

```

4780 GOTO 5000
4790 :
4800 REM
4801 REM
4802 REM
4803 REM
4810 :
4820 PRINT
4825 W=V+1
4830 FOR X
4835 FOR I
4840 PRINT
4850 NEXT:
4860 PRINT
4870 FOR I
4880 IF MD
(I+1);:GOT
4890 PRINT
4900 NEXT
4910 PRINT
4920 FOR I
4925 PRINT
4930 IF MD
Q";:GOTO 4
4935 PRINT
4940 NEXT:PRINT
4950 PRINT"#####";
4960 FOR I=2 TO 24 STEP 2
4965 PRINT"|";
4970 IF MD*(I+W-1)="#####" THEN PRINT"#####";
M$(I)"|";:GOTO 4980
4975 PRINT M$(I);
4980 NEXT:PRINT"#####"

```

# Metadata

RNDr. Ondřej Zýka



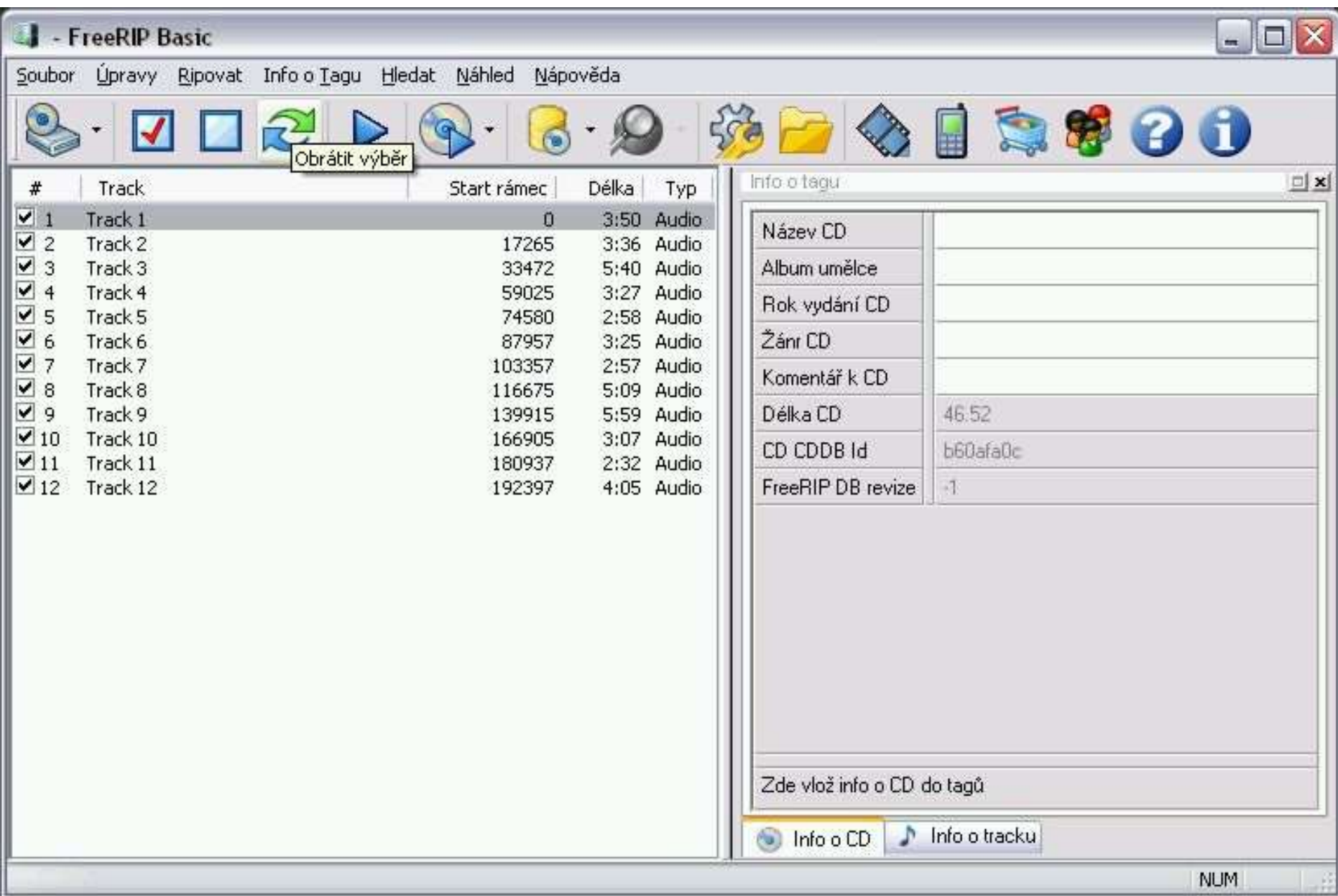
# Metadata

Jedna z kompetencí Data managementu

Cíle kompetence:

- Zajistit jednotné porozumění a užití termínů
- Provázat informace na různých úrovních (byznys, aplikační, technické)
- Integrovat všechna metadata organizace
- Zajistit snadný přístup ke všem metadatům

# Užitečnost metadat



The screenshot shows the FreeRIP Basic application window. The title bar reads "- FreeRIP Basic". The menu bar includes "Soubor", "Úpravy", "Ripovat", "Info o tagu", "Hledat", "Náhled", and "Nápověda". The toolbar contains various icons, including a CD, a checkmark, a square, a refresh button, a play button, a CD with a play button, a folder, a magnifying glass, a gear, a folder, a film strip, a mobile phone, a shopping cart, a group of people, a question mark, and an information icon. A tooltip "Obrátit výběr" is visible over the refresh button.

The main window is divided into two panes. The left pane displays a table of tracks:

#	Track	Start rámeček	Délka	Typ
✓ 1	Track 1	0	3:50	Audio
✓ 2	Track 2	17265	3:36	Audio
✓ 3	Track 3	33472	5:40	Audio
✓ 4	Track 4	59025	3:27	Audio
✓ 5	Track 5	74580	2:58	Audio
✓ 6	Track 6	87957	3:25	Audio
✓ 7	Track 7	103357	2:57	Audio
✓ 8	Track 8	116675	5:09	Audio
✓ 9	Track 9	139915	5:59	Audio
✓ 10	Track 10	166905	3:07	Audio
✓ 11	Track 11	180937	2:32	Audio
✓ 12	Track 12	192397	4:05	Audio

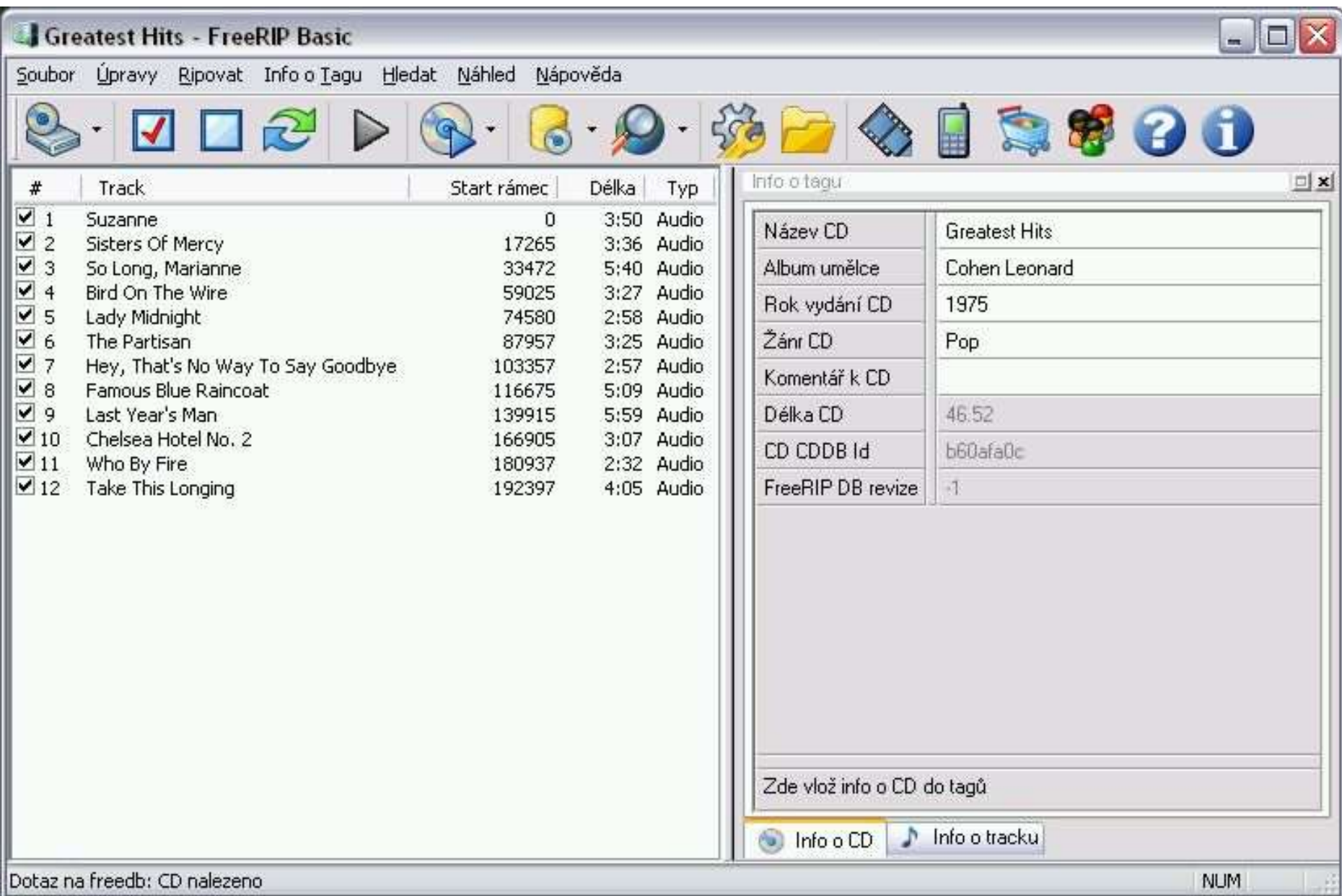
The right pane is titled "Info o tagu" and contains a form for entering CD metadata:

Název CD	
Album umělce	
Rok vydání CD	
Žánr CD	
Komentář k CD	
Délka CD	46.52
CD CDDb Id	b60afa0c
FreeRIP DB revize	-1

At the bottom of the right pane, there is a text field with the placeholder "Zde vlož info o CD do tagů". Below this field are two buttons: "Info o CD" and "Info o tracku".

The status bar at the bottom right of the window displays "NUM".

# Užitečnost metadat



The screenshot shows the 'Greatest Hits - FreeRIP Basic' application window. The main area displays a list of 12 tracks from a CD, each with a checked selection box, track number, title, start time, duration, and type. To the right, the 'Info o tagu' (Tag Info) panel shows metadata for the CD, including title, artist, release year, genre, and CD ID. At the bottom, a status bar indicates 'Dotaz na freedb: CD nalezeno' (Query to freedb: CD found) and a 'NUM' label.

#	Track	Start rámeček	Délka	Typ
<input checked="" type="checkbox"/>	1 Suzanne	0	3:50	Audio
<input checked="" type="checkbox"/>	2 Sisters Of Mercy	17265	3:36	Audio
<input checked="" type="checkbox"/>	3 So Long, Marianne	33472	5:40	Audio
<input checked="" type="checkbox"/>	4 Bird On The Wire	59025	3:27	Audio
<input checked="" type="checkbox"/>	5 Lady Midnight	74580	2:58	Audio
<input checked="" type="checkbox"/>	6 The Partisan	87957	3:25	Audio
<input checked="" type="checkbox"/>	7 Hey, That's No Way To Say Goodbye	103357	2:57	Audio
<input checked="" type="checkbox"/>	8 Famous Blue Raincoat	116675	5:09	Audio
<input checked="" type="checkbox"/>	9 Last Year's Man	139915	5:59	Audio
<input checked="" type="checkbox"/>	10 Chelsea Hotel No. 2	166905	3:07	Audio
<input checked="" type="checkbox"/>	11 Who By Fire	180937	2:32	Audio
<input checked="" type="checkbox"/>	12 Take This Longing	192397	4:05	Audio

Info o tagu	
Název CD	Greatest Hits
Album umělce	Cohen Leonard
Rok vydání CD	1975
Žánr CD	Pop
Komentář k CD	
Délka CD	46.52
CD CDDb Id	b60afa0c
FreeRIP DB revize	-1

