



## TDDD43 Advanced Data Models and Databases

<http://www.ida.liu.se/~TDDD43>

6hp

## Advanced Data Models and Databases

<http://www.ida.liu.se/~patla00/courses/AdvDB/>

6hp



# Teachers

- Examiner: Patrick Lambrix
- Lectures: Patrick Lambrix, Huanyu Li
- Labs: Ying Li, Huanyu Li
- Director of studies: Patrick Lambrix



# Course literature

- Articles (on web/handout)
- Lab descriptions (on web)



# Databases / Data sources

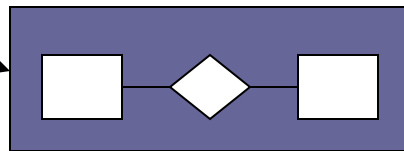
- One (of several) ways to store data in electronic format
- Used in everyday life: bank, hotel reservations, library search, shopping

# Databases

- Database management system (DBMS): a collection of programs to create and maintain a database
- Database system = database + DBMS

# Databases / Data sources

Information

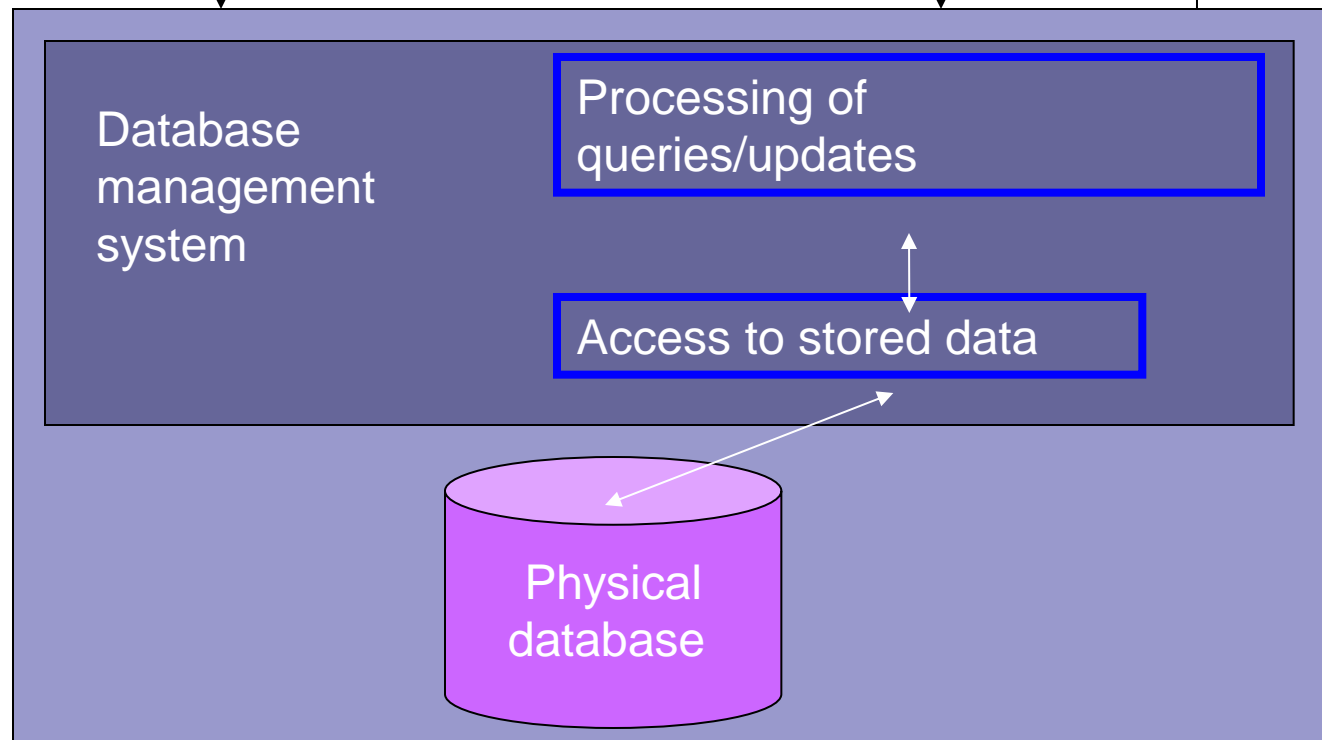


Model

Queries

Answer

Database system





# What information is stored?


- Model the information
  - Entity-Relationship model (ER)
  - Unified Modeling Language (UML)

