Datos y Analítica <b>RA05</b> Analista de Negocio <b>RA06</b> Arquitecto de Datos	
Salidas	Plan de Proyecto. Declaración de trabajo (documento o contrato en donde se especifica los objetivos y entregables para el proyecto a realizar).	

### 5.1.2. Entendimiento del negocio

T2

Tabla 2 Ciclo de Elaboración, fase de Entendimiento del negocio  
Fuente: Elaboración Propia

Tarea	Entender los objetivos del proyecto y sus requerimientos desde la perspectiva del negocio, para convertir este conocimiento en un proceso de minería de datos y en un plan preliminar que logre cumplir con los objetivos.	
Descripción	Se determinan los objetivos del negocio y se evalúa su situación como por ejemplo conocer los riesgos/ contingencias, requerimientos. Además, determinar los objetivos de la minería de datos.	
Actividades	Determinación de los objetivos de negocio: <ul style="list-style-type: none"> <li>Recopilación del contexto del negocio</li> </ul>	



UAECD

	<ul style="list-style-type: none"> <li>• Objetivos de negocio</li> <li>• Criterios de éxito de negocio</li> </ul> <p>Evaluación de la situación:</p> <ul style="list-style-type: none"> <li>• Inventario de recursos, requerimientos, suposiciones y restricciones y contingencias.</li> <li>• Definición de Roles</li> <li>• Terminología</li> <li>• Análisis de costo beneficio.</li> </ul> <p>Determinar objetivos de minería de datos Criterios de éxito de minería de datos</p>
Roles	RA02 Líder de Negocio RA03 Usuarios de Negocio RA04 Gerente de Datos y Analítica RA05 Analista de Negocio RA06 Arquitecto de Datos RA11 Coordinador de Metadatos RA13 Científico de Datos RA15 Analista de Datos RA16 Analista de Datos Geográficos
Salidas	Documento de entendimiento de negocio.

### 5.1.3. Descubrimiento y Entendimiento de datos

T3 Tabla 3 Ciclo de Elaboración, fase de Entendimiento de datos.  
Fuente: Elaboración propia

Tarea	Acceder y explorar los datos con el fin de determinar la calidad de la información.
Descripción	Se recolectan los datos iniciales con el fin de describirlos, explorarlos y verificar su calidad.
Actividades	<p><b>Recolectar datos iniciales:</b> Conseguir y acceder a los datos (si estos son adquiridos de diferentes fuentes será necesario integrarlos). Los datos pueden provenir de fuentes como:</p> <ul style="list-style-type: none"> <li>• <b>Datos existentes:</b> incluye una gran variedad de información, por ejemplo: datos transaccionales, datos de encuestas y otros.</li> <li>• <b>Compra de datos:</b> Se usa en caso de necesitar datos que no se</li> </ul>



UAECD



encuentren disponibles en las fuentes de información de la organización.

- **Datos adicionales:** En caso de que los tipos de datos anteriores no sean suficientes se debe iniciar una exploración adicional de otros posibles almacenes de datos disponibles en la organización.

**Descripción los datos:** Describir los datos teniendo en cuenta los siguientes criterios

- Fuente
- Tipos de datos: Numérico, categórico, booleano, fecha, geográfico. Para el caso de los datos geográficos tener en cuenta el tipo de geometría y sistema de referencia.
- Dominios: Ejm. Sexo, rango etario, tipo de producto, etc.
- Volumen / Cantidad de datos
- Velocidad de cambio de los datos
- Variedad: Estructurado, No estructurado, semi estructurado
- Veracidad: Fiabilidad de los datos (Algunas de las fuentes utilizadas pueden tener características variables de fiabilidad)
- Valor: Valor que puede ser obtenido del análisis de los datos

**Exploración de datos:** Aborda las preguntas de minería de datos utilizando técnicas de consulta, visualización y reportes.

