

Analítica de Datos

Profesor: Jose Aguilar

Sitio de Trabajo: CIDITICs-Yachay EP
Lab. Robótica y Sistemas Inteligentes, EPN

Contacto: aguilar@ula.ve,

Información del curso: www.ing.ula.ve/~aguilar

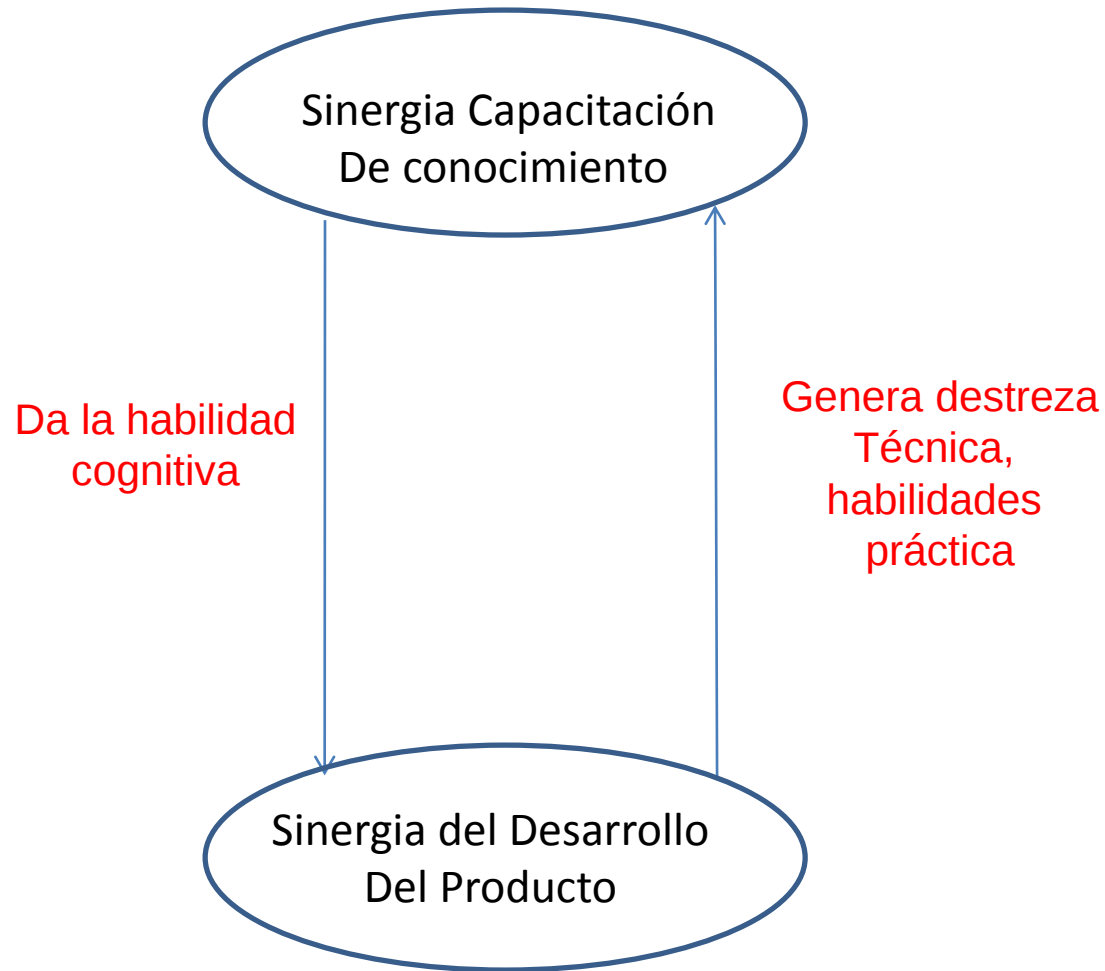
OBJETIVO

Introducir a la Analítica de Datos y a la importancia en las organizaciones

OBJETIVOS ESPECIFICOS

- **Establecer una cartera de proyectos en Analítica de Datos**
- **Introducir una metodología para concebir proyectos en Analítica de Datos**
- **Establecer alianzas de trabajo en Analítica de Datos**

Dinámicas del Curso



Dinámicas Sinergias

- Nos moveremos en las dos esferas
- Se introducirán los conceptos de base
- ¿Cómo se trabaja?:
 - Todo el material en línea, internet
 - Aprender hacienda, aprendizaje activo
 - Espacios de discusión y debate

Sinergia Desarrollo de Productos (SDP)

- Se trabajará en productos,
- Los productos (obras) irán plasmando el conocimiento adquirido
- Los productos (obras) al final deberán contener todo el contenido adquirido en el curso inmerso en ellos

Dinámica del curso

Día	Horario	Actividades
18.08.16	9H50-10H	Bienvenida
18.08.16	10H-11H	- Introducción a la Analítica de Datos - Introducción a la Metodología MIDANO
18.08.16	11H15-13H30	- Descripción de la organización - Caracterización detallada de los procesos de la organización
18.08.16	14H30-16H	- Análisis de Procesos Organizacionales - Introducción a la especificación de Tareas de Analítica de Datos
18.08.16	16H15-18H	- Análisis de factibilidad de los procesos, para extraer y explotar los datos como conocimiento - Selección de los procesos organizacionales de interés, para realizar tareas de Analítica de Datos
19.08.16	9H15-10H45	- Introducción a la Prospectiva Organizacional
19.08.16	11H-13H	- Definición de las posibles tareas de Analítica de Datos en esos procesos - Realización de un ejercicio de prospección tecnológica sobre dichas tareas de Analítica de Datos identificadas
19.08.16	14H-16H	- Definición de los posibles pasos en conjunto con CIDITICs
	16H-16H20	Finalización del Curso. Despedida

Introducción a la Analítica de Datos

Ideas introductorias



Según Steve Haeckel

“El **cociente de inteligencia** de una empresa está determinado por la medida en que su **infraestructura informática conecta la información, la comparte y le da estructura.**”

Ideas introductorias



SISTEMA NERVIOSO DIGITAL

Proporciona a sus usuarios una
profunda comprensión y una
capacidad para aprender, que
no podrían conseguir por otros
medios

Resemantizar a las Organizaciones



Sin fronteras organizacionales

Transferir parte de nuestros costos a nuestros proveedores y usuarios

Fidelidad de nuestros usuarios



Identificar las oportunidades de negocios

Estrategias de aplicación de Tecnología de Información

Análisis de la Cadena de Valor

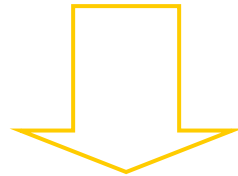
La sociedad del Conocimiento



- Ausencia de fronteras, porque el conocimiento viaja aun con menos esfuerzo que el dinero
- Disponible para todos, en virtud de que la auto-formación cada día es más fácil de adquirir.
- La mayoría de los empleados cada día serán menos de tiempo completo para la organización.
- Nacimiento de nuevas instituciones teorías, problemas, a un ritmo vertiginoso.

CONOCIMIENTO

“En los últimos 10 años se han producido más conocimientos que en los 10.000 años anteriores”.



Bill Gates

**Estamos en la Civilización
del Conocimiento**

CONOCIMIENTO

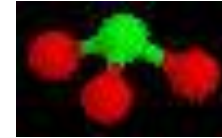
ERA INDUSTRIAL



VALORES PREDOMINANTES

- Poder
- Control
- Disciplina
- Especialización
- Estructura jerarquizada

ERA DEL CONOCIMIENTO



VALORES PREDOMINANTES

- Descentralización
- Información
- Innovación
- Calidad
- Trabajo en equipo

Los trabajadores trabajan más con sus mentes que con sus manos (Knowledge Worker)



SOCIEDAD DEL USO DE CONOCIMIENTO

Embudo del Conocimiento



GESTIÓN DEL CONOCIMIENTO

**"EL CONOCIMIENTO ES EL ACTIVO
INTANGIBLE QUE MAYOR
COMPETITIVIDAD GENERA A LAS
NACIONES Y A LAS
ORGANIZACIONES, EN LA
ECONOMÍA GLOBAL"**

"DEBE SER GERENCIADO"

CONOCIMIENTO ORGANIZACIONAL

- **EXPLÍCITO:** Referencial, replicable, entrenable, reposa en textos, bases de datos, objetos, redes, como información.
- **TÁCITO:** Vivencial, basado en valores, creencias, actitudes, sentimientos humanos, no es fácil de replicar.

CONOCIMIENTO ORGANIZACIONAL

Conocimiento documentado

- Artículos
- Propuestas
- Presentaciones
- Políticas de la organización

Conocimiento explícito no documentado

- Información de proyectos
- Información de clientes
- Políticas de la organización
- Experiencia de los empleados
- Procedimientos para las reuniones
- Procedimientos para la contratación
- Mejores prácticas para propuestas