# What information is stored? - ER

- entities and attributes
- entity types
- key attributes
- relationships
- cardinality constraints
  
- EER: sub-types

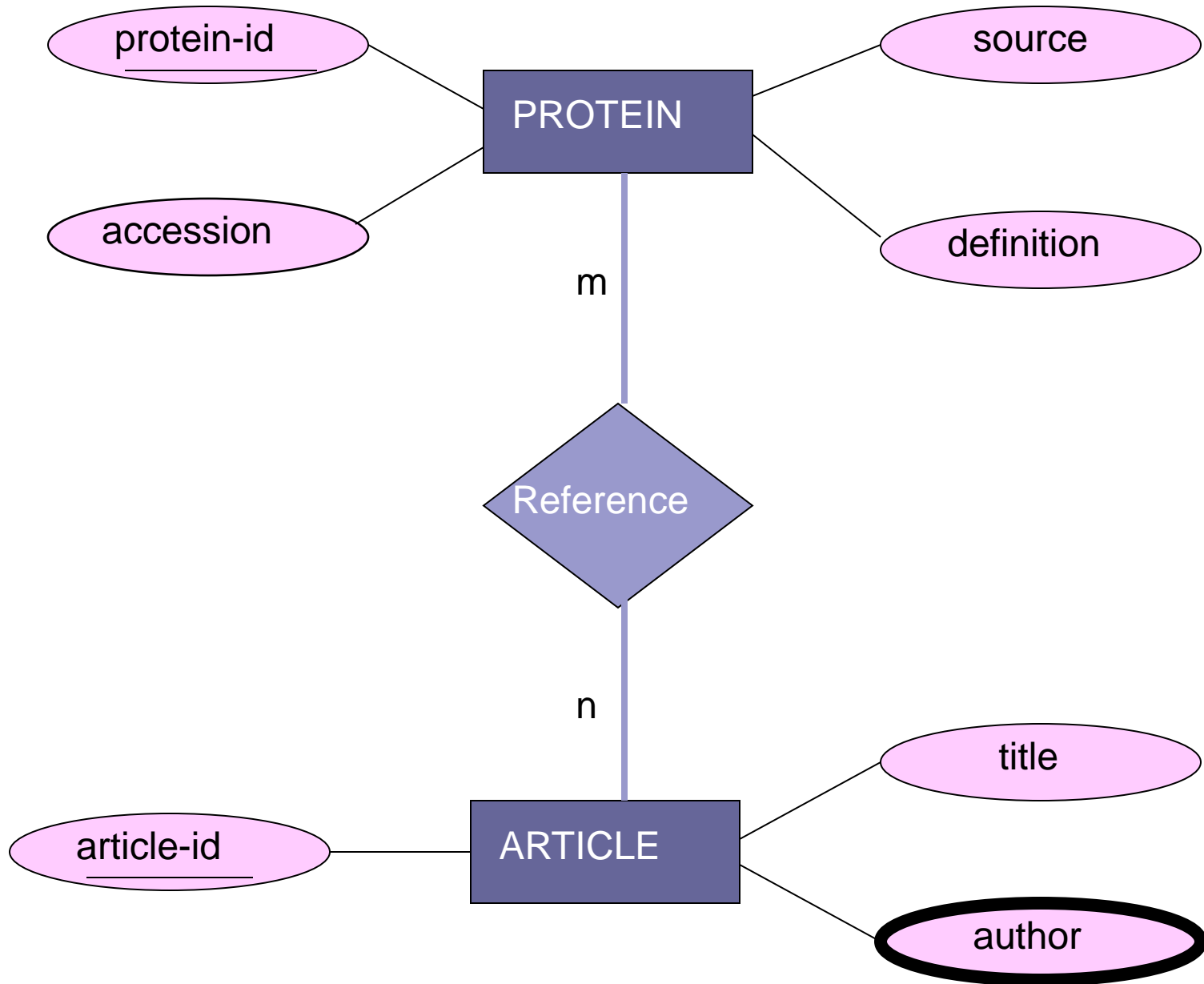


1 tgctacccgc gcccgggctt ctggggtgtt cccaaccac ggcccagccc tgccacaccc  
61 cccgccccg gcctccgag ctgcatg ggcggggggt gctcgtctg ggcgctccg  
121 agcccgtaa cctgctgctg gccgaccgc tccccgacgg cgcggccacc ggcgcgggc  
181 tgctggtgcc cgcgtcgccg cccgctctgt tgctgcctcc cgccagcgaa agccccgagc  
241 cgctgtctca gcagtggaca ggggcatgg gtctgctgat ggcgctcctc gtgctgtca  
301 tcgtggcggg caatgtgctg gtgatcgtg ccatcgccaa gacgccgagg ctgcagacgc  
361 tcaccaacct ctcatcatg tcctggcca ggcggacct ggtcatgggg ctgctggtg  
421 tgccgttcgg ggcaccatc tgggtgtggg gccgctggga gtacggctcc ttcttctcg  
481 agctgtggac ctgagtggac gtgctgtgct tgacggccag catcgagacc ctgtgtgtca  
541 ttgccctgga ccgctacc ccatcacct cggccttccg ctaccagagc ctgctgacgc  
601 ggcgcggggc gcggggcctc gtgtgcaccg tgtgggcat ctgggccctg gtgtcctcc  
661 tgccatcct catgactgg tggcgggagg agagcgacga ggcgcccgc tgctacaacg  
721 accccaagt ctgcgactc gtcaccaacc gggcctacgc catcgctctg tccgtagtct  
781 ccttctacgt gccctgtgc atcatggcct tctgtacct gcgggtgttc cgcgagggcc  
841 agaagcagg gaagaagatc gacagctgc agcggcgtt cctcggcggc ccagcgggc  
901 cggcctcgc ctgcctcgc cccgtccccg cggcggcgc gccgcccga cccccgccc  
