

Aquisição e análise de dados

Integração I – Bases de dados

Armazenamento de dados

- Tipos de dados e volume de dados
- Abordagens típicas de armazenamento
- Onde, quando e como é que surgem problemas
- Quando é que **Dados = Informação?**

Bases de dados

- Tipos de bases de dados
- Características de uma boa base de dados
- Vantagens de utilização e requisitos de informação
- Caso de estudo I : U.S.G.S. - S. Francisco Bay
- Caso de estudo II : BarcaWin2000

Armazenamento de dados I

Tipos de dados

Metadados

- Instituições
- Equipa(s)
- Projecto
- Sistema(s)
- Campanhas
- Produtos
- Formatos
- Disponibilidade/custo

Dados brutos

Nome da estação

Coordenadas da estação

(...)

Data da amostra

Hora da amostra

Profundidade da amostra

(...)

Nome do parâmetro

Unidades do parâmetro

(...)

Valor medido (resultado)

(...)

Armazenamento de dados II

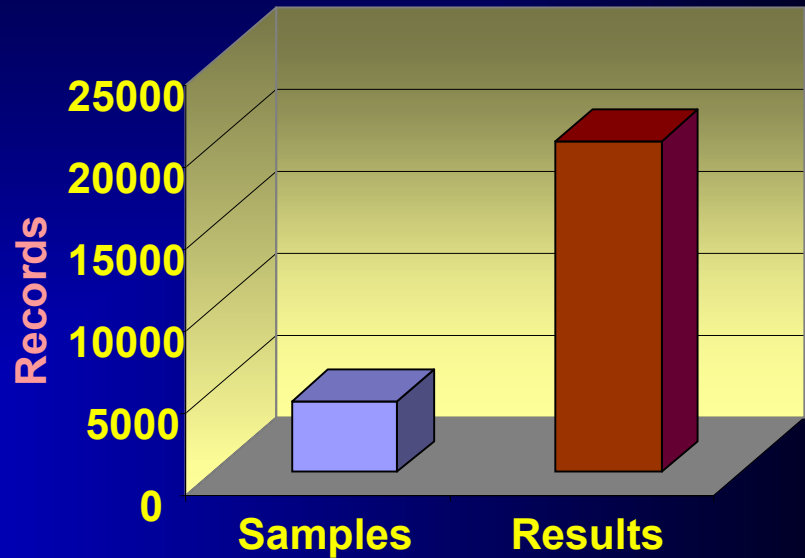
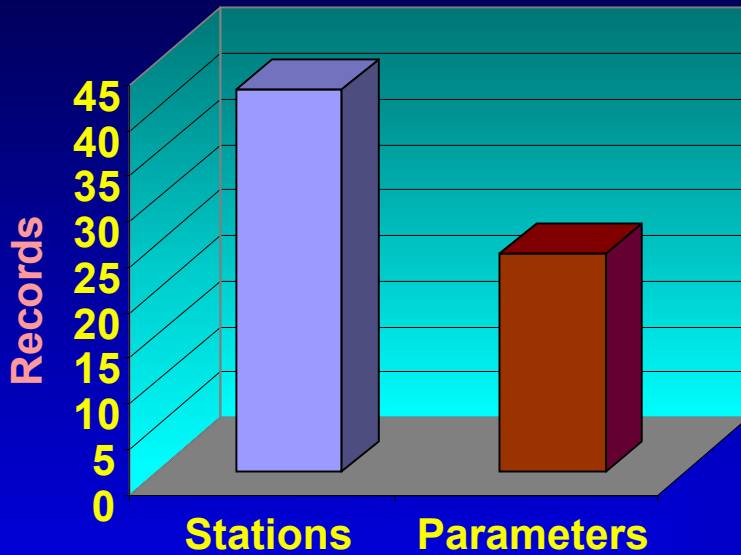
Volume de dados

Projecto	Agência(s)	Sistema	Custo (€)	Duração (anos)	Dados	F/D (€ D-1)
Tejo	UNDP, FCT, Solvay	Tejo	1,200,000	3	68,000	18
INCO*1	EC	Sanggou	300,000	3	60,000	5
OAERRE*2	EC	Ria Formosa	170,000	3	52,000	3

