

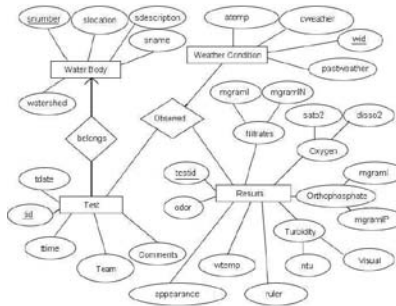
DESIGNING RELATIONAL DATABASES FOR REAL DATASETS

Author: Michael Folding, *Computer Science*

Faculty Advisor: Elena Braynova, *Computer Science*

In this work we analyze an Excel database for water quality of selected local Worcester water bodies and design a relational database for it. The initial data was stored in an Excel database and had 26 fields. The database size was 583 records.

Based on the data analysis the following ER model for a relational database was designed: four Entity Sets - *Test*, *Results*, *Water Body*, *Weather Condition* and two Relationship Sets - *belongs* and *Observed*. Some of the attributes are composite. The ER model was converted to the corresponding Relational Model. Our Database consists of 10 tables.



The tables' relationships are defined by appropriate primary and foreign key constraints.

The tables are defined and initialized by data in SQL.

```
CREATE TABLE WATERBODY
( description
  varchar(25),
  snumber
  varchar(25),
  watershed
  varchar(50),
  sname  varchar(50),
  slocation
  varchar(75),
  primary key(snumber));
```

We run SQL queries on the constructed database and study their efficiency. Here are some examples of the SQL queries:

Show site name, date, and air temperature on test done after 9/1/2004 where the air temperature was greater than 18 Celsius?

```
Select  w.sname, t.tdate,
w1.atemp
From    waterbody w,
belongs b, test t,
observed o, weather w1
Where
w.snumber=b.snumber and
b.tid=t.tid and
t.tid=o.tid and
o.wid=w1.wid and
t.tdate>20040901 and
w1.atemp>18;
```