- Explorar distribución de los atributos clave, relaciones entre atributos, resultados de agregaciones simples, propiedades de subpoblaciones significativas y análisis estadísticos simples.
- Estos análisis pueden abordar directamente los objetivos de extracción de datos; también pueden contribuir o refinar la descripción de los datos y los informes de calidad, y alimentar la transformación y otros pasos de preparación de datos necesarios para un análisis más detallado<sup>7</sup>.
- Explorar los datos con las tablas, cuadros y otras herramientas de visualización.
- Formular hipótesis y dar forma a las tareas de transformación de datos que tienen lugar durante la preparación de los datos.

---

<sup>7</sup> [http://gforge.icesi.edu.co/ASUM-DM\\_External/index.htm#cognos.external.asum-DM\\_Teaser/tasks/sps\\_explore\\_data\\_E8427F39.html?proc=\\_OeKIHI6EeW\\_y7k3h2HTng&path=\\_OeKIHI6EeW\\_y7k3h2HTng,\\_OeHEyVt6EeW\\_y7k3h2HTng,\\_OeFPrFt6EeW\\_y7k3h2HTng,\\_OeFPnlt6EeW\\_y7k3h2HTng](http://gforge.icesi.edu.co/ASUM-DM_External/index.htm#cognos.external.asum-DM_Teaser/tasks/sps_explore_data_E8427F39.html?proc=_OeKIHI6EeW_y7k3h2HTng&path=_OeKIHI6EeW_y7k3h2HTng,_OeHEyVt6EeW_y7k3h2HTng,_OeFPrFt6EeW_y7k3h2HTng,_OeFPnlt6EeW_y7k3h2HTng)



	<b>Verificación de calidad de los datos</b> <ul style="list-style-type: none"> <li>• ¿Están completos?</li> <li>• ¿Son correctos?</li> <li>• ¿Qué tipos de errores hay?</li> <li>• ¿Con qué frecuencia ocurren errores?</li> <li>• ¿Existen nulos?</li> <li>• ¿Hay datos duplicados? ¿Con qué frecuencia? ¿Cómo se representan?</li> <li>• ¿Hay errores en el formato de los datos?</li> <li>• ¿Hay inconsistencias en los dominios o códigos en los datos?</li> <li>• ¿Hay errores entre los metadatos del conjunto de datos y la información encontrada?</li> <li>• ¿Falta información clave para poder utilizar los datos? (Ejem. Sistema de referencia)</li> </ul>
Roles	RA06 Arquitecto de Datos RA07 Administrador de Datos RA08 Administrador de Base de Datos RA11 Coordinador de Metadatos RA13 Científico de Datos RA15 Analista de Datos RA16 Analista de Datos Geográficos RA03 Usuarios de Negocio RA05 Analista de Negocio
Salidas	Reporte de entendimiento de datos que debe contener: <ul style="list-style-type: none"> <li>• Reporte de recolección de datos.</li> <li>• Reporte de descripción de datos.</li> <li>• Reporte de exploración de datos.</li> <li>• Reporte de calidad de datos.</li> </ul>



## 5.1.4. Preparación y gestión de datos

### 5.1.4.1. Diseñar y validar la infraestructura

T4

Tabla 4 Ciclo de Elaboración, fase de Preparación y gestión de datos:  
Diseñar y validar la infraestructura.

**Fuente:** Elaboración Propia

Tarea	Diseñar estrategias de infraestructura técnica y seguridad
Descripción	Verificar que hardware, software y demás componentes técnicos se necesitan. También se diseñará la infraestructura de seguridad.
Actividades	<p><b>Diseño de infraestructura técnica:</b></p> <ul style="list-style-type: none"> <li>Diseñar la arquitectura de los entornos requeridos (Ejm: Analítico, Control de calidad y producción) en términos del hardware y software que se requieran</li> <li>Validar la infraestructura técnica propuesta</li> </ul> <p><b>Diseño de infraestructura de seguridad:</b></p> <ul style="list-style-type: none"> <li>Diseñar los mecanismos de autenticación y autorización requeridos</li> <li>Documentar el modelo de seguridad y validarlo con los interesados (stakeholders) del proyecto</li> </ul>
Roles	Arquitecto empresarial
Salidas	Documento de entendimiento de negocio.