Conocimiento tácito no documentado

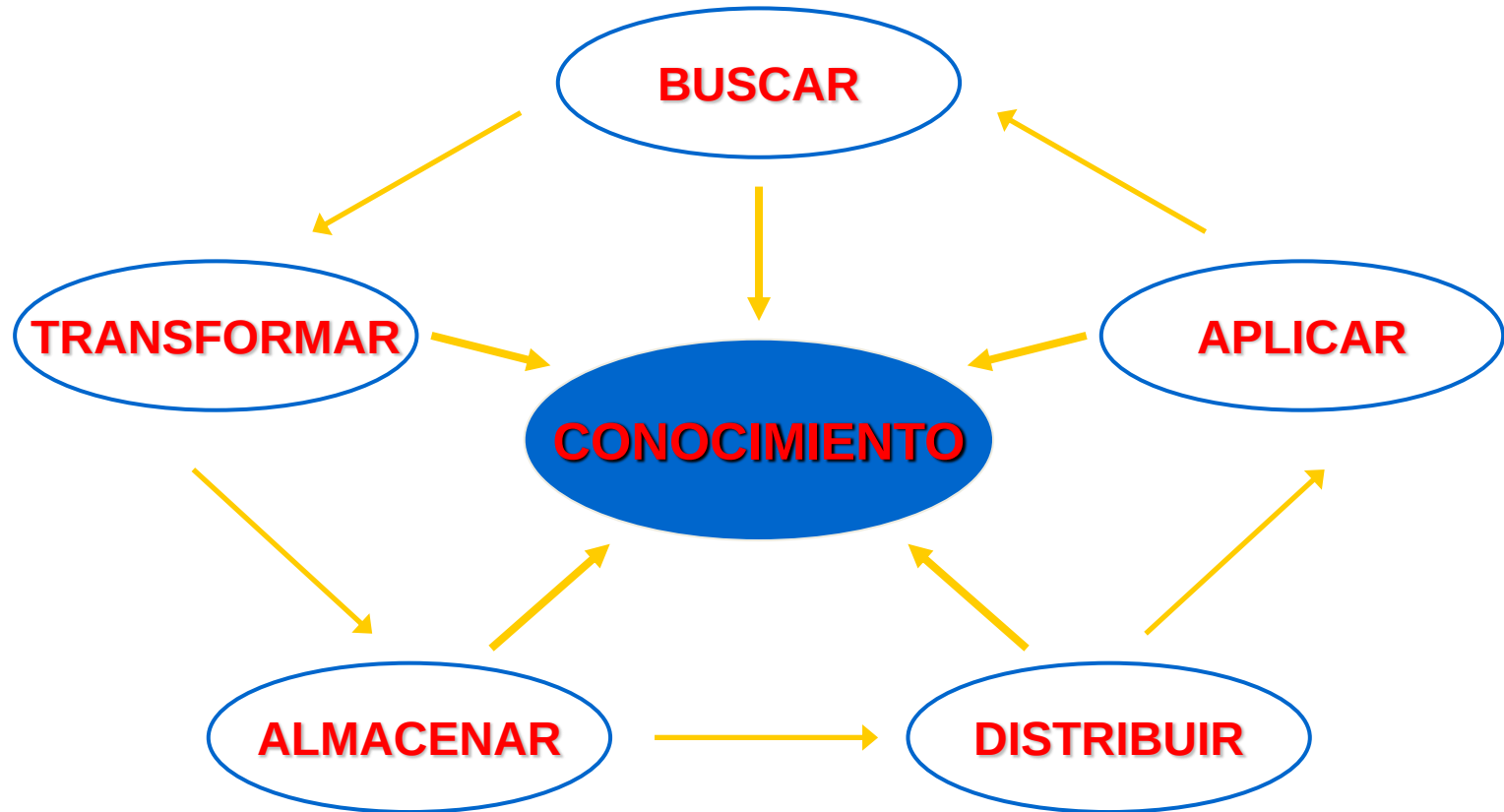
- Mejores prácticas para proyectos
- Mejores prácticas para atención a clientes

Fácil

Posible

Difícil

CONOCIMIENTO ORGANIZACIONAL



Rol del Conocimiento en una ORGANIZACION

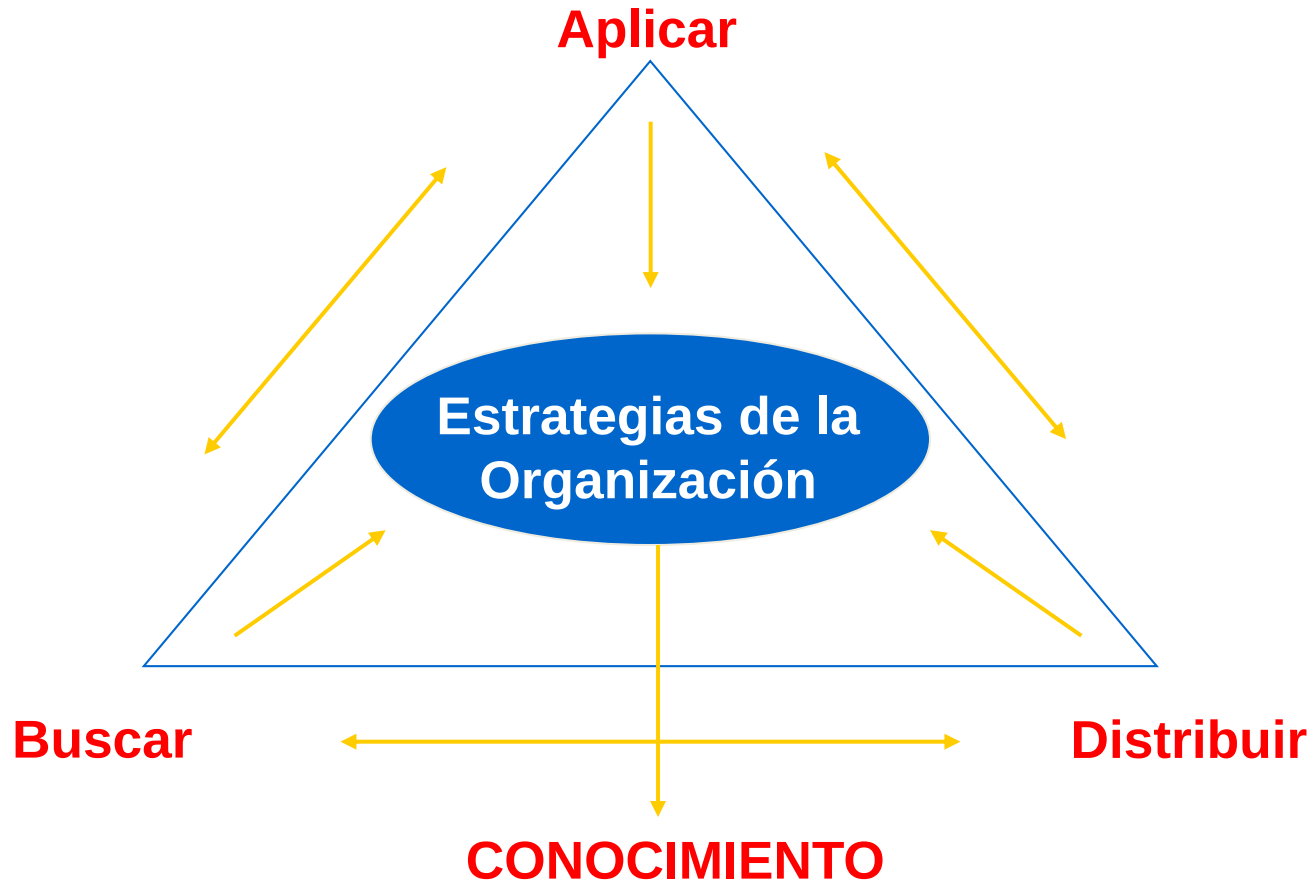
El único progreso sostenible que posee una organización viene de:

1. Lo que colectivamente sabe
2. Cuan eficientemente utiliza lo que sabe
3. Que tan rápido adquiere y utiliza nueva información

(Drucker, 2011; Bart, 2015)



CONOCIMIENTO ORGANIZACIONAL



Gestión del Conocimiento

La gestión del conocimiento es **crear la posibilidad para todo el mundo de ver exactamente qué está pasando, dónde vamos bien.**

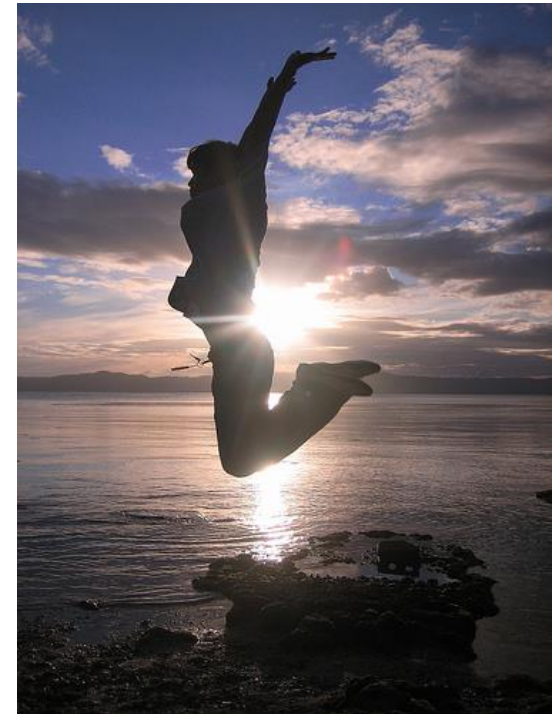
Bill Gates, presidente de Microsoft

La gestión del conocimiento es la capacidad de la organización para **crear nuevos conocimientos, diseminarlos y encapsularlos en productos, servicios y sistemas.**

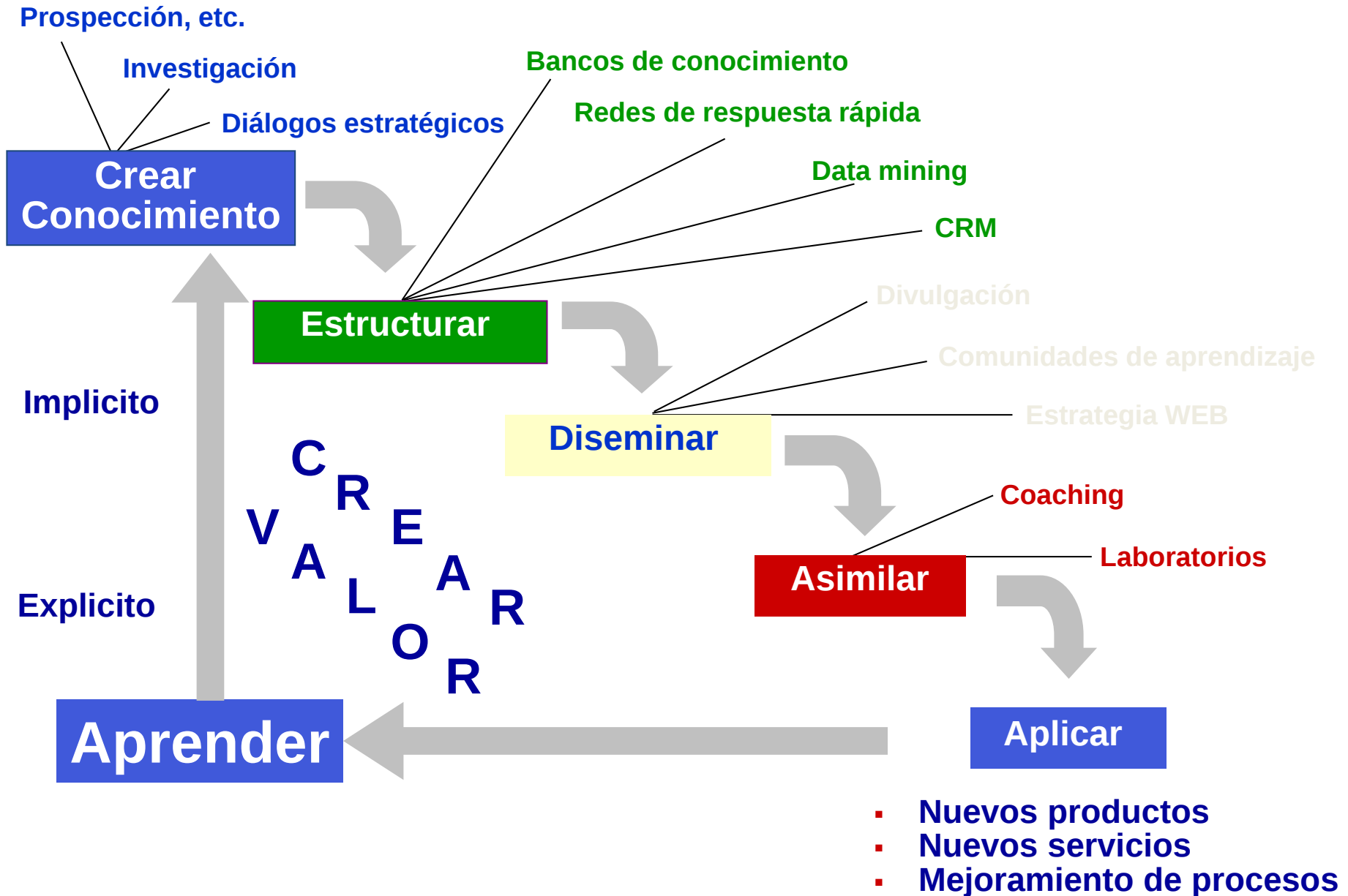
Ángel L. Arbonies, presidente de Cluster del Conocimiento