*1 – Um sistema, 600 K€ / 2, não contabiliza dados históricos

*2 – Não contabiliza dados históricos

OAERRE data for Ria Formosa



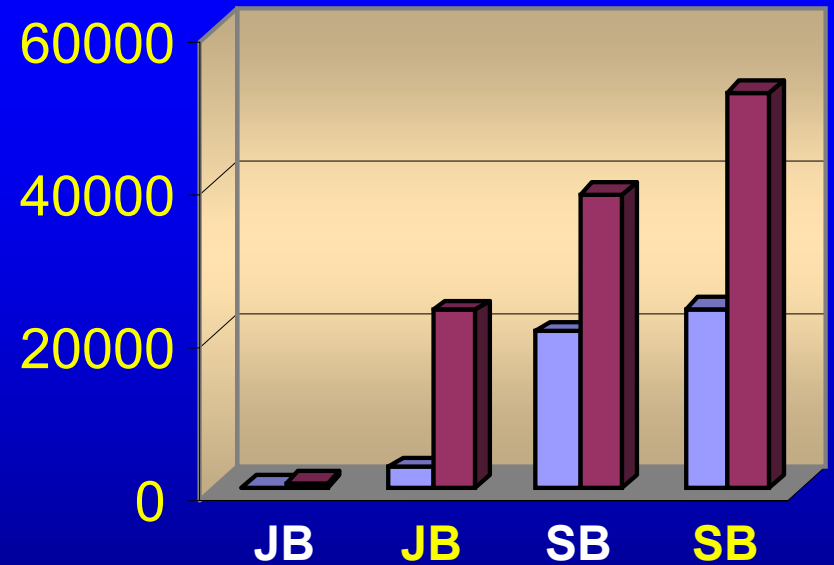
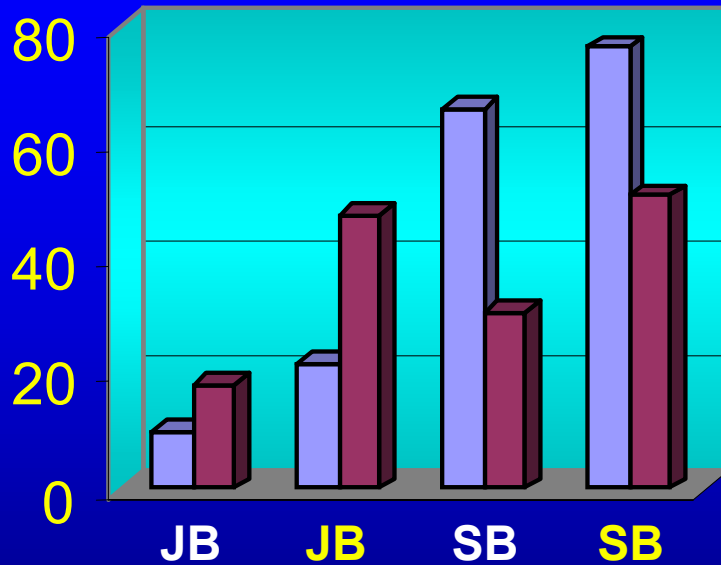
■ Stations	42	■ Samples	4331
■ Parameters	24	■ Results	20215

Historical data sources

- ✓ Manuela Falcão
- ✓ John Icely
- ✓ Alice Newton
- ✓ Carlos Vale
- ✓ Sundry authors

Period: 1985-86, 1987-88, 1989, 1992-93, 2000


Inco-DC data for Sanggou Bay and Jiaozhou Bay



■ Stations	10	22	66	77	■ Samples	80	2728	20536	23621
■ Parameters	18	48	31	51	■ Results	677	23327	38717	51793