### 5.1.4.2. Configurar entornos

T5

Tabla 5 Ciclo de Elaboración, fase de Preparación y gestión de datos:  
Diseñar y validar la infraestructura.

**Fuente:** Elaboración Propia

Tarea	Configurar Entornos
Descripción	Configurar los diferentes entornos requeridos para las tareas de análisis ya sea en sitio o en la nube según el diseño y requerimientos establecidos.



**UAEC**

Actividades	<b>Configurar entorno analítico:</b> Entorno donde se ejecutan las tareas relacionadas con la elaboración y pruebas de los modelos. Configurar entorno de control de calidad Configurar entorno de producción
Roles	<b>RA19</b> Arquitecto Empresarial <b>RA21</b> Administrador de tecnología <b>RA20</b> Arquitecto de Soluciones
Salidas	Documento de entendimiento de negocio.

#### 5.1.4.3. Preparar datos

**T6** Tabla 6 . Ciclo de Elaboración, fase de Preparación y gestión de datos: Preparar datos.  
**Fuente:** Elaboración Propia

Tarea	Preparar datos
Descripción	Preparación de los datos para las tareas de análisis. Por lo general esta es una de las actividades que más consume tiempo en el proceso.
Actividades	<ul style="list-style-type: none"> <li>• Seleccionar datos: Decidir qué datos son útiles para el análisis.</li> <li>• Limpiar datos: Selección de subconjuntos, asignación de valores por defecto, etc.</li> <li>• Construir datos: Producción de atributos derivados o transformación de datos.</li> <li>• Integrar datos: Integración de datos de múltiples fuentes.</li> <li>• Formatear datos: Transformación de datos respecto a formatos sin cambiar su significado.</li> </ul>
Roles	<b>RA13</b> Científico de Datos <b>RA14</b> Investigadores especializados <b>RA15</b> Analista de Datos <b>RA16</b> Analista de Datos Geográficos <b>RA09</b> Arquitecto / Diseñador de ETL <b>RA10</b> Desarrollador de ETL
Salidas	Documento detallado de preparación de datos



UAECD

### 5.1.5. Validación de datos

T7 Tabla 7 Ciclo de Elaboración, fase de Validación de datos.  
Fuente: Elaboración Propia

Tarea	Validación de datos
Descripción	<p>La validación de los datos ocurre inmediatamente después de la preparación de los datos y antes del modelado. Este paso es necesario porque durante la preparación de datos existe una gran posibilidad de ocurrencia de errores, especialmente en escenarios complejos. La validación de datos garantiza que el modelado se realice en los datos correctos<sup>8</sup>.</p> <p><b>Causas comunes de errores:</b></p> <ul style="list-style-type: none"> <li>• Falta de comprensión adecuada de los datos y por tanto aplicación de lógica errónea durante la preparación.</li> <li>• Errores en los programas ejecutados para la preparación de los datos que generan salidas defectuosas.</li> <li>• Formatos de datos especiales que pueden generar errores durante la preparación y que no generan alertas de error por los programas utilizados para la preparación de datos.</li> </ul> <p><b>Ejemplo:</b> en algunas ocasiones al aplicar herramientas de geoprocetamiento como Dice <sup>9</sup> puede suceder que el área de los polígonos se vea afectada.</p>
Actividades	Validar los datos resultantes de la etapa de preparación. En caso de encontrar errores, retornar a la etapa de preparación de datos.
Roles	<p><b>RA13</b> Científico de Datos</p> <p><b>RA14</b> Investigadores especializados</p> <p><b>RA15</b> Analista de Datos</p> <p><b>RA16</b> Analista de Datos Geográficos</p> <p><b>RA18</b> Ingeniero de Datos</p> <p><b>RA09</b> Arquitecto / Diseñador de ETL</p> <p><b>RA10</b> Desarrollador de ETL</p>
Salidas	Documento de validación de los datos generados en el proceso de preparación.