Atributos Claves de la Gestión del Conocimiento

- Generar nuevos Conocimientos
- Acceder conocimiento de fuentes externas
- Usar dicho conocimiento en la toma de decisiones
- Arraigar el conocimiento en procesos productos y servicios
- Representar el conocimiento en documentos, programas de computación, y base de datos
- Facilitar el crecimiento del conocimiento a través de cultura y otros incentivos
- Transferir conocimiento existente a otras partes de la organización
- Medir el valor de los activos del conocimiento y/o su impacto en el administración del conocimiento

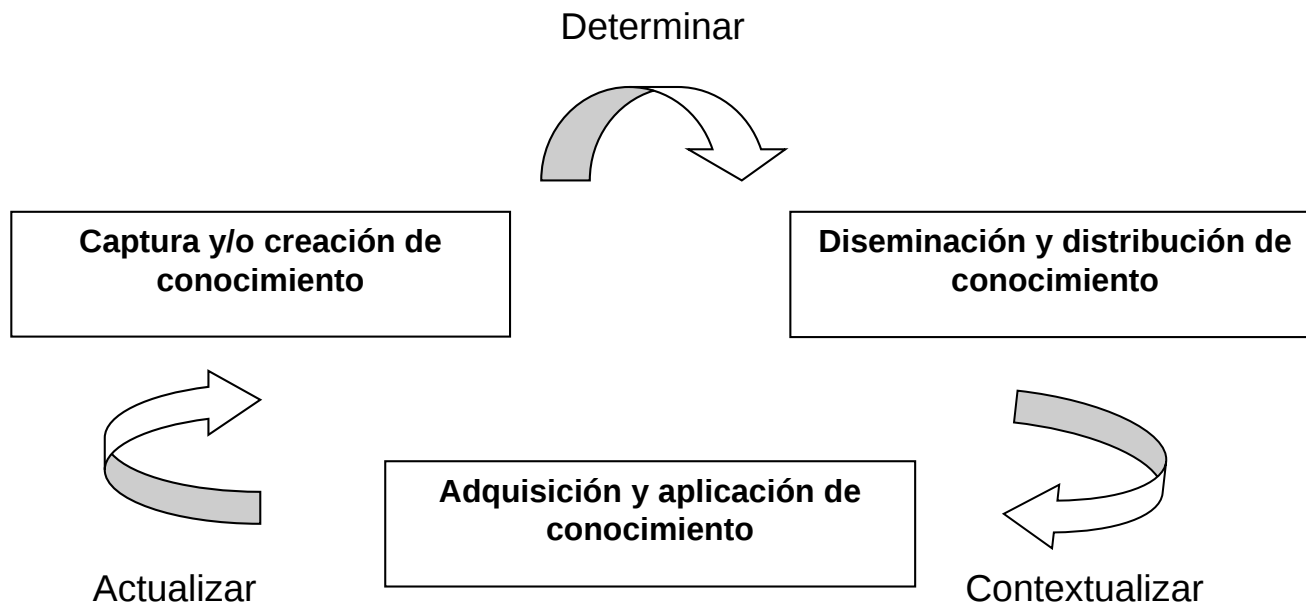


Modelo de Gerencia del Conocimiento Skandia



Ciclo de la Gestión del Conocimiento

1. Captura y/o creación de conocimiento
2. Diseminación y distribución de conocimiento
3. Adquisición y aplicación de conocimiento

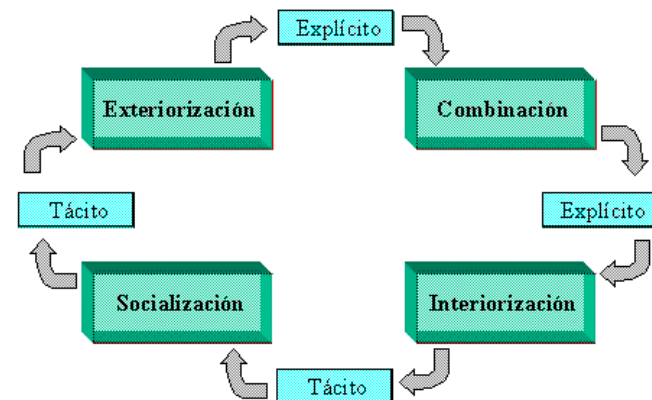


Modelo Espiral de Nonaka y Takeuchi



Los 4 modos de conversión de conocimiento entre el *Conocimiento Tácito* (CT) y el *Conocimiento Explícito* (CE)

1. de CT a CT = Socialización
2. de CT a CE = Externalización
3. de CE a CE = Combinación
4. de CE a CT = Internalización



Administración Inteligente

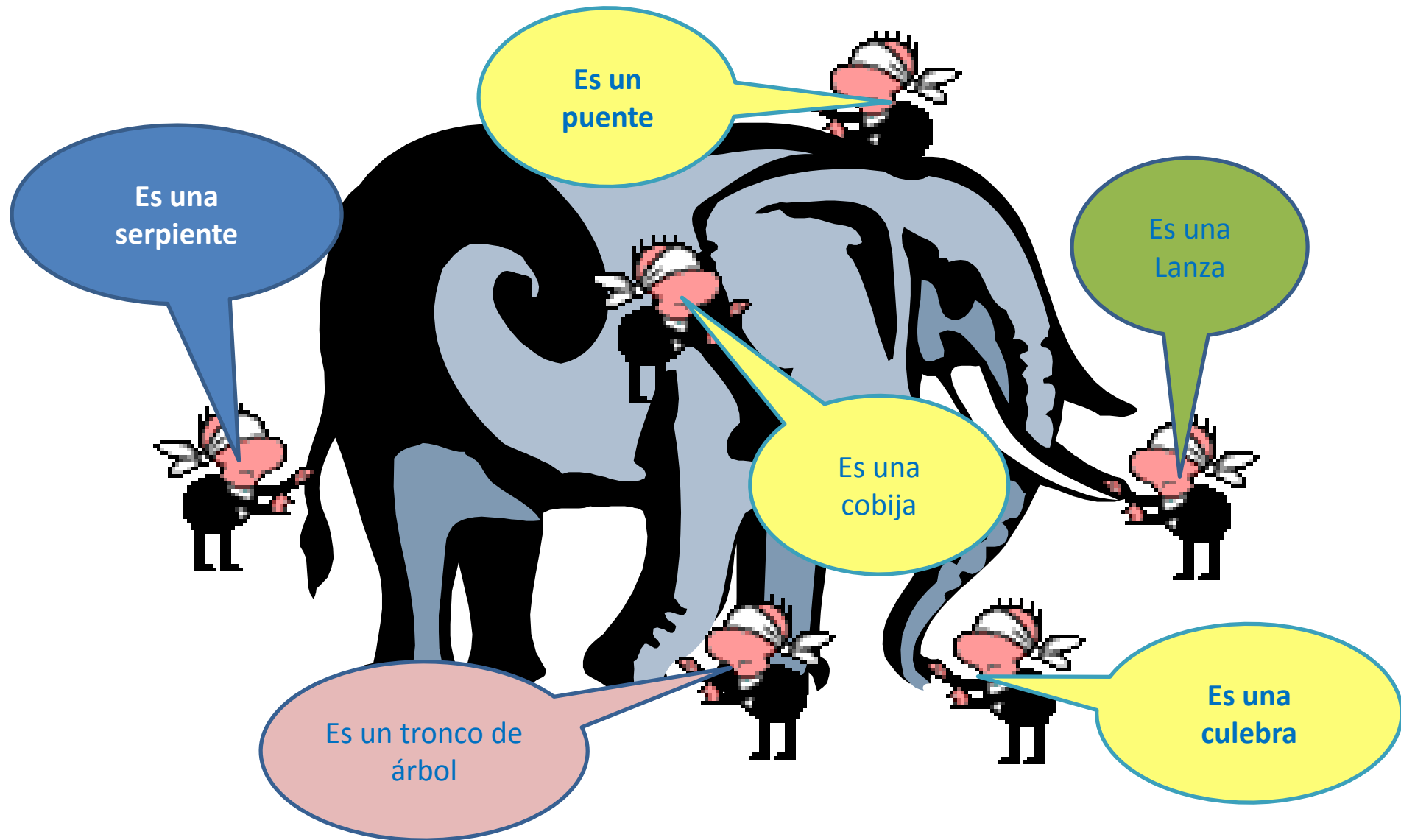
No se puede administrar lo que no se puede ver

La capacidad de ver toda la organización es el aspecto más importante de la **administración inteligente**.