Historical data
 Full dataset

Abordagens típicas de armazenamento



Microsoft Excel - SG9401.XLS

File Edit View Insert Format Tools Data Window Help

100% Times New Roman

A8 = 3

	A	B	C	D	E	F	G	H	I
1	Data of N, P, pH, COD, Chlorophyll-a and POM in Sungo Bay, Investigated in Jan. 199								
2									
3	Station	sample	depth(m)	temp (o	time	D2-N (µg	NH4-N (µg	NO3-N (µg	NO2-N (µg
4	1	surface	24	4.8	1.13 10	0.18	0.04	2.6	2.82
5		bottom				0.17	0.04	2.38	2.59
6	2	surface	15	4.7	1.13 11	0.16	0.01	1.92	2.09
7		bottom				0.16	0.49	1.5	2.15
8	3	surface	14	4.8	1.13 10	0.25	0	1.55	1.8
9		bottom				0.45	0.09	1.62	2.16
10	4	surface	16	4.8	1.13 09	0.35	0.35	2.93	3.63
11		bottom				0.4	0.32	1.67	2.39
12	5	surface	13	4.4	1.14 09	0.26	0	1.71	1.97
13		bottom				0.36	0.19	1.03	1.58
14	6	surface	12	4.4	1.14 10	0.38	0	0.77	1.15
15		bottom				0.47	0.22	0.9	1.59
16	7	surface	10	4.7	1.13 11	0.21	0.2	2.16	2.57
17		bottom				0.19	0.04	1.27	1.5
18	8	surface	11	4.7	1.13 12	0.44	0	1.6	2.04
19		bottom				0.46	0.12	2.03	2.61



- **Organização Temporal:** Dados organizados por data (1 conjunto – múltiplas estações)
- **Organização Espacial:** Dados organizados por estação (1 conjunto - múltiplas datas)

Problemas

Onde

- Em qualquer estudo realizado por várias equipas
- Em conjuntos de dados provenientes de projectos diferentes
- Em conjuntos de dados interdisciplinares
- Conclusão: Em todos os estudos de dimensão razoável (e.g. mais que três estações, mais que duas campanhas, alguns parâmetros)

Quando

- No armazenamento e disponibilização atempada
- No controle de qualidade e validação dos dados
- Na conversão de dados em informação

Como

- Impossibilidade de pesquisa eficaz
- Dificuldades (substanciais) no tratamento de dados conjunto
- Perda de informação
- Redundância e/ou perda de dados

Dados e informação

Exemplos para um estuário

Conjunto de dados

1. Salinidade em várias estações, em diferentes ocasiões de amostragem
2. Caudais fluviais
3. Dados de marégrafo



Informação

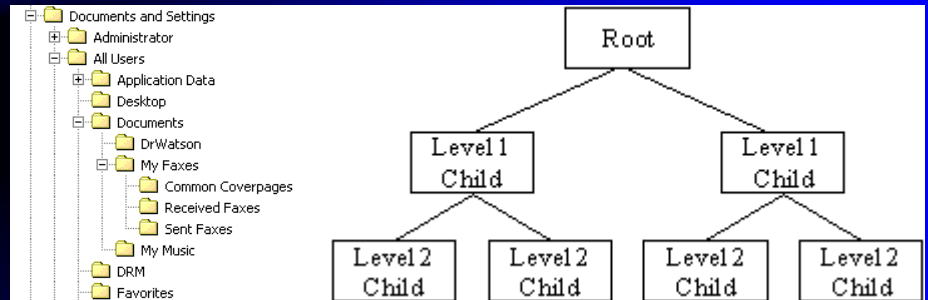
1. Estratificação estuarina
2. Constantes harmónicas
3. Tempo de residência
4. Número de estuário

1. Matéria particulada em suspensão
2. Taxas de produção primária
3. Concentração de N e P
4. Cargas de nutrientes
5. Batimetria



1. Estimativa da produtividade global do sistema
 2. Compreensão do papel de N e P na regulação da prod. 1aria
 3. Análise da remoção de nutrientes
-