Zde vlož info o CD do tagů

Info o CD Info o tracku

Dotaz na freedb: CD nalezeno NUM

# Metadata

- Metadata jsou data popisující data, aplikace, procesy a použití dat. Mohou být reprezentovány jednoduchým popisem, ale také složitou strukturou.
- Metadata jsou strukturované informace, které nám umožňují najít informace o datech, spravovat je, kontrolovat je a porozumět jim.
- Metadata
  - Byznys metadata
  - Strukturální metadata
  - Aplikační metadata
  - Metadata datových transformací
  - Operativní metadata
- Metadata
  - Ve specializovaných systémech
    - Byznys slovník
    - Datové modely
  - Metadata u jednotlivých datových objektů
    - XML
    - MP3 tagy

# Typy metadat

- Strukturální metadata
  - Datové modely
  - Historie datových modelů
  - Popisy extraktů
  - Popisy databází, tabulek a sloupců
  - Technické parametry a byznys popisy
- Aplikační metadata
  - Popis systémů
  - Popis obrazovek a funkcí aplikace
  - Procesní modely aplikací
  - Vazba na datové struktury

# Typy metadat

- Metadata datových transformací
  - Popis na byznys úrovni, slovní popis
  - Technologický popis (SQL, ETL, ELT)
  - Popis jednotlivých transformací
  - Vazba databáze – databáze
  - Logická úroveň – transformované byznys objekty
  - Vazba tabulka – tabulka
  - Vazba sloupec – sloupec
- Operativní metadata
  - Velikost dat
  - Doba trvání procesů a transformací
  - Počet transformovaných objektů (řádků)
  - Kdy a kdo inicioval transformace a procesy

# Procesy kompetence Metadata

- Understand Meta Data Requirements (P)
- Define the Meta Data Architecture (P) (same as 2.7)
- Develop & Maintain Meta Data Standards (P)
- Implement a Managed Meta Data Environment (D)
- Create & Maintain Meta Data (O)
- Integrate Meta Data (C)
- Manage Meta Data Repositories (C)
- Distribute & Deliver Meta Data (C)
- Support Meta Data Reporting and Analysis (O)



# Příklady metadat

- Business Glossary : contains recursive relationship to Business terms.
- Business tags: Contains various affiliation to that term or terms.
- Data Dictionary: contains information from data model tools for the definition of metadata elements and their technical definitions provided by data or enterprise architecture.
- Conceptual data models:
- Logical data models
- Physical data models
- Databases
- Validation rules and data quality rules
- ETL, business rules and their relationship to attributes and entities
- Reports description
- Source to target mapping artifacts (relationships)
- Reporting requirements (relationships)
- Business processes and their relationship to technology
- People hierarchy and their relationship
- Owner relationship

# Business metadata

- Jednotný slovník organizace
- Komunikace
  - Mezi odděleními
  - Mezi Byznysem a IT
  - Řešení výjimek
- Požadavky
  - Schvalovací proces
  - Diskuse
  - Více druhů slovníků

Biometric risk	Entry value
Book value	Equity risk
Break-up basis	European embedded value
Business risk	Excess capital
Calamity risk	Exit value
Captive insurer	Expected loss (only for credit insurance)
Carrying amount	Expected policyholder deficit
Casualty insurance	Expected shortfall
Catastrophe risk	Expense risk
Central estimate	Extreme value model
Claims risk	Fair value
Compliance risk	Financial conglomerate
Concentration risk	Financial group
Confidence level	Foreign exchange risk
Contagion	Funeral insurance
Contingent capital	Fungible capital
Cost of capital approach	General insurance
Counterparty credit risk	Going concern basis
Credit insurance	Group insurance
Credit risk	Guarantee
Critical illness insurance	Guaranteed benefit
Current entry value	Guaranteed element

Interest rate risk	Prorogation
Internal model	Product design risk
Lapse	Property and casualty insurance
Lapse risk	Property insurance
Lead supervisor	Provision
Legal risk	Prudential filter
Liability insurance	Prudent person approach
Life insurance	Pure endowment insurance
Liquidity risk	Quantile approach
Longevity risk	Rating agency driven capital
Loss given default	Real-estate risk
Management risk	Regulatory capital
Marine, aviation, and transport insurance	Regulatory surplus