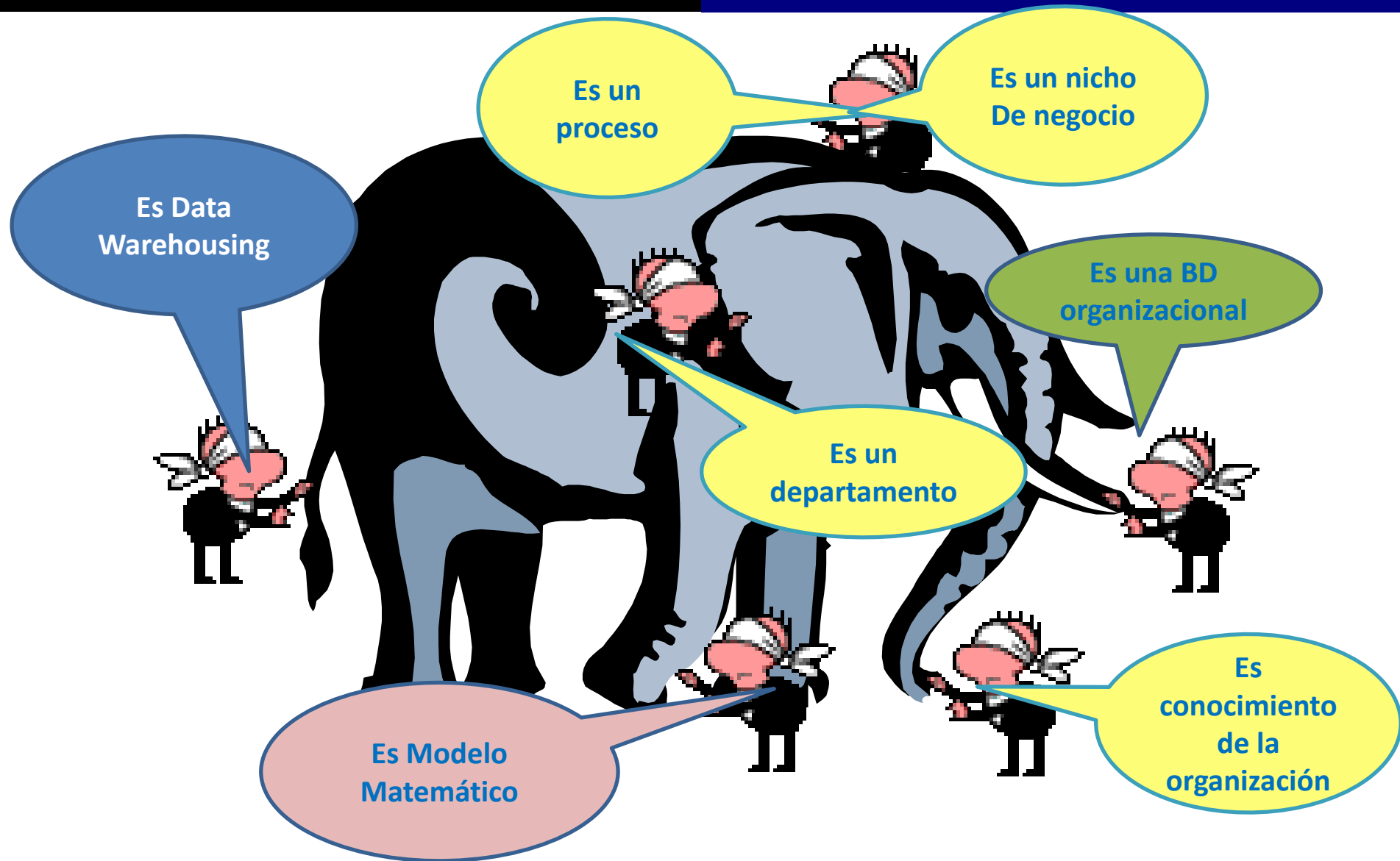
Las organizaciones tienen que ser **flexibles y adaptables** a cambios en el proceso y modelo de los negocios, así como también a nuevas tecnologías.



Un ciego describiendo un Elefante



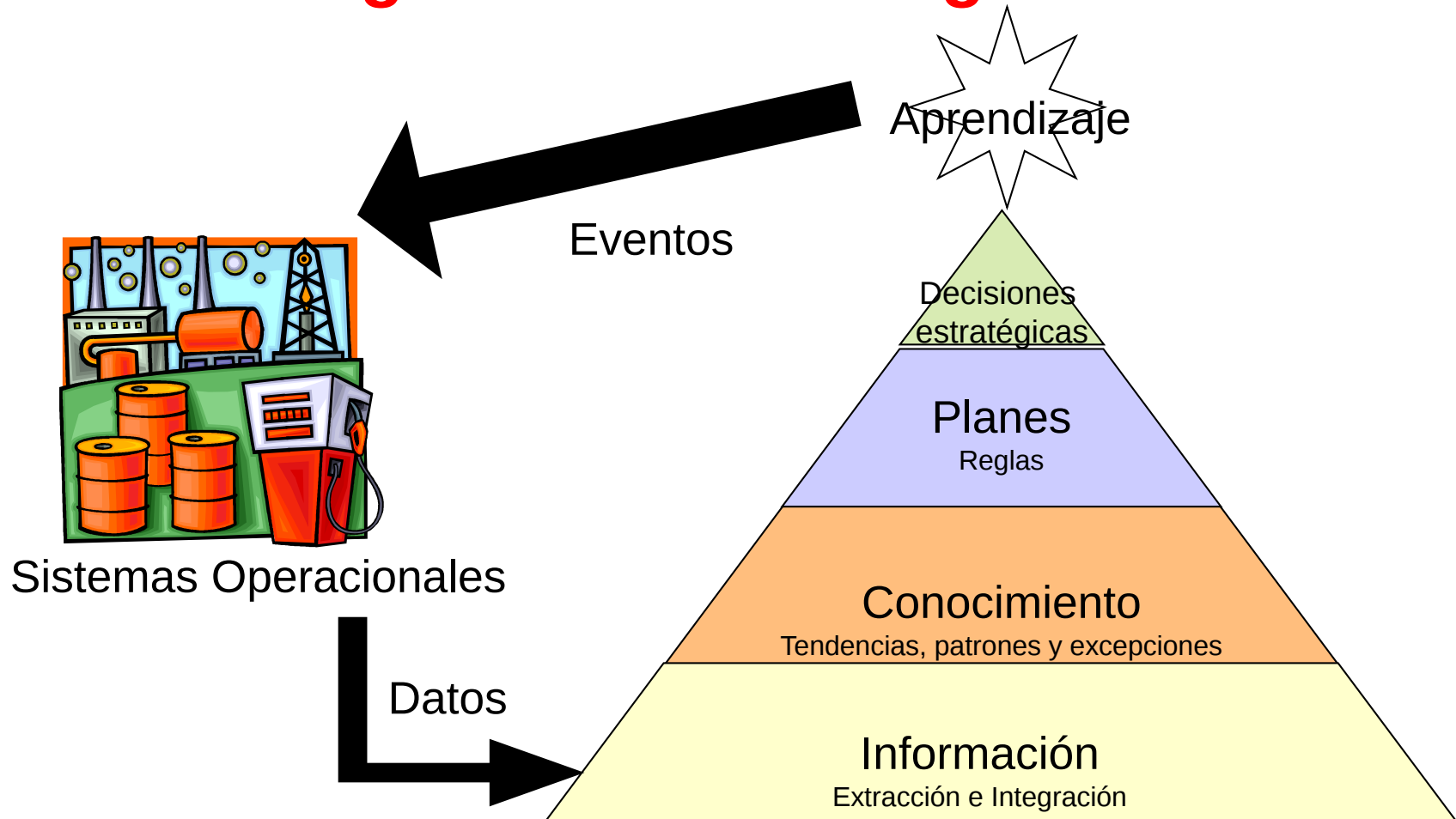
Un gerente describiendo su organización Sin AD



Dato, información, conocimiento e inteligencia en una organización



Dato, información, conocimiento e inteligencia en una organización

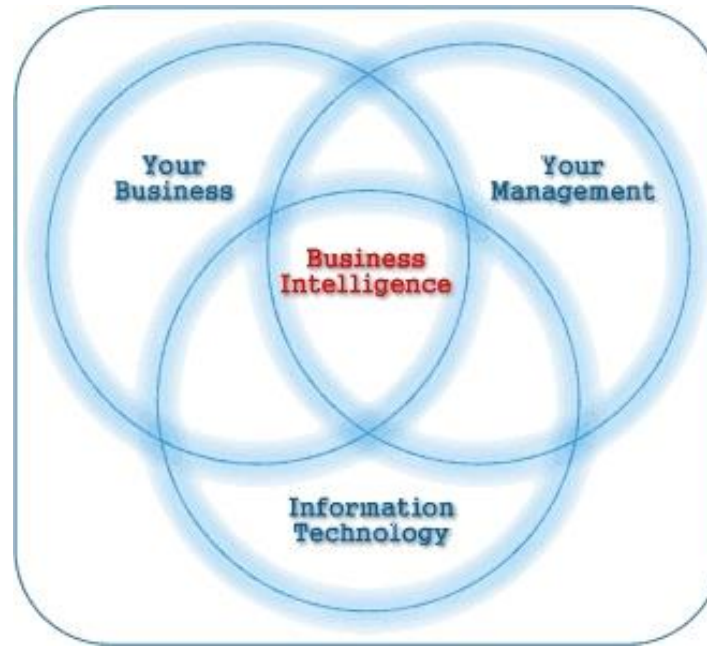


¿Qué es Inteligencia de Negocios?



*Conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el **análisis de datos existentes en una organización.**

*Abarca la comprensión del funcionamiento actual de la organización, y la anticipación de acontecimientos futuros, con el objetivo de **ofrecer conocimientos para respaldar las decisiones organizacionales.**

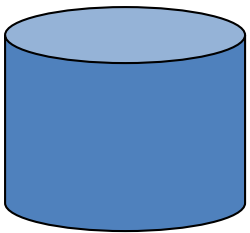


¿Qué es Inteligencia de Negocios?



