

# **Üzleti intelligencia megoldások, avagy vezetői döntéstámogatás (XXI.)**

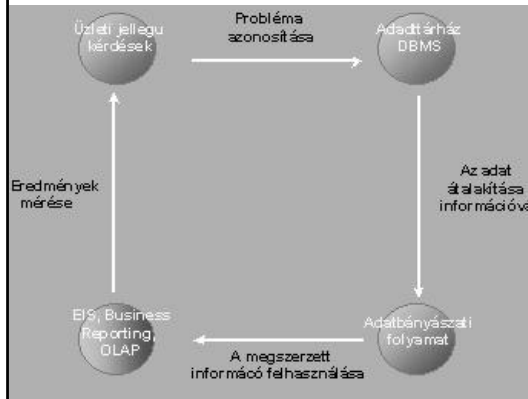
**Pásztor Márta Zsuzsanna  
Szent István Egyetem  
Gazdasági Informatika Tanszék**

**2004. május**

## **Vezetői döntéstámogatás**

- Executive Information Systems (EIS)
- Management Information Systems (MIS)
- Decision Support Systems (DSS)
- Enterprise Information Systems
- Business Intelligence (BI)
  - Data Warehouse (DW)
  - Data Mining (DM)
  - Online Analytical Processing (OLAP)
  - Geographic Information Systems (GIS)

# A teljes információs rendszer



(Barry, Linhoff 1997)

Rendszerelemek: az adat

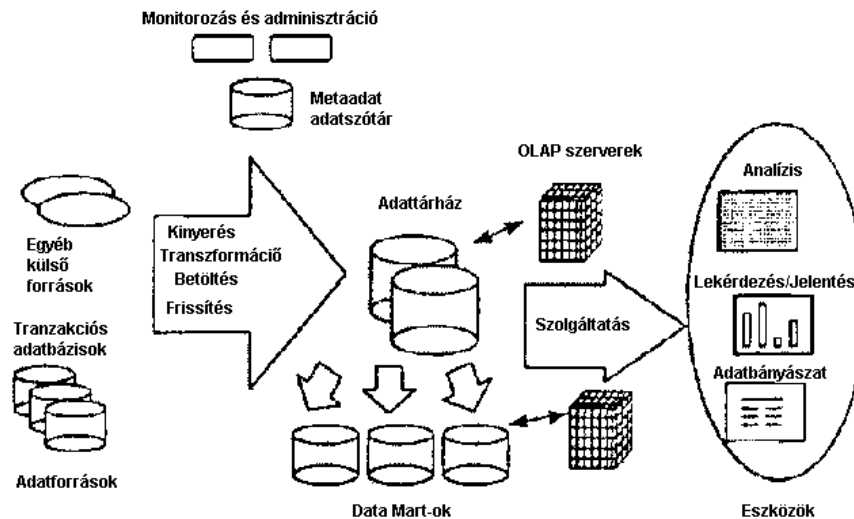
- tárolás - tárház (Inmon)
- feldolgozás - bányászat
- megjelenítés - vizualizáció (Codd)

## Az adattárház

„Az adattárház egy témaorientált, integrált, történeti, nem változó adatgyűjtemény, amelyet elsősorban a döntéshozatalban használnak.” (W.H. Inmon)

- Adatkinyerés a tranzakciós és/vagy más (vállalat-működtetési) rendszerekből
- A kinyert adatok átformálása feldolgozás vagy beszámoló készítés számára
- A riportok (beszámolók) elérhetővé tétele a döntéshozók számára

# Adattárház komponensek



## OLAP

"olyan szoftver technológia, amely analitikusok, üzletemberek, vezetők számára lehetővé teszi, hogy vállalkozásuk adatainak dimenziók szerint rendezett mértékeit gyors, konzisztens és interaktív módon vizsgálhassák" (OLAP Council)

- Multidimenzionális nézet  
(jellemzően adattárház alapon)
- Adatelemzés  
(egyszerű „lefűrés”-től az adatbányászati eredmények felhasználásáig)

## Adatbányászat

„nagy adatbázisokban rejlő rejtett, és kevésbé rejtett információk felismerése és kinyerése”

- Előkészítés: adatforrások azonosítása, adatgyűjtés
- Átalakítás: adattisztítás, hiányzó adatok pótlása
- Felfedezés: modellalkotás, -értékelés, elemzés
- Bemutatás: adat-vizualizáció
- Használat: alkalmazás

## Projekt módszertanok

### ■ CRISP-DM

- SPSS technológia
- hat szakaszt különböztet meg
- alkalmazási helytől és eszköztől független
- Cross-Industry Standard Process for Data Mining

### ■ SEMMA

- SAS technológia
- öt lépésből áll
- a cég adattárház alapú információ szolgáltató rendszerébe épül be
- az elemzési szakasz módszertana, a lépések kezdőbetűiből áll össze az elnevezés

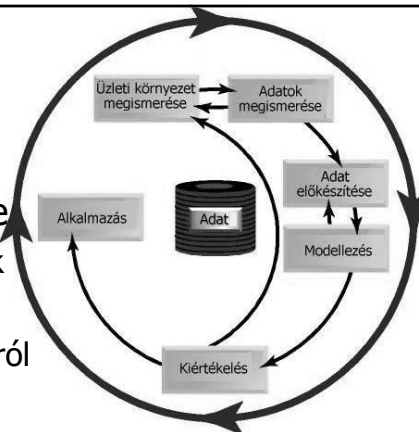
## CRISP-DM (1-2)

### ■ Üzleti környezet megismerése

- projekt célok és követelmények
- adatbányászati probléma leírás
- előzetes terv az elérendő célokról

### ■ Adatok megismerése

- kezdeti adatgyűjtemény összeállítása
- minőségi problémák azonosítása
- elsődleges adatösszefüggések vagy részhalmazok definiálása