<sup>8</sup> <https://sharing.luminis.eu/blog/the-forgotten-step-in-crisp-dm-and-asum-dm-methodologies/>

<sup>9</sup> <http://desktop.arcgis.com/en/arcmap/10.6/tools/data-management-toolbox/dice.htm>



## 5.1.6. Análisis

T8 Tabla 8 . Ciclo de Elaboración, fase de Análisis.  
Fuente: Elaboración Propia

Tarea	Análisis
Descripción	Construcción del modelo a través de múltiples iteraciones en donde se van refinando los diferentes parámetros o en ocasiones se determinan nuevos requerimientos previos para la preparación de los datos.
Actividades	<p>Seleccionar técnica del modelo según los datos disponibles, objetivos de análisis y requerimientos específicos de modelamiento. Consideraciones a tener en cuenta:</p> <ul style="list-style-type: none"> <li>• ¿Se requiere separación de datos entre pruebas y entrenamiento?</li> <li>• ¿Existen datos suficientes para tener resultados confiables del modelo?</li> <li>• ¿El modelo requiere algún nivel especial de calidad de datos?, ¿Los datos actuales cumplen con esos niveles de calidad?</li> <li>• ¿Los datos existentes son apropiados o se necesitan transformaciones adicionales?</li> </ul> <p>Generar diseño de la prueba: Generar mecanismo que permita evaluar la calidad y validez del modelo.</p> <p>Construir modelo: Ejecutar la herramienta del modelo seleccionado con los datos. Crear uno o más modelos. Para cada modelo determinar:</p> <ul style="list-style-type: none"> <li>• ¿Se obtuvieron conclusiones significativas?</li> <li>• ¿Se encontraron hallazgos (insights) relevantes o patrones inusuales?</li> <li>• ¿Existieron problemas durante el procesamiento? ¿El tiempo de procesamiento es razonable?</li> <li>• ¿El modelo tuvo problemas con la calidad de datos?</li> <li>• ¿Se encontraron inconsistencias significativas en los cálculos?</li> </ul> <p><b>Evaluar modelo:</b> Determinar cuál o cuáles modelos son acertados o efectivos para ser considerados finales.</p> <ul style="list-style-type: none"> <li>• Realizar evaluación integral de los modelos: <ul style="list-style-type: none"> <li>○ Efectividad</li> <li>○ ¿Los resultados son lógicos?</li> <li>○ ¿Los resultados son muy simples o triviales?</li> </ul> </li> <li>• Llevar registro de los parámetros utilizados para el afinamiento del modelo</li> </ul>



Roles	RA13 Científico de Datos RA14 Investigadores especializados
Salidas	Documento de análisis.

### 5.1.7. Prototipo

T9 Tabla 9 Ciclo de Elaboración, fase de Prototipo.  
Fuente: Elaboración Propia

Tarea	Prototipo
Descripción	Construir un prototipo para los casos cuando los requerimientos impliquen la disposición de un prototipo de aplicación adicional a los modelos seleccionados.
Actividades	Definir arquitectura del prototipo: Puede ser aplicaciones basadas en software comercial o software desarrollado a la medida. Diseñar prototipo: Definir flujo de trabajo del prototipo. Integrar o implementar prototipo: Ejem. Visualizaciones. Evaluar prototipo: Determinar si la aplicación cumple con los criterios de éxito de negocio.
Roles	RA12 Desarrollador de BI RA13 Científico de Datos RA04 Gerente de Datos y Analítica RA05 Analista de Negocio RA06 Arquitecto de Datos RA20 Arquitecto de Soluciones
Salidas	Documento de arquitectura y diseño del prototipo. Prototipo.