Hace a las organizaciones más inteligentes

Agregar Datos



Database, Data Mart, Data Warehouse, ETL Tools,



Presentar Datos



Dashboards,, OLAP Cubes



Enriquecer Datos



Informar Decisiones



Analítica de Datos

- Con las grandes cantidades de datos disponibles actualmente, las organizaciones se centran en la explotación de los datos para obtener una ventaja competitiva.
- En el pasado, las organizaciones podrían emplear equipos de estadísticos y analistas para explorar conjuntos de datos de forma manual, pero el volumen y la variedad de datos han superado con creces la capacidad de análisis manual.
- Las computadoras son más poderosas, el trabajo en red es omnipresente, y se han desarrollado algoritmos que pueden conectar conjuntos de datos para permitir análisis más amplios y profundos que antes era imposible.

La convergencia de estos fenómenos ha dado lugar a la aplicación generalizada cada vez mayor de los principios de la analítica de datos.

- La AD se utiliza en general para analizar comportamientos de clientes, de empleados, etc.
- La industria financiera utiliza la AD para la detección del fraude, administración de personal, para decidir créditos, etc.
- Walmart, Amazon, etc. usan la AD en sus negocios: marketing, en la gestión de la cadena de gestión.

Los objetivos principales de AD son ayudar a *ver los problemas de negocio desde una perspectiva de los datos*, y comprender los principios de *extracción de conocimiento útil a partir de los datos*.

Analítica de Datos

Análisis de los datos **es la ciencia** de la recogida, almacenamiento, extracción, limpieza, transformación, agregación y análisis de datos, **con el fin de descubrir información y conocimiento.**

- Analítica utiliza modelos descriptivos, de identificación, predictivos, de optimización, prescriptivo, etc., con el fin de **producir conocimiento a partir de datos**, que se utilizará para guiar la **toma de decisiones.**
- El **alto grado de datificación** incrustado en la sociedad exige nuevas herramientas y mecanismos para la manipulación y la representación de los datos que facilitan la **extracción de conocimiento significativo.**

Analítica de Datos



Los datos son el nuevo petróleo de la economía



Es la ciencia que examina datos en bruto con el propósito de buscar conocimiento, sacar conclusiones, generar información, entre otras cosas.

Es usado en muchos ámbitos:

- La industria para tomar mejores decisiones empresariales
- Las ciencias para verificar o reprobando modelos o teorías existentes.
- ...

Análisis de los datos no es simplemente la minería de datos

La minería de datos navega a través de grandes conjuntos de datos utilizando un software sofisticado para **identificar patrones**.



El análisis de datos se centra en la inferencia para sacar una conclusión sobre la base de lo que se conoce, para descubrir relaciones ocultas y establecer nuevas

Análisis de los datos no es simplemente la minería de datos

Datos disponibles para un agricultor:

1. Los patrones climáticos históricos
2. Los datos de cultivo de plantas y la productividad de cada cepa
3. Las especificaciones de los Fertilizantes
4. Las especificaciones de los Plaguicidas
5. Los datos de productividad del suelo
6. Los datos del ciclo de plagas
7. Los costos, la fiabilidad, etc., de las máquinas
8. Los datos de conducción del agua
9. Los datos históricos de oferta y demanda
10. Los precios del Mercado

Análisis de los datos no es simplemente la minería de datos

Analítica Datos es la aplicación de un procedimiento algorítmico para obtener conocimiento, desde un conjuntos de datos.

- En cuanto a los datos del tiempo y plagas, se podrían establecer correlaciones entre un cierto tipo de hongo cuando el nivel de humedad alcanza un cierto punto.
- Si las futuras proyecciones meteorológicas para los próximos meses predicen un bajo nivel de humedad, **por lo tanto, habrá** riesgo bajo de ese hongo.

Para el agricultor, esto podría significar ser capaz de plantar **un determinado tipo de producto agrícola no resistente a ese hongo**, con un mayor rendimiento y precio en el mercado, **sin tener que comprar un determinado fungicida.**

Todos los datos se usan para responder a:

- ¿Cuál es el patrón de siembra agrícola para obtener el mejor precio?
- ¿cuáles son los productos agrícolas con mejor rendimiento?

Análisis de los datos no es simplemente la minería de datos

Análisis de datos: cualquier intento de dar sentido a los datos.

Actividad heurística, donde se realiza una exploración de todos los datos, para ganar conocimiento sobre el negocio

Es la ciencia o el proceso de análisis de datos para **sacar conclusiones** acerca de lo que está pasando en el proceso, evento, etc. que los datos están describiendo.

Minería de datos: se refiere a la ciencia de la recopilación de todos los datos del pasado, y luego la **búsqueda de patrones** en ellos, previamente desconocidos.

Una vez que se encuentren, **se validan** mediante la aplicación de los patrones detectados a nuevos subconjuntos de datos.

Analítica de Datos

- Enfoques de des-síntesis de los datos e información para responder a preguntas
- Método de crear datos, cifras (conocimiento) para resolver problemas
- Proceso sistemático de utilizar los datos para hacer frente a preguntas
- Desglose de los temas a través de la utilización de los datos y la información conocida

Analítica de Datos

Utilizando datos para **aumentar el valor del negocio**

Dato =

- Grande y pequeño
- Interno y externo
- Estructurado y no estructurado
- Tradicional y "Nuevo"
- "Libre" y comprado

Analítica de Datos

Los datos pueden "hablar"

El análisis de datos contiene aspectos del razonamiento científico:

- Define
- Interpreta
- Evalúa
- Ilustra
- Discute
- Explica
- Clarifica
- Compara
- Contrasta

Guía básica para el análisis de datos:

- Analizar es “NO” narrar
- Descomponer los objetivos en preguntas de investigación
- Identificar los fenómenos que han de investigarse
- Visualizar las respuestas "esperadas"
- Validar las respuestas con los datos
- Afirmar soportado por datos

- Con la AdD:
 - Se es objetivo
 - Se es preciso
 - Se es veraz
- Separa los hechos de las opiniones
- Pero se debe evitar el "mal" razonamiento.
Por ejemplo, errores en la interpretación.

La gestión usando AdD

El éxito de la analítica sólo puede medirse en términos de lo bien que ayudan a lograr objetivos estratégicos

Por lo tanto, se debe:

- Identificar los objetivos de negocio
- Recoger los datos necesarios para medir sus objetivos
- Analizar los datos
- Sacar conclusiones sobre la base de la información generada

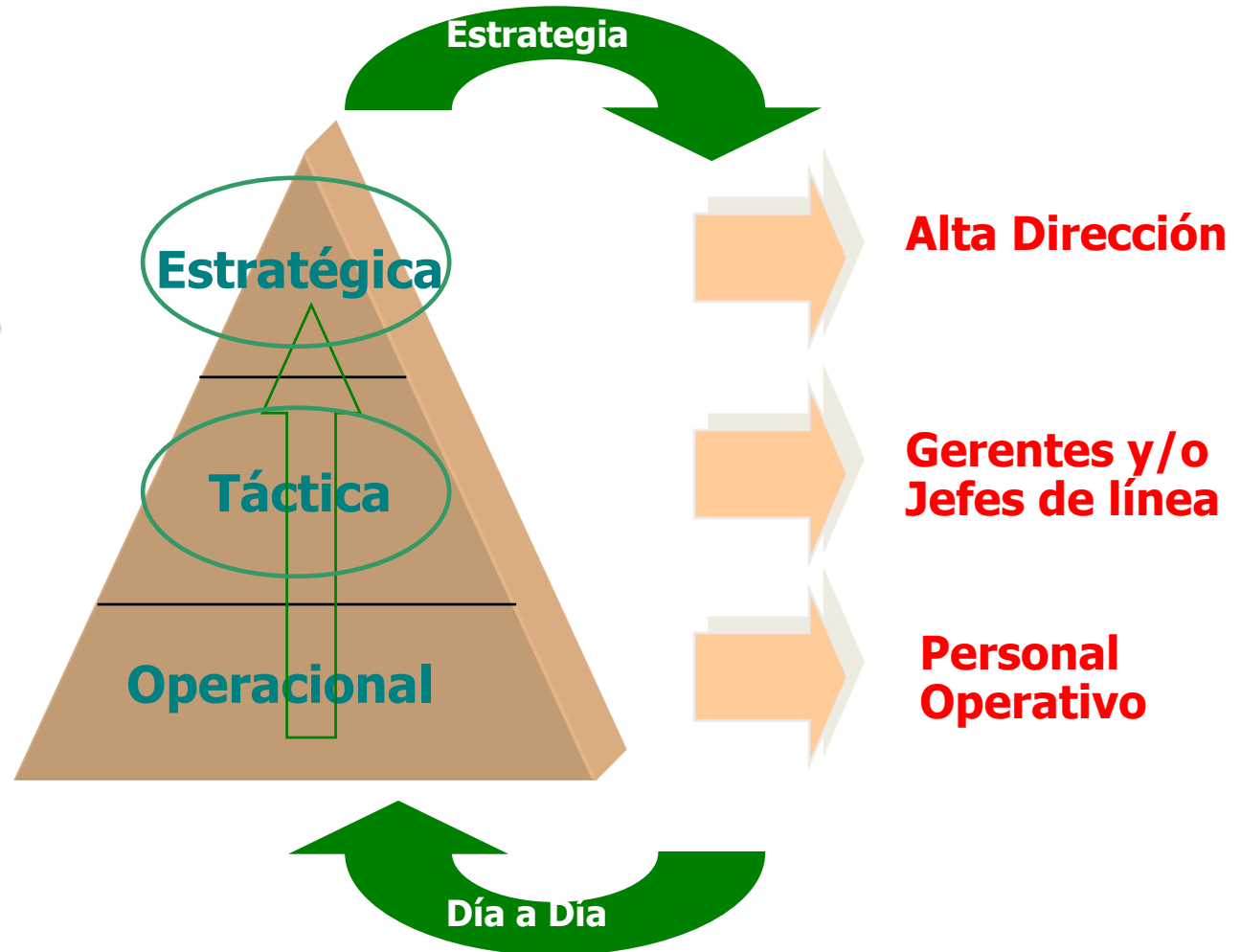
Un análisis debe tener cuatro elementos:

- Los datos/información
- Razonamiento Científico (¿qué? ¿quien? ¿dónde? ¿qué pasa?)
- Hallazgo (¿qué resultados?)
- Lección/conclusión (¿y por qué? ¿cómo?, ...)