961 ccgcccgc cggccacc gcccgctgg ccaacgggag tgcgggtaag cggcggcct  
1021 cgcgctctgt ggcctacgc gagcagaagg cgtcaagac gctgggcatc atcatgggag  
1081 tcttaccgt ctgctggctg cccttctcc tggccaact ggtgaaggcc tccaccgag  
1141 agctggtgcc cgaccgctc tctgtctct tcaactggct gggctacgcc aactcggcct  
1201 tcaacccat catctactc cgcagccccg acttccgcaa ggcctccag ggactgctc  
1261 gctgcgagc cagggtgcc cggcggcgc acgcgacca cggagaccgg ccgagcggc  
1321 cgggctgtct gggcggccc ggacccccgc catcgcccgg ggcgctcgc gacgagcgc  
1381 acgacgatgt cgtcggggc acgcccgc cgcgctgct ggagccctgg gccggctga  
1441 acggcggggc ggcggggac agcactcga gctggacga gccgtcccgc cccggtctg  
1501 cctcggatc caaggtgtg ggcccggc gggcggcggc ctccgggac ggctcccag  
1561 gggaacgagg agatctgtg ttacttaaga ccgatagcag gtgaactcga agcccacaat  
1621 cctcgtctga atcatccgag gcaaagagaa aagccacgga ccgtgcaca aaaaggaaa  
1681 ttgggaagg gatgggagag tggctgctg atgtcctg ttg



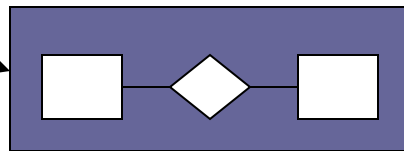
DEFINITION Homo sapiens adrenergic, beta-1-, receptor  
ACCESSION NM\_000684  
SOURCE ORGANISM human  
REFERENCE 1  
AUTHORS Frielle, Collins, Daniel, Caron, Lefkowitz,  
Kobilka  
TITLE Cloning of the cDNA for the human  
beta 1-adrenergic receptor  
REFERENCE 2  
AUTHORS Frielle, Kobilka, Lefkowitz, Caron  
TITLE Human beta 1- and beta 2-adrenergic  
receptors: structurally and functionally  
related receptors derived from distinct  
genes