## 5.1.8. Validación y evaluación

T10 Tabla 10 Ciclo de Elaboración, fase de Validación y evaluación.  
Fuente: Elaboración Propia

Tarea	Validación y evaluación
Descripción	Determinar cuál o cuáles modelos son acertados o efectivos para ser considerados finales acordes a los criterios de éxito del negocio.
Actividades	<p><b>Evaluar resultados:</b> Evaluar los resultados según los criterios de éxito de negocio.</p> <ul style="list-style-type: none"> <li>Entender los resultados del proceso de análisis</li> <li>Interpretar los resultados en términos de su aplicación</li> <li>Verificar el efecto en los objetivos planteados para el análisis</li> <li>Verificar los resultados del análisis en comparación al conocimiento base de los hechos con el fin de evaluar si la información descubierta es novedosa y útil</li> <li>Evaluar los resultados respecto a los criterios de éxito de negocio y los objetivos de negocio originales.</li> <li>Determinar si hay nuevos objetivos de negocios para abordar después en este u otros proyectos.</li> <li>Generar recomendaciones sobre futuros proyectos de análisis.</li> <li>Validar y aprobar los modelos según los resultados de la validación.</li> </ul> <p><b>Revisión del proceso:</b></p> <ul style="list-style-type: none"> <li>Proveer un resumen del proceso realizado</li> <li>Generar recomendaciones sobre el proceso realizado, fallas y sugerencias de mejora.</li> </ul> <p><b>Determinar pasos siguientes</b></p> <ul style="list-style-type: none"> <li>Opción 1: Pasar al ciclo de publicación</li> <li>Opción 2: Volver a ejecutar el ciclo de análisis para refinamiento o reemplazo de los modelos.</li> </ul>
Roles	<p><b>RA04</b> Gerente de Datos y Analítica</p> <p><b>RA13</b> Científico de Datos</p> <p><b>RA14</b> Investigadores especializados</p> <p><b>RA20</b> Arquitecto de Soluciones</p>
Salidas	Documento de validación y evaluación.





## 5.1.9. Nueva adquisición de datos

T11 Tabla 11 . Ciclo de Elaboración, fase de Nueva adquisición de datos.  
Fuente: Elaboración Propia

Tarea	Nueva adquisición de datos
Descripción	<p>Identificar e incorporar nuevos datos que anteriormente no fueron tenidos en cuenta para que alimente el modelo y prototipo del ejercicio en cuestión.</p> <p>La Validación y evaluación del prototipo de analítica construido, puede arrojar la necesidad de incluir en el modelo del prototipo mismo, datos adicionales no tenidos en cuenta anteriormente, los cuales deberían ayudar a responder el requerimiento analítico del negocio.</p>
Actividades	<p>Listar los hallazgos más significativos en la validación y evaluación de la fase preliminar.</p> <ul style="list-style-type: none"> <li>• Consolidar el resultado de la validación</li> <li>• Consolidar el resultado de la evaluación</li> <li>• Identificar variables adicionales accesibles, que sean requeridas para el mejoramiento o precisión del modelo</li> </ul> <p>Aplicar las demás fases del ciclo de elaboración</p> <ul style="list-style-type: none"> <li>• Ejecutar la fase de “Entendimiento del negocio” y/o de “Descubrimiento y entendimiento de datos” en la medida de que sean requeridas para las nuevas variables incorporadas</li> <li>• Ejecutar la fase de “Preparación y gestión de datos”, junto con sus subfases</li> <li>• Ejecutar la fase de “Análisis”</li> <li>• Ejecutar la fase de “Prototipo”</li> <li>• Ejecutar la fase de “Validación y evaluación”</li> </ul> <p>Revisión del proceso</p> <ul style="list-style-type: none"> <li>• Proveer un resumen del proceso realizado</li> <li>• Generar recomendaciones sobre el proceso realizado, fallas y sugerencias de mejora.</li> </ul> <p>Determinar pasos siguientes</p> <ul style="list-style-type: none"> <li>• Opción 1: Pasar al ciclo de Entendimiento del negocio</li> <li>• Opción 2: Pasar al ciclo de Descubrimiento y entendimiento de datos</li> </ul>
Roles	<p>RA06 Arquitecto de Datos</p> <p>RA07 Administrador de Datos</p> <p>RA08 Administrador de Base de Datos</p> <p>RA11 Coordinador de Metadatos</p> <p>RA13 Científico de Datos</p> <p>RA15 Analista de Datos</p>