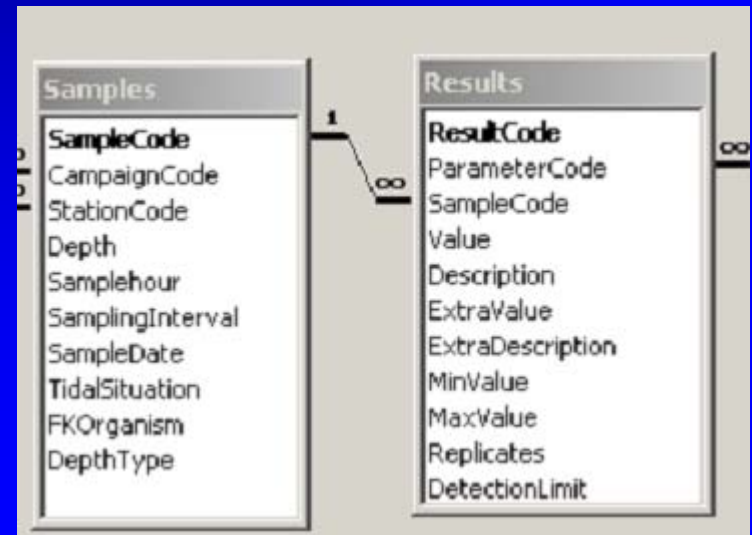
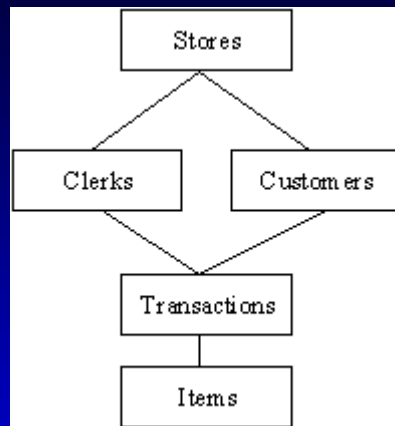
Tipos de bases de datos



● Hierárquica

● Distribuida

● Relacional



<http://www.extropia.com/tutorials/sql/>

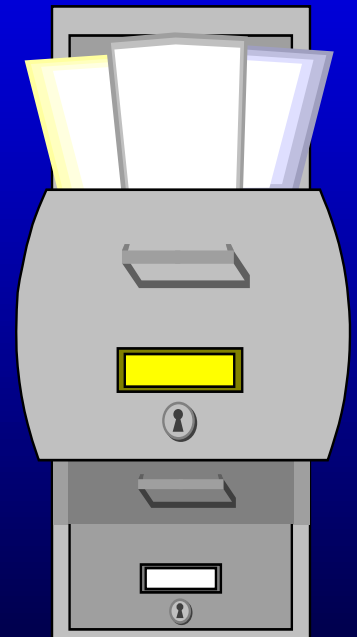
Key features of a successful database

- **Security of data for individual participants**
- **Clearly defined rules for access**
- **User-friendly software**
- **Net gain from investment by participants in DBMS**
- **Flexible (semi-automatic) data input**
- linkage to habitual software used
- **Output not limited to pre-defined options - export of data**



Advantages in the use of a database

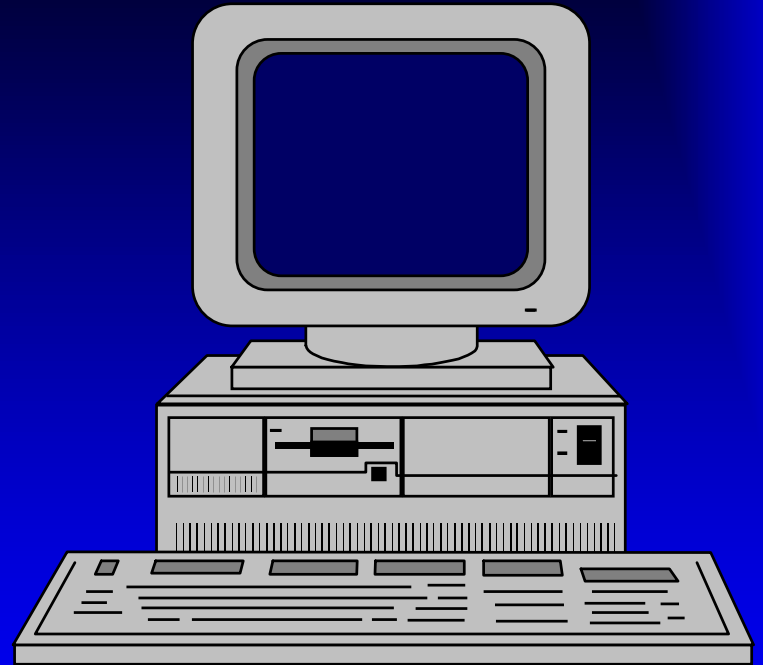
- **Rapid access to all data collected**
- **Capacity for rapid validation tests on the dataset**
- **Effective storage of data in an integrated system**
- **Conditional search operations on the dataset**
- **Evaluation needs of/from researchers, and adaptative change for a sampling program or required data set**
- **Interactive web search**