# Entity-relationship



# Databases / Data sources

Information

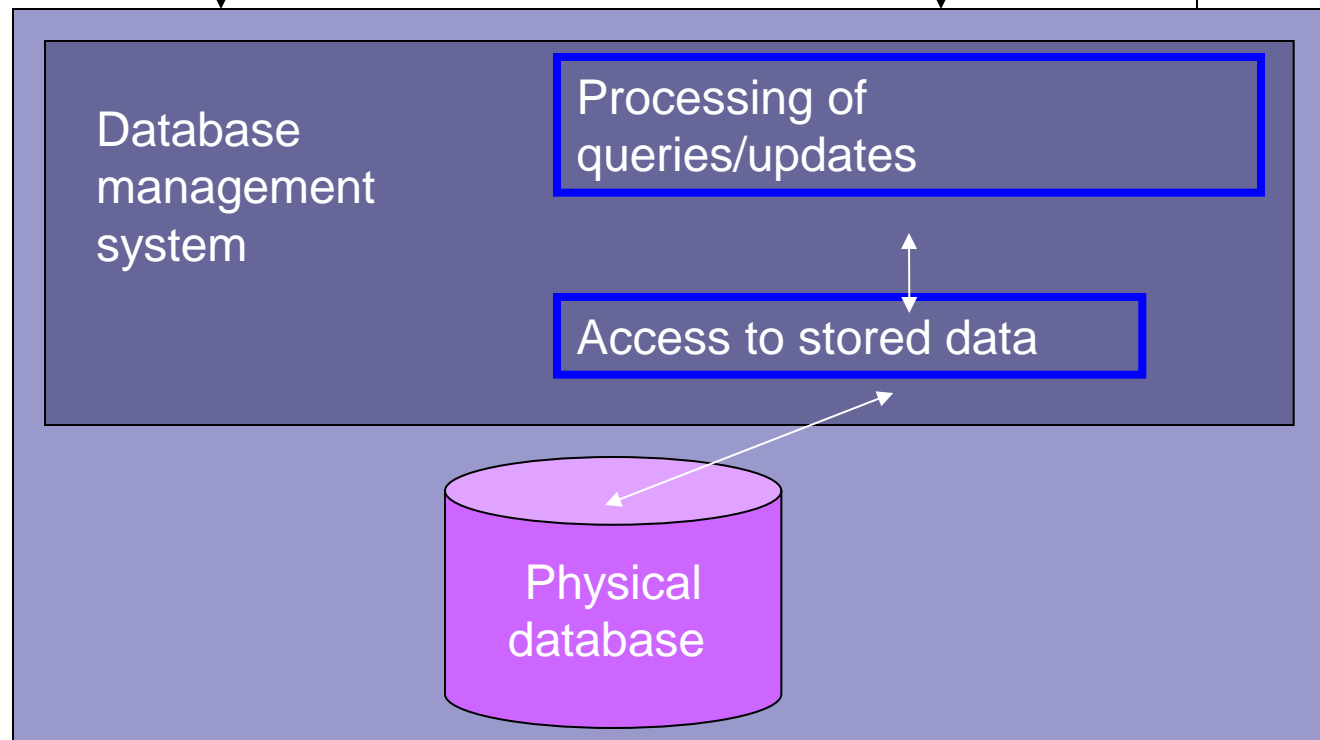


Model

Queries

Answer

Database system

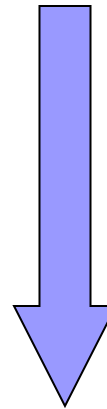


How is the information stored?  
(high level)

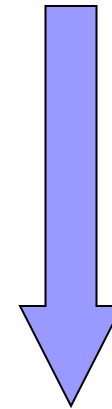
How is the information accessed?  
(user level)

- Text (IR)
- Semi-structured data
- Data models (DB)
- Rules + Facts (KB)

structure



precision



# Course overview

- Information Retrieval (HT1 – lectures)
- Semi-structured data, XML and RDF (HT1 - lectures + labs)
- NoSQL databases (HT2 - lectures + lab)
- Semantic Web, Ontologies, OWL (HT1+HT2 - lectures + lab)
- Data integration (HT1+HT2 – lectures + lab)

# Info

- Results reported in connection to exams
- Info about handing in labs on web; strong recommendation to hand in as soon as possible
- Sign up for labs via web **in pairs (TDDD43); individual (PhD students)**
- Apply for special account for NoSQL lab (NSC) – we will let you know when

# Examination

- TDDD43

- Written exam

- Labs

- PhD students

- Take home exam

- Labs





# Changes w.r.t. last time

- Minor clarifications in labs

# My own interest and research

- Modeling of data
  - Ontologies (for Life sciences, animal health surveillance, materials design, crime scene investigation, sports analytics)
- Ontology engineering
  - Ontology alignment  
(Winner Anatomy track OAEI 2008 / Organizer OAEI tracks since 2013)
  - Ontology debugging and completion  
(Founder and organizer WoDOOM/CoDeS 2012-2016)
  - Ontology visualization  
(Founder and organizer VOILA since 2015)

# My own interest and research

- Sports Analytics
  - Course in VT2
  - <https://www.ida.liu.se/research/sportsanalytics/>
- Former work: knowledge representation, data integration, knowledge-based information retrieval, object-centered databases
- <http://www.ida.liu.se/~patla00/research.shtml>