UAECD

	<b>RA16</b> Analista de Datos Geográficos <b>RA03</b> Usuarios de Negocio <b>RA05</b> Analista de Negocio
Salidas	Documento de Nueva adquisición de datos.

## 5.2. Ciclo de Publicación

### 5.2.1. Publicación

T12 Tabla 12 Ciclo de Publicación, fase de Publicación.  
Fuente: Elaboración Propia

Tarea	Publicación
Descripción	Publicar el modelo final para ser utilizado por los usuarios y preparar su operación continua.
Actividades	<p>Realizar Transferencia de Conocimiento Operacional sobre la solución. Preparar para el mantenimiento subsecuente:</p> <ul style="list-style-type: none"> <li>• Ha de asegurar que todas las funciones y actividades previas están en su lugar antes de la publicación.</li> <li>• Determinar actividades de mantenimiento</li> <li>• Determinar actividades de monitoreo</li> </ul> <p>Publicar el modelo final en el entorno de producción.</p> <ul style="list-style-type: none"> <li>• Cargar datos para producción</li> <li>• Probar que el entorno de producción funciona correctamente</li> <li>• Migrar modelo a producción</li> <li>• Habilitar (lanzar) el acceso al modelo de producción</li> <li>• Comunicar detalles de la solución implementada a los usuarios</li> <li>• Revisar los detalles del lanzamiento con el fin de aprender de este para futuras publicaciones</li> </ul> <p>Cerrar proyecto de analítica</p> <ul style="list-style-type: none"> <li>• Crear reporte final</li> <li>• Reunión de cierre del proyecto</li> <li>• Ejecutar acciones encontradas en la reunión de cierre</li> <li>• Lecciones aprendidas</li> <li>• Cierre del proyecto</li> </ul>
Roles	<b>RA06</b> Arquitecto de Datos <b>RA13</b> Científico de Datos <b>RA12</b> Desarrollador de BI <b>RA17</b> Desarrollador Web



UAECD

Salidas	Documento resumen de lanzamiento de publicación del modelo final. Reporte final del proyecto de analítica (Resumen, recomendaciones, oportunidades).
---------	---

### 5.2.2. Publicar Hallazgos

T13 Tabla 13 . Ciclo de Publicación, fase de Publicar hallazgos.  
Fuente: Elaboración Propia

Tarea	Publicar Hallazgos
Descripción	Transferencia de conocimiento al usuario final de los hallazgos encontrados durante el proyecto analítico.
Actividades	Determinar los hallazgos generados por el modelo. Orientar y educar al usuario final con respecto a la solución del modelo mostrando su eficacia y eficiencia.
Roles	<b>RA06</b> Arquitecto de Datos <b>RA03</b> Usuarios de Negocio <b>RA15</b> Analista de Datos <b>RA13</b> Científico de Datos <b>RA04</b> Gerente de Datos y Analítica
Salidas	Documento con hallazgos finales

### 5.2.3. Monitorear

T14 Tabla 14 Ciclo de Publicación, fase de Monitorear.  
Fuente: Elaboración Propia

Tarea	Monitorear
Descripción	Monitorear los resultados de los modelos implementados de forma continua para garantizar su precisión y que aún cumplan con los objetivos de extracción de datos de la organización y los objetivos de negocio
Actividades	Monitorear resultados Determinar necesidades de ajuste del modelo
Roles	<b>RA21</b> Administrador de tecnología <b>RA03</b> Usuarios de Negocio <b>RA19</b> Arquitecto Empresarial