Tipos de análisis de datos que se pueden realizar



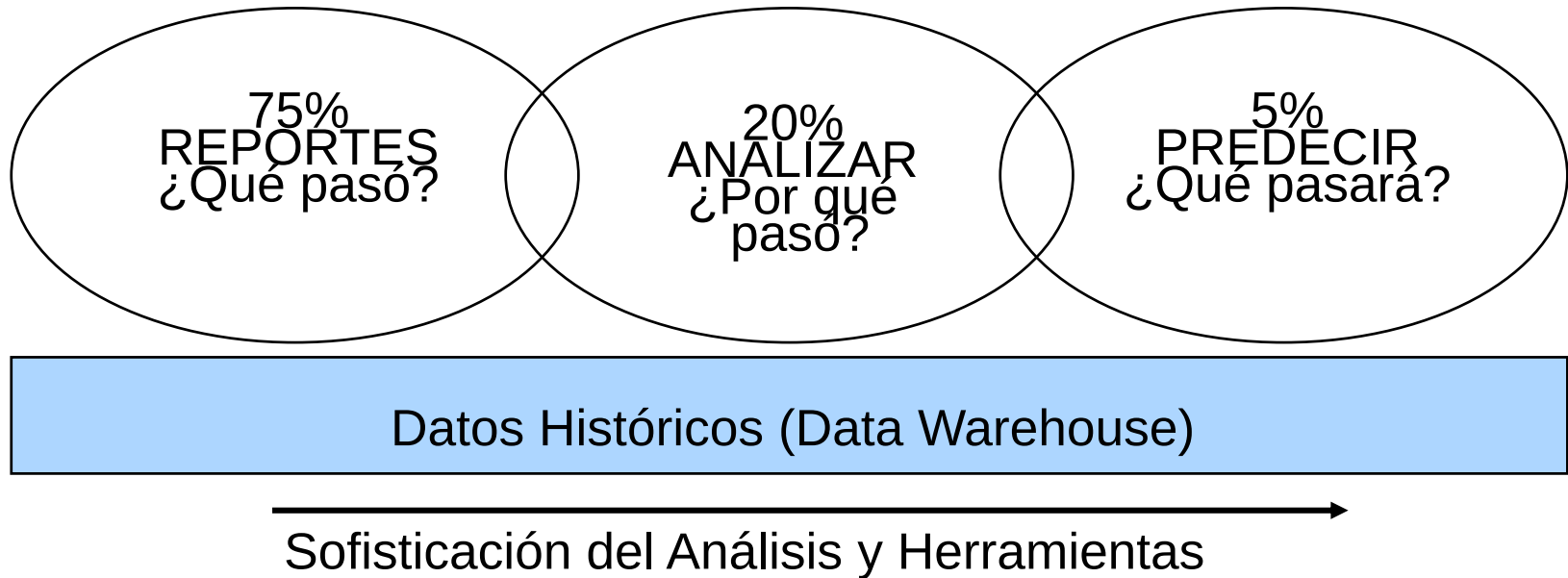
Conforme se avanza de los tácticos hacia los estratégicos, los análisis son más complejos, riesgosos y pueden dar una mayor recompensa.



Tipos de análisis de datos que se pueden realizar

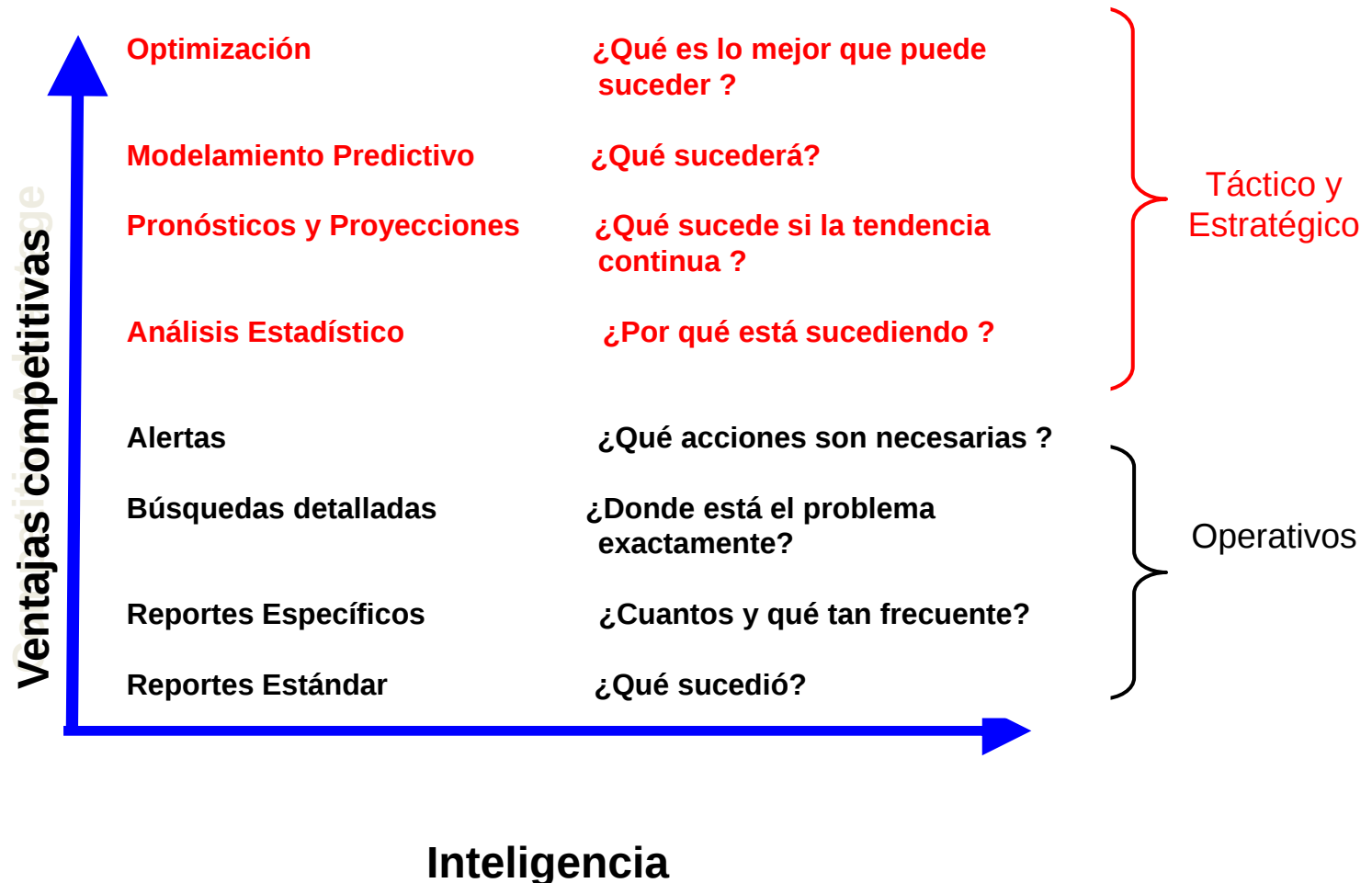


Dominios del Análisis Táctico y Estratégico

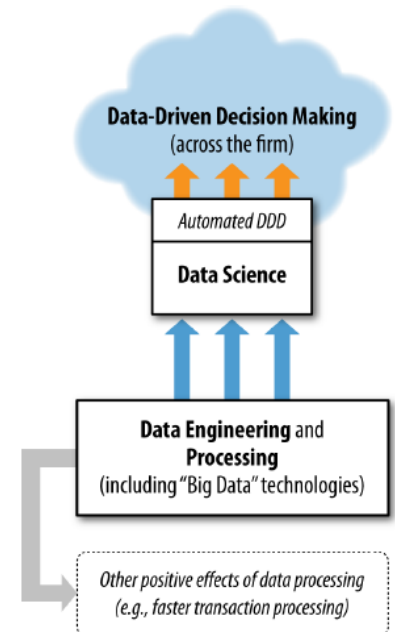


- **Decisiones Tácticas, siguiente semana o mes**
- **Decisiones Estratégicas, siguiente semestre o año**

Tipos de análisis de datos que se pueden realizar



la ciencia de datos requiere de principios, procesos y técnicas para la comprensión de los fenómenos a través del análisis (automatizado) de los datos.



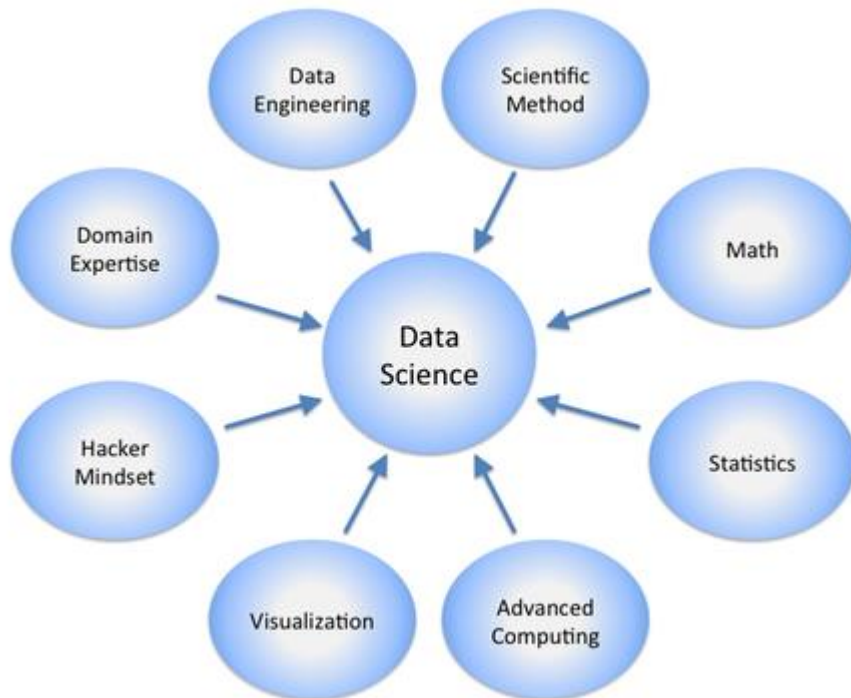
Conocer los datos

La ciencia de datos es un procedimiento que **consume tiempo y requieren mucho trabajo**, pero que es absolutamente necesario para la AdD con éxito.

Expertos de dominio deben ser consultados para explicar las anomalías, los valores perdidos, el significado de los números enteros que representan categorías en lugar de cantidades numéricas, y así sucesivamente.