# Operace s metadaty

## Náklady

- Shromažďování
- Integrace

## Přínosy

- Prezentace
- Analýzy

# Shromažďování metadat

- Byznys metadata vyžadují lidskou interakci
  - Shromažďování
  - Validace
- Všechny ostatní metadata je nutné shromažďovat automaticky
  - Modely
  - Schémata
- Využití metadat pro vývoj (generování komponent)
  - Model driven architecture
- Shromažďování metadat z hotových řešení
  - Analýza kódu
  - Reverzní inženýrství
- Vysoká závislost na technologiích

# Integrace metadat

- V podniku jsou pouze jedna metadata
- Provázat metadata od definice na business úrovni až k technickým detailům, od zdrojů dat k reportům.
- Často existují lokální ostrůvky kompetence
  - Lokální slovníky
  - Lokální popisy vazeb, struktur, závislostí
  - Často špatně technologicky podporováno
  - Integrace na základě emailů, excelů a množství jednání
  - Pouze znalosti zaměstnanců
- Integrace mezi byznys daty a strukturálními daty
- Integrace mezi strukturálními metadaty a metadaty datových toků.

# Prezentace metadat

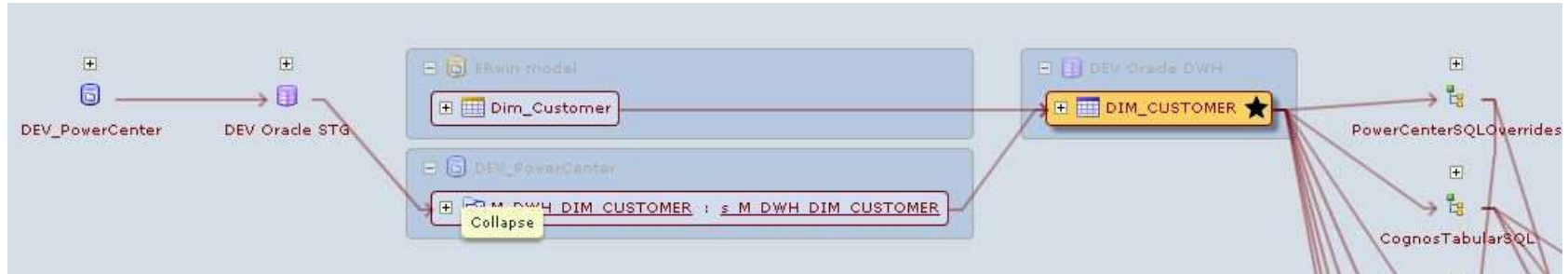
- Rychlé zapojení nových pracovníků
- Zjednodušení komunikace
- Schopnost najít potřebná data a informace
  
- Intranet a generování dokumentace
- Pluginy do aplikací a prohlížečů pro napojení na metadata
  
- Data mohou být důvěrná a musí být chráněná
- Metadata musí být maximálně veřejně a jednoduše dostupná

# Analýza metadat

- Historie
  - Kdo a kdy naposledy upravil proceduru `procedure_name` tak, že nepoužívá tabulku `table_name`?
- Data Lineage
  - Upstream
    - Které aplikace používají centrálních číselník měn?
  - Downstream
    - Která všechna data se podílejí na ohodnocení spolehlivosti dodavatele?
- Impact analysis
  - Které všechny tabulky a aplikace se budou muset upravit, když přejdeme z kódování ISO88592 na kódování UTF8?
  - Pokud místo Y/N začneme používat A/N, co všechno musíme zkontrolovat?

# Metadata - analýza

- Lineage analýza



- Katalóg

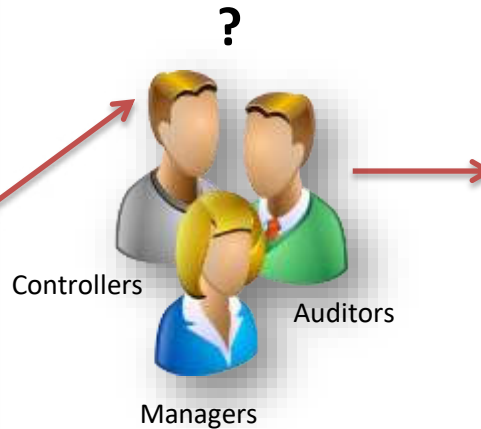
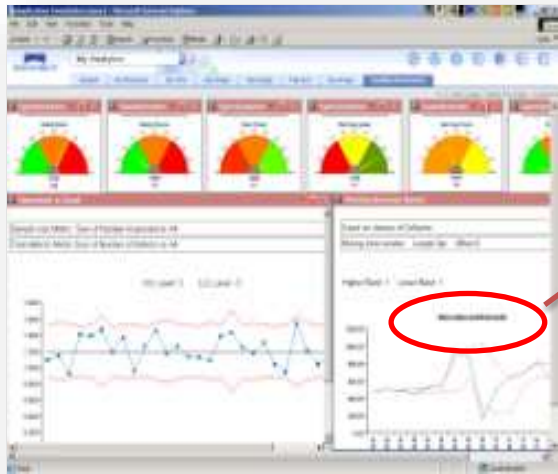
## Where – used analýza