Salidas	Reporte de alertas generadas a partir del monitoreo.
---------	--

5.2.4. Operar

T15 Tabla 15 Ciclo de Publicación, fase de Monitorear.  
Fuente: Elaboración Propia

Tarea	Operar el sistema
Descripción	Desarrollar las actividades necesarias para garantizar la operación del sistema.
Actividades	Operar sistema (Em: Automatizar tareas rutinarias, gestión de errores y logs, monitorear, generar y gestionar incidencias. Etc) Monitorear sistema en producción (Ejm: logs, uso de cómputo y almacenamiento, desempeño, etc) Gestionar tareas de mantenimiento (Ejm: ventanas de mantenimiento, backups, archivamiento de datos, disaster recovery, etc) Gestionar incidencias Gestionar mejoras. Optimizar sistema Mantener la documentación. Ejecutar tareas de gestión del conocimiento (Ejm. Capacitar usuarios y administradores) Revisar los procedimientos de mantenimiento para el ambiente de producción y generar un cronograma para las actividades de mantenimiento más críticas Identificar señales de alerta y problemas comunes.
Roles	RA07 Administrador de Datos RA08 Administrador de Base de Datos RA21 Administrador de tecnología RA03 Usuarios de Negocio
Salidas	Reporte de operación



### 5.2.5. Soporte

**T16** Tabla 16 Ciclo de Publicación, fase de Soporte.  
Fuente: Elaboración Propia

Tarea	Soporte del sistema
Descripción	Desarrollar las actividades necesarias para garantizar el debido soporte a los usuarios
Actividades	<ul style="list-style-type: none"> <li>• Proveer punto de contacto y horarios disponibles para llamadas de soporte al usuario.</li> <li>• Gestionar incidentes.</li> <li>• Gestionar mesa de ayuda.</li> <li>• Gestionar las comunicaciones con los usuarios</li> </ul>
Roles	<b>RA21</b> Administrador de tecnología <b>RA07</b> Administrador de Datos <b>RA08</b> Administrador de Base de Datos
Salidas	Reporte de soporte

### 5.2.6. Gestión de Infraestructura

**T17** Tabla 17 Ciclo de Publicación, fase de Gestión de Infraestructura.  
Fuente: Elaboración Propia

Tarea	Gestión de Infraestructura
Descripción	Realizar las actividades necesarias para garantizar el mantenimiento de la infraestructura
Actividades	<ul style="list-style-type: none"> <li>• Gestionar configuración</li> <li>• Gestionar capacidad</li> <li>• Gestionar disponibilidad del sistema</li> <li>• Gestionar publicación de versiones en producción</li> <li>• Gestionar seguridad</li> </ul>
Roles	<b>RA21</b> Administrador de tecnología <b>RA07</b> Administrador de Datos <b>RA08</b> Administrador de Base de Datos
Salidas	Reporte de gestión de infraestructura



### 5.2.7. Integración de aplicaciones

**T18** Tabla 18 Ciclo de Publicación, fase de Integración de aplicaciones.  
Fuente: Elaboración Propia

Tarea	Integración de aplicaciones
Descripción	Para los casos en que aplique, integrar la solución analítica con otros sistemas a través de un esquema orientado a servicios
Actividades	Entender los requerimientos de negocio. Realizar el mapeo de los procesos en componentes. Derivar requerimientos específicos Producir la arquitectura de integración. Planear y ejecutar la integración <sup>10</sup>
Roles	<b>RA21</b> Administrador de tecnología <b>RA20</b> Arquitecto de Soluciones <b>RA09</b> Arquitecto / Diseñador de ETL <b>RA10</b> Desarrollador de ETL
Salidas	Reporte de integración de aplicaciones

### 5.2.8. Gestión del ciclo de vida y mejora continua

**T19** Tabla 19 Ciclo de Publicación, fase de Gestión del ciclo de vida y mejora continua.  
Fuente: Elaboración Propia

Tarea	Gestión del Ciclo de Vida y mejora continua
Descripción	Realizar las actividades necesarias para manejar el ciclo de vida de la solución.