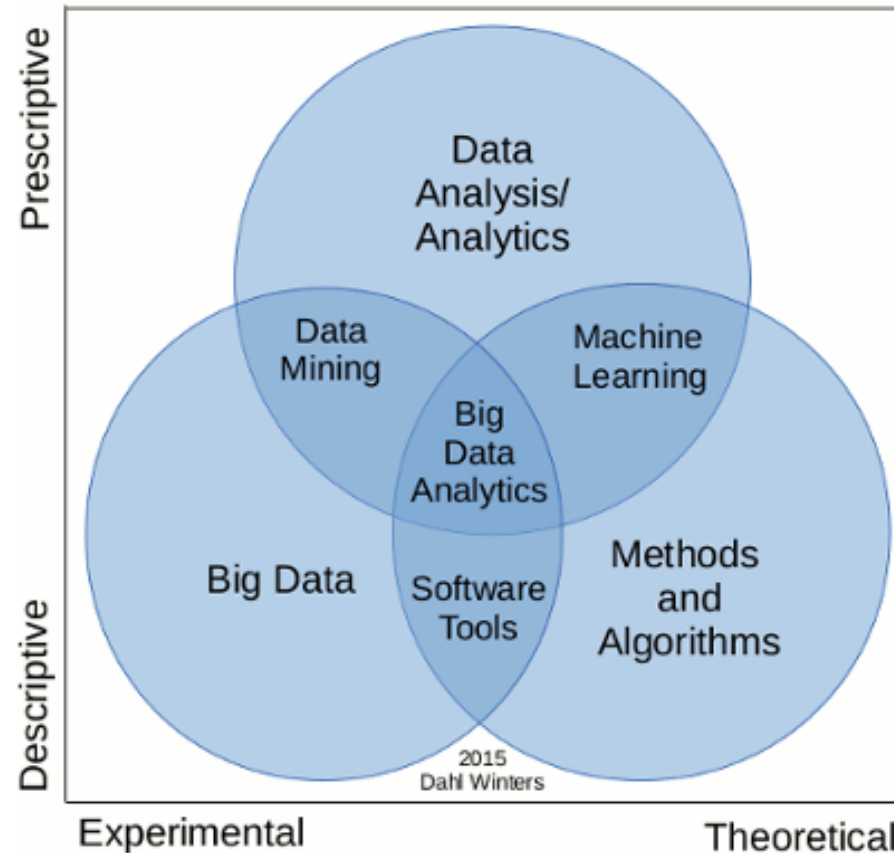
Los datos deben pasar por **procesos de preparación, limpieza, integración, y agregación**.

La ciencia de los datos



Combinación de las matemáticas, estadísticas, etc., para resolver el problema de captura de datos, además de la limpieza, la preparación y la alineación de los datos.

The Fields of Data Science



Limpieza

se refiere a una serie de procesos en los cuales la **calidad de los datos es mejorada**, enfrentando los problemas mencionados como datos mal capturados, anómalos y vacíos, etc.

- normalización de formatos,
- remoción de anomalías,
- corrección de errores
- eliminación de duplicados.

Transformación

En esta etapa se **transforman las variables de entrada en nuevas variables de interés**, a través de diversos métodos.

Una transformación de variables puede ser la combinación entre variables

- concatenación de cadenas,
- multiplicación entre variables,
- otras operaciones aritméticas, etc.

Reducción

Consiste en **decidir qué datos deben ser utilizados para el análisis**. El criterio que se sigue para realizar reducción de variables incluye la relevancia con respecto a los objetivos que se persiguen, y limitaciones técnicas tales como los volúmenes máximos de datos o bien tipos de datos concretos.

Así que en este paso se reduce la cantidad de variables a sólo las necesarias para modelar el proceso en estudio.

- Realizar análisis estadísticos para reducir variables que posean una alta relación lineal, como por ejemplo un análisis de correlación.
- Identificar las posibles variables que se pueden reducir.
- Justificar la reducción de las mismas
- Construir la nueva vista minable con las nuevas variables reducidas

- Privacidad
- Seguridad
- Decisiones basado en datos incompletos
- Decisiones con datos inexactos
- Usando sólo los datos que apoyan nuestras decisiones
- Llegar a la conclusión errónea de los datos: por ejemplo, los precios de las acciones

Ciencias de los Datos

Medición de la Gestión de Conocimiento con AdD

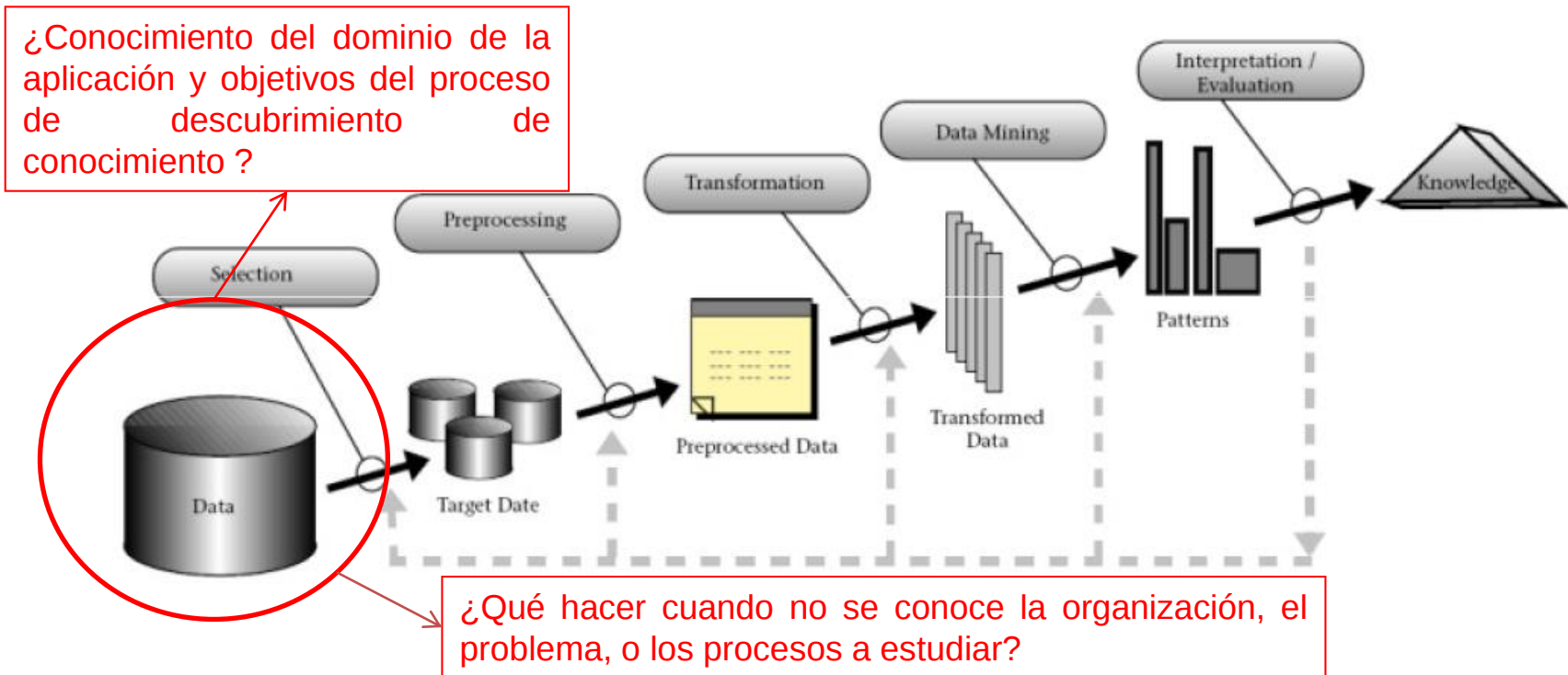
- La auditoria en la GC
- Sistemas de memoria organizacional
- El rehúso y la innovación
- Lecciones aprendidas y mejores practicas
- Amnesia Corporativa

Introducción a la Metodología MIDANO

MIDANO

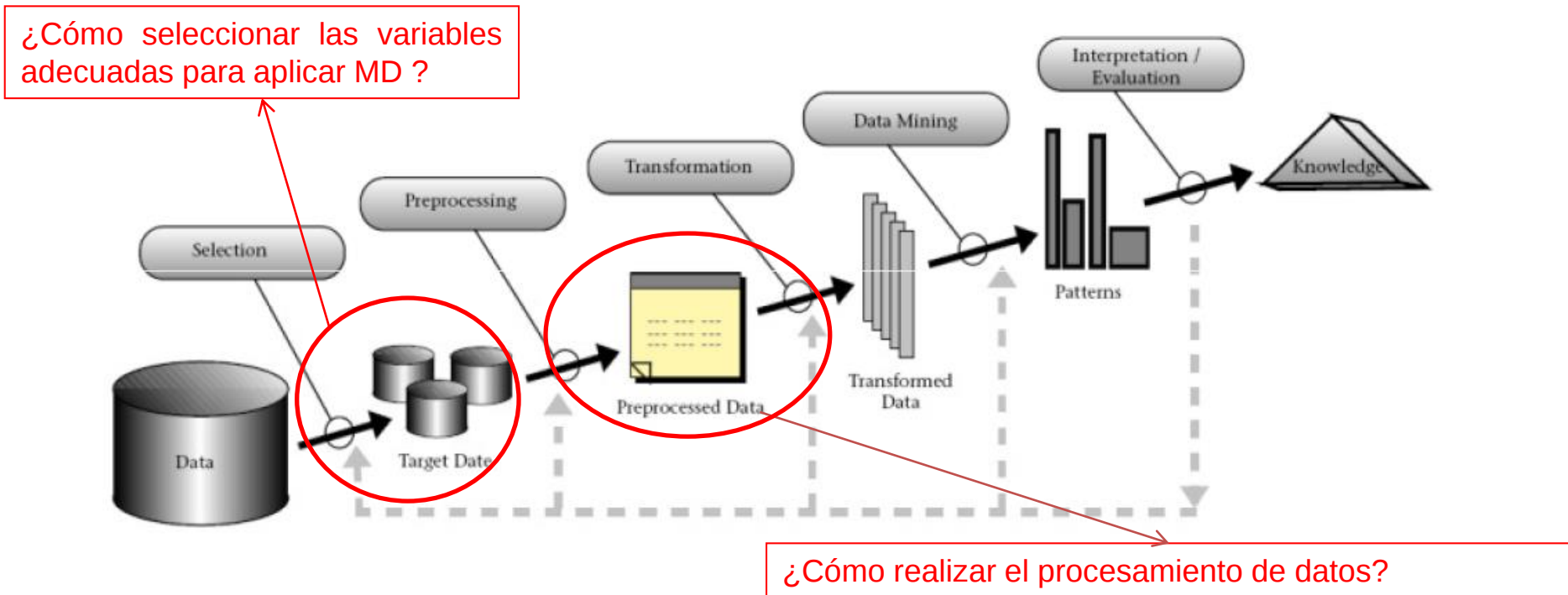
“Metodología para el desarrollo de aplicaciones de minería de datos basada en el análisis organizacional”