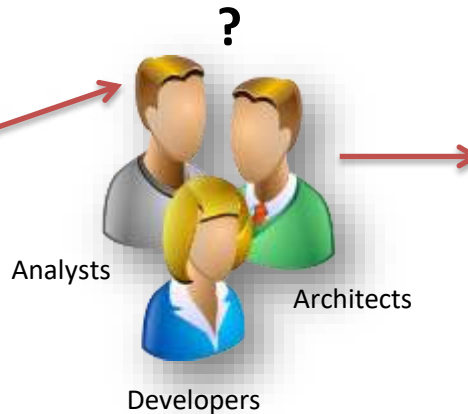
Name	Type
DEV Oracle DWH	Repository
DWH	OracleSchema
DIM_LINK ★	OracleTable
PKDIM_LINK	OracleIndex
INSERT_DUMMY_VALUE	OracleProcedure
ERwin model	Repository
ST R23 Telco Insight 2-3.0.71_ph	ERwinModel
DIM_LINK	ERwinTable
DEV_PowerCenter	Repository
DAILY_SPA	Folder
M_DWH_DIM_LINK	Mapping
Shortcut_to_DIM_LINK	TargetDefinitionIns...



# Přínosy metadat



- 1. Jak je pojem definován?
- 2. Odkud se vzala data?
- 3. Jak jsou data aktuální?



- 1. Co vše musím upravit při změně zdrojového systému?
- 2. Které všechny reporty musím opravit, když změním definici sloupce?
- 3. Co se stane, když havaruje toto ETL?

# Nástroje

- Byznys slovník
  - Semanta
  - Collibra
  - Informatica Metadata Manager
- Informatica Metadata Manager
- Oracle Metadata Directory
- IBM InfoSphere Metadata Workbench
- Adaptive Metadata Manager™
- InfoLibrarian™
- Meta Integration
- ASG Rochade
- SAP PowerDesigner



# Co si zapamatovat

- Co to jsou metadata
- Co to jsou byznys metadata
- Jak se liší byznys metadata od technických metadat
- Co jsou zdroje technických metadat
- Co to jsou operační metadata
- Které čtyři činnosti jsou nutné pro správu metadat
- Jaké typy analýz metadat se používají

```

4780 GOTO 5000
4790 :
4800 REM -----
4801 REM --- DARSTELLUNG ---
4802 REM --- DES MANUALS ---
4803 REM -----
4810 :
4820 PRINT" ";
4825 W=V+1:IF W<8 THEN W=W+14
4830 FOR X=1 TO 2:PRINT"XXXXXXXXXXXX";
4835 FOR I=0 TO 23
4840 PRINT MD$(I+W);
4850 NEXT:PRINT:NEXT
4860 PRINT"XXXXXXXXXXXX";
4870 FOR I=0 TO 23
4880 IF MD$(I+W)=CHR$(32) THEN PRINT M$(
(I+1));:GOTO 4900
4890 PRINT MD$(I+W);
4900 NEXT
4910 PRINT:PRINT"XXXXXXXXXXXX";
4920 FOR I=2 TO 24 STEP 2
4925 PRINT"|";
4930 IF MD$(I+W-1)="  " THEN PRINT"
";:GOTO 4940
4935 PRINT " ";
4940 NEXT:PRINT" "
4950 PRINT"XXXXXXXXXXXX";
4960 FOR I=2 TO 24 STEP 2
4965 PRINT"|";
4970 IF MD$(I+W-1)="  " THEN PRINT"
"
M$(I)" ";:GOTO 4980
4975 PRINT M$(I);
4980 NEXT:PRINT" "

```



**Diskuse**

- Otázky
- Poznámky
- Komentáře
- Připomínky