<sup>10</sup> [https://www.researchgate.net/publication/220306398\\_An\\_enterprise\\_application\\_integration\\_methodology\\_for\\_e-government](https://www.researchgate.net/publication/220306398_An_enterprise_application_integration_methodology_for_e-government)



Actividades	Monitorear, medir y verificar los beneficios ofrecidos por la solución Evaluar el desempeño de la solución Revisar el plan de ciclo de vida del sistema Monitorear y mantener la calidad de servicio de la solución. Ejem: Auditoría de calidad de datos, gestión de calidad de datos, encuestas a usuarios, gestión de calidad de la documentación, gestión de calidad del proceso de gestión de datos.
Roles	RA08 Administrador de Base de Datos RA02 Líder de Negocio RA04 Gerente de Datos y Analítica RA05 Analista de Negocio RA19 Arquitecto Empresarial
Salidas	Reporte de gestión de ciclo de vida.



---

## Conclusiones

- ✓ Definir una metodología para hacer ejercicios asociados con la analítica de datos (soluciones de analítica) aporta significativamente al éxito de un proyecto, debido a que los elementos involucrados en el ciclo de vida de los ejercicios de analítica mantienen unas directrices, reglas y lineamientos establecidos para cumplir con un propósito en particular, la metodología de analítica permite establecer un plan de trabajo adecuado.
- ✓ La metodología de ASUM DM es la evolución de CRISP DM, y surge como resultado de la compilación y estructuración de una serie de aplicaciones de analítica de negocios e inteligencia de negocios, lo cual, implica una organización/estructuración adecuada de sus fases y sus elementos involucrados.





---

## Referencias

- Angée, S., Lozano Argel, S., Montoya Munera, E., Ospina Arango, J., & Tabares Betancur, M. (2018). Towards an Improved ASUM-DM Process Methodology for Cross-Disciplinary Multi-organization Big Data & Analytics Projects. *Communications in Computer and Information Science*. doi:[https://doi.org/10.1007/978-3-319-95204-8\\_51](https://doi.org/10.1007/978-3-319-95204-8_51)
- Bahrepour, M. (2018). *luminis.eu*. Obtenido de <https://sharing.luminis.eu/blog/the-forgotten-step-in-crisp-dm-and-asum-dm-methodologies/>
- CAOBA - Universidad de Los Andes. (2017). *Perfil Alianza Caoba*. Bogotá: Universidad de Los Andes.
- Esri. (2019). *arcgis.com*. Obtenido de <http://desktop.arcgis.com/en/arcmap/10.6/tools/data-management-toolbox/dice.htm>
- IBM Corporation. (2015). *icesi.edu.co*. Obtenido de [http://gforge.icesi.edu.co/ASUM-DM\\_External/index.htm](http://gforge.icesi.edu.co/ASUM-DM_External/index.htm)
- IBM Corporation. (2016). *Analytics Solutions Unified Method. Implementations with agile principles*. New York: IBM Analytics. Recuperado el 04 de 2019, de <ftp://ftp.software.ibm.com/software/data/sw-library/services/ASUM.pdf>
- Janssen, M., & Cresswell, A. (2005). An enterprise application integration methodology fore-government. *Journal of Enterprise Information Management*. doi:10.1108/17410390510623990
- kdnuggets. (2017). *www.kdnuggets.com*. Obtenido de <https://www.kdnuggets.com/2017/02/analytics-grease-monkeys.html>





[www.  
ideca.  
gov.  
co](http://www.ideca.gov.co)

Latitud: 4.603557, Longitud: -74.094105  
Bogotá, Cundinamarca, Colombia.