MIDANO



“Metodología para el desarrollo de aplicaciones de minería de datos basada en el análisis organizacional”

MIDANO



“Metodología para el desarrollo de aplicaciones de minería de datos basada en el análisis organizacional”

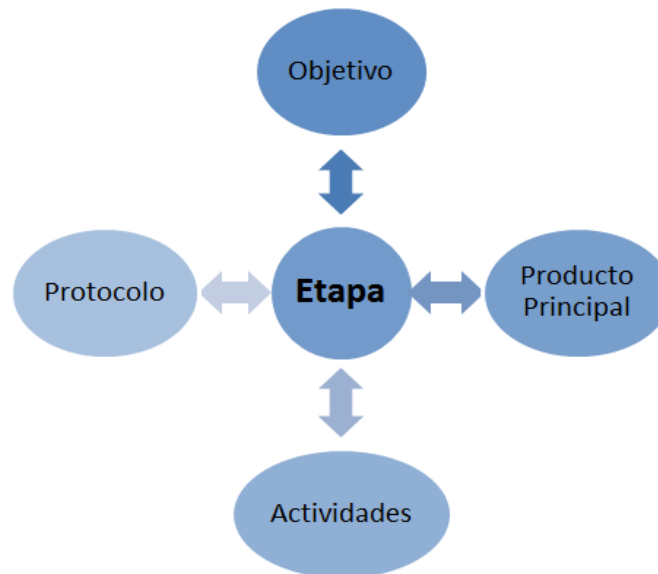
MIDANO

MIDANO consta de tres fases.



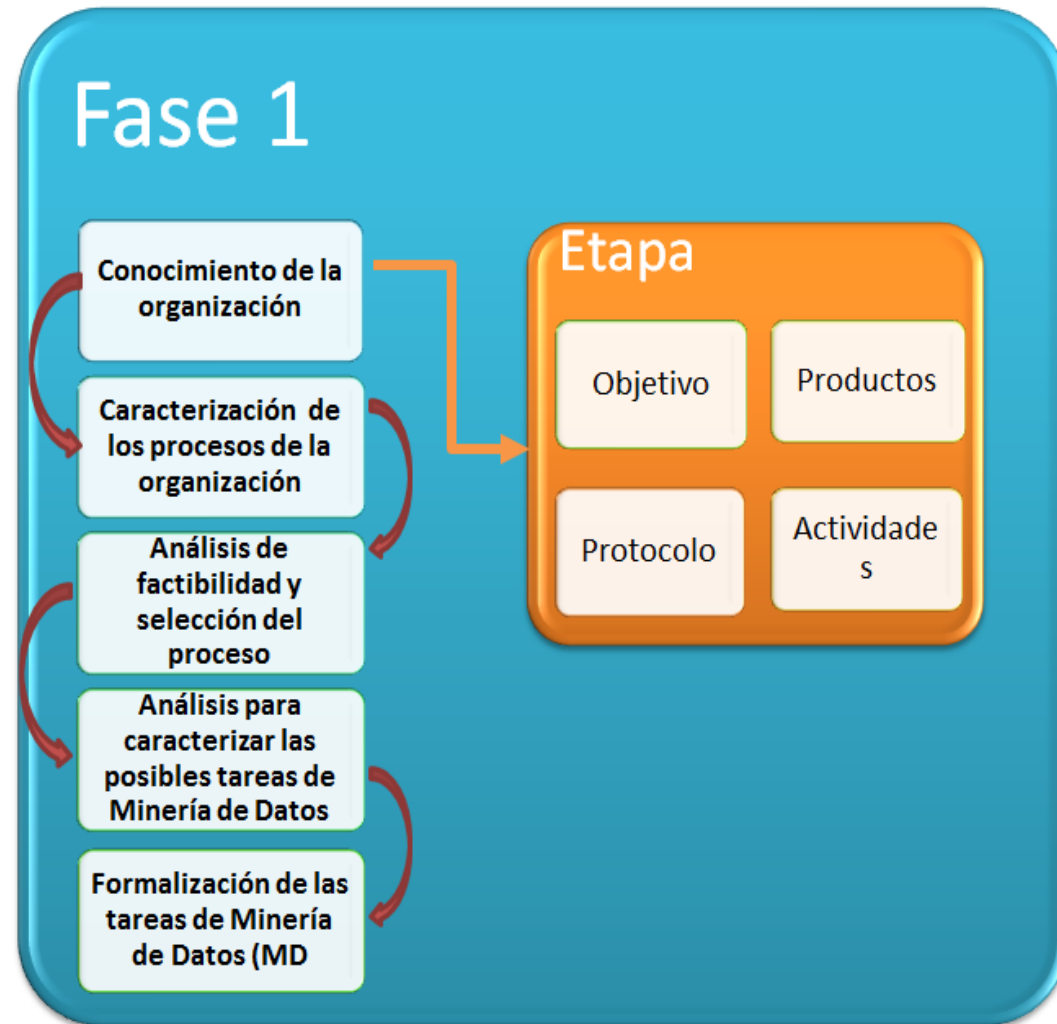
MIDANO

Cada fase de la metodología está dividida en etapas, los elementos que describen cada etapa. En cada etapa se especifican cuatro aspectos principales, como se describe a continuación.



Fase 1: Conocimiento de la Organización

Esta fase tiene como finalidad realizar un proceso de ingeniería de conocimiento, orientado a organizaciones/empresas, de las cuales no se conoce o se tiene poca información del (de los) problema(s), o los procesos a estudiar.



Etapas 1: Conocimiento de la Organización

1. Objetivo { • Conocer la organización/empresa, sus objetivos, procesos, objetos y actores

2. Protocolo de la Fase:

- Descripción de los elementos de la institución/empresa y sus características. Objetivos, Procesos , Objetos y Actores.
- Descripción de las relaciones entre estos elementos.
- Organización de estos elementos.

Etapa 1: Conocimiento de la Organización

Preguntas y ejemplos para determinar los elementos de la institución/empresa

Elemento	Preguntas	Ejemplos
Objetivos	¿Cuál es la razón de ser de la institución?	Conocer, determinar, establecer, la finalidad de la institución/empresa.
Procesos	¿Cuales son las actividades que permiten alcanzar los objetivos de la institución?	Procesos de producción o administrativos.
Objetos	¿Qué cosas o entidades se manipulan en los procesos de la institución?	Pueden ser físicos o abstractos, departamentos, documentos, herramientas, plantas.
Actores	¿Quiénes ejecutan los procesos?	Personas, sistemas, máquinas, etc.

Etapa 2: Caracterización detallada de los procesos de la organización

1. Objetivo {
- Conocer los procesos sobre los cuales se puede enfocar el proyecto de minería de datos.

2. Protocolo de la Fase:

- Familiarización con los procesos sobre los cuales se puede realizar la ingeniería de conocimiento
- Identificación de la fuente de conocimiento
- Familiarización con los ambientes computacionales donde se encuentran los datos a ser utilizados en cada proceso.

Etapas 2: Caracterización detallada de los procesos de la organización

1. Familiarización con los procesos sobre los cuales se puede realizar la extracción de conocimiento

- ¿Qué productos generan esos procesos?
- ¿Qué beneficios proporcionan esos procesos a la organización?
- ¿Qué problemas tienen actualmente?
- ¿Importancia de esos procesos para la organización, o impacto sobre otros procesos?
- ¿Qué impacto generaría la mejora de esos procesos o el estudio de los mismos?

2. Identificar la fuente del conocimiento

- ¿Cuáles son los actores o personas que intervienen en los procesos?
- ¿Quién o quiénes son las personas expertas en los procesos?
- ¿Existen documentos que permitan conocer esos procesos?
- ¿Existen sistemas computacionales que intervengan o interactúen en el proceso?

3. Familiarización con los ambientes computacionales donde se encuentran los datos a ser utilizados en cada proceso explicado

- ¿Dónde se encuentran los datos almacenados del proceso en cuestión?
- ¿Cómo se almacenan los datos del proceso?
- ¿Qué variables son observadas del proceso?
- ¿Cuáles son las variables más importantes de esos datos para la organización?

Etapla 3: Análisis de factibilidad y selección del proceso

1. Objetivo
- 
- Analizar los procesos con la información proporcionada/recogida, con la finalidad de conocer la factibilidad de la aplicación de la minera de datos sobre cada uno de ellos

2. Protocolo de la Fase:

- Revisión de los procesos propuestos por los expertos
- Disponibilidad del experto o grupo de expertos
- Análisis de las fuentes de información sobre los procesos

Etapas 4: Análisis para caracterizar las posibles tareas de Minería de Datos

1. Objetivo {
- Caracterizar las posibles tareas de minería de datos a realizar en el(los) proceso(s) seleccionado(s) en la fase anterior (objetivos, requerimientos, factibilidad, etc.), con la finalidad de escoger las tareas de MD de interés a desarrollar.

2. Protocolo de la Fase:

- Selección y descripción de los actores.
- Descripción de los escenarios actuales y posibles escenarios futuros de la institución/empresa.
- Especificación de los requerimientos para los posibles escenarios futuros (donde se puedan aplicar tarea(s) de MD)
- Elaboración de los casos de uso para los requerimientos funcionales

Etapa 5: Formalización del Problema/tareas de Minería de Datos (MD)

1. Objetivo

- Definir el(los) problema(s) formales de MD.

2. Protocolo de la fase

- Desarrollo de un informe, con la conceptualización del proceso a estudiar, la caracterización de sus problemáticas operacionales y del uso de la MD en dicho proceso.