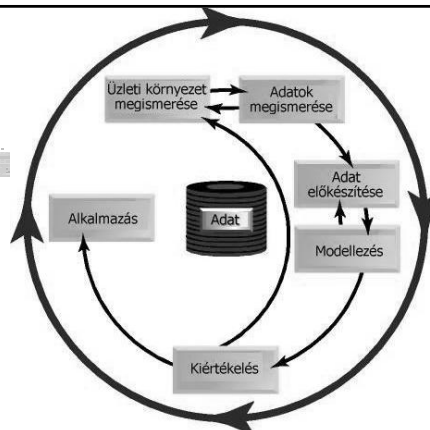
## CRISP-DM (3-4)

### ■ Adatelőkészítés

- a végső adathalmaz összeállítása
- kiválasztás és transzformáció
- valamint az adattisztítás

### ■ Modellezés

- többféle modellezési technika és eljárás kipróbálása
- adatelőkészítési fázisba való visszatérés adatpótlás miatt



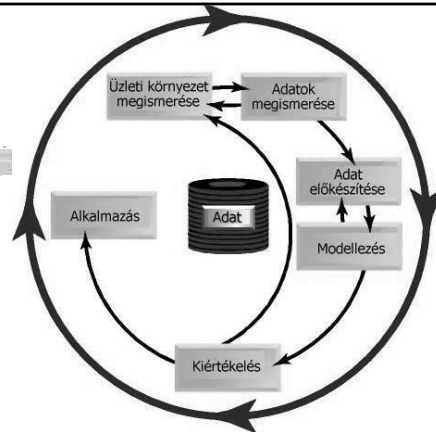
## CRISP-DM (5-6)

### ■ Kiértékelés

- a felépített modellek közül a legjobb kiválasztása
- az elérhető üzleti eredmények számszerűsítése

### ■ Alkalmazás

- a legjobb modell eredményeinek bemutatása
- azon megjelenítési formák kidolgozása, melyek legjobban tükrözik a számítási eredményeket



## SEMMA

- **Sampling** (mintavételezés): külső és belső rendszerekből gyűjtött adatok leválogatása
- **Exploration** (feltárás): hiányzó vagy extrém elemek kiszűrése, esetleges adatösszevonások
- **Modification, Manipulation** (módosítás): hiányzó adatok pótlása, változók összevonása vagy elhagyása
- **Modelling** (modellezés): modellalkotás különféle számítási módszerekkel
- **Assessment** (felmérés, kiértékelés): modellek összehasonlítása, válasz az üzleti kérdésekre

## **Tipikus adatbányászati kérdések az üzleti életben**

- Ki a jó (megbízható) adós, biztosított?  
Kik vásárolnak sokat (sokszor, nagy értékben)?
- Milyen terméket ajánljunk a fogyasztónak?  
Mely termékeket vásárolják együtt?
- Mi az esélye, hogy az adott fogyasztó válaszol a megkeresésünkre?
- Hova telepítsük üzletünket/raktárunkat?

## **Alkalmazott módszerek 1.**

### **A) Kapcsolat-keresés:**

1. Asszociációs keresés
2. Gyakori minták kinyerése
3. Idősoros minták
4. Bayes-hálók

### **B) Klasszifikáció:**

1. Klasszifikáló neuronális hálók
2. Döntési fák
3. Bayesi klasszifikáció
4. K-legközelebbi szomszéd
5. Lineáris diszkriminancia elemzés
6. Emlékezet alapú következtetés
7. Asszociációs klasszifikáció

## **Alkalmazott módszerek 2.**

### **C) Klaszterezés:**

1. K-közép
2. Kapcsolatelemzés
3. Klaszterező neuronális háló

### **D) Statisztikai módszerek:**

1. Hagyományos módszerek: középérték, gyakoriság
2. Lineáris regresszió
3. Nem lineáris regresszió
4. Próbák: t-, F-  $\chi^2$
5. Főkomponens analízis
6. Faktoranalízis

## **Alkalmazott módszerek 3.**

### **E) Előrejelzés:**

1. Előrejelző neuronális hálók
2. Radiális bázisfüggvény
3. Fuzzy következtetés
4. Idősor elemzés
5. Emlékezet alapú következtetés



## Fontosabb termékek és az általuk használt elemzési módszerek

Cég	Termék	Kapcsolat-keresés				Klasszifikáció							Klaszterezés			Statistikai módszerek						Előrejelzés				
		1.	2.	3.	4.	1.	2.	3.	4.	5.	6.	7.	1.	2.	3.	1.	2.	3.	4.	5.	6.	1.	2.	3.	4.	5.
Cognos	4Thought						X									X						X				
Cognos	Scenario						X									X										
IBM	Intelligent Miner	X	X	X		X	X		X				X	X	X	X	X	X	X	X	X	X	X		X	X
ISoft	Alice / AC2						X									X		X	X							
Oracle	Darwin					X	X		X							X	X	X				X				X
SAS	SAS Enterprise Miner					X	X				X		X	X	X	X	X	X	X	X	X	X			X	X
Silicon Graphics Inc.	MineSet	X					X	X					X										X			
SPSS Inc.	Clementine, AnswerTree, Neural Connection	X	X	X		X	X					X			X	X	X				X	X			X	
WizSoft	WizWhy / WizRule	X										X														

## Modellösszehasonlítás

- Hogyan értékeljük a különféle módszerekkel kapott eredményeket?
  - összevonás vagy
  - választás
- Melyik (létezik-e) a legjobb modell?
  - múlt-jelen-jövő(?)
  - céltalanság tétele
- Szakértők és a programok együttműködése
  - nem vak bizalom(!)
  - tapasztalat