Relational database design

Normal form rules

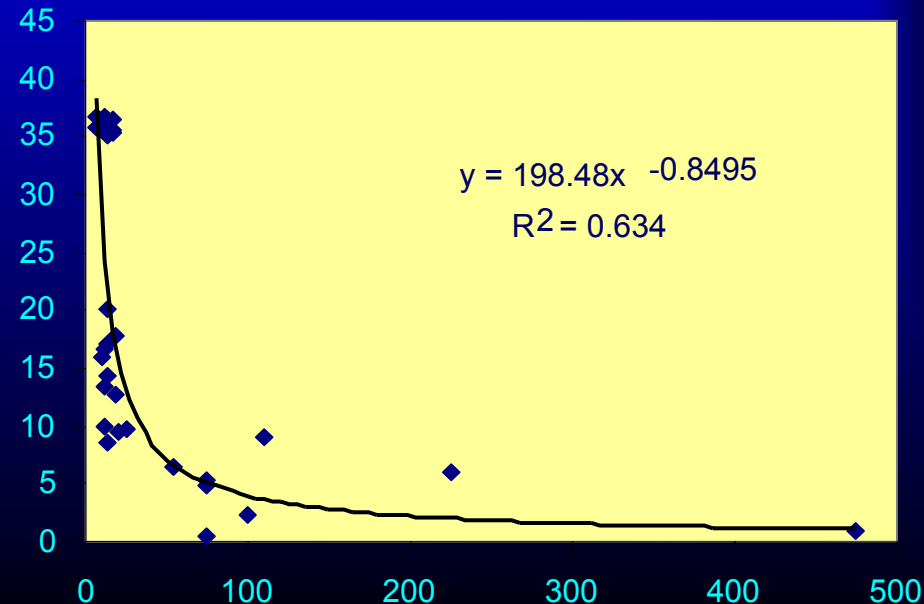
- **1NF: Only one data item may occur at a row-column intersection**
- **2NF: No non-key field may depend solely on part of a primary key (composite keys)**
- **3NF: Any non-key field alteration will not affect other non-key fields**



What information do we want to retrieve?

Example: Usage in the OAERRE project

- Archiving historical and project data
- Generation of water quality data for GIS
- Definition of homogeneous model boxes
- Determination of initial conditions
- Analysis of data relationships for models (e.g. N/P ratios)
- Deriving relationships (e.g. SPM and POC)
- Model calibration and validation



S. Francisco Bay United States Geological Survey



- **Collects and organizes water quality data from a large number of sampling programs**
- **Provides distributed access to raw data**
- **Provides data assimilation mechanisms**
- **Uses a web-based interface**
- **Has no cost restrictions**
- **Fairly easy to use**

S. Francisco Bay

S. Pablo Bay

Suisun Bay

South Bay



USGS Water Quality Database

Water Quality of San Francisco Bay

Step By Step Query Construction Zone -- Step 1 of 3: Parameter Selection

Choose the parameters for which you want to acquire data. Click on the highlighted link to see helpful information about the parameter's checkbox () to choose that data type. Then click the button at the bottom to continue on to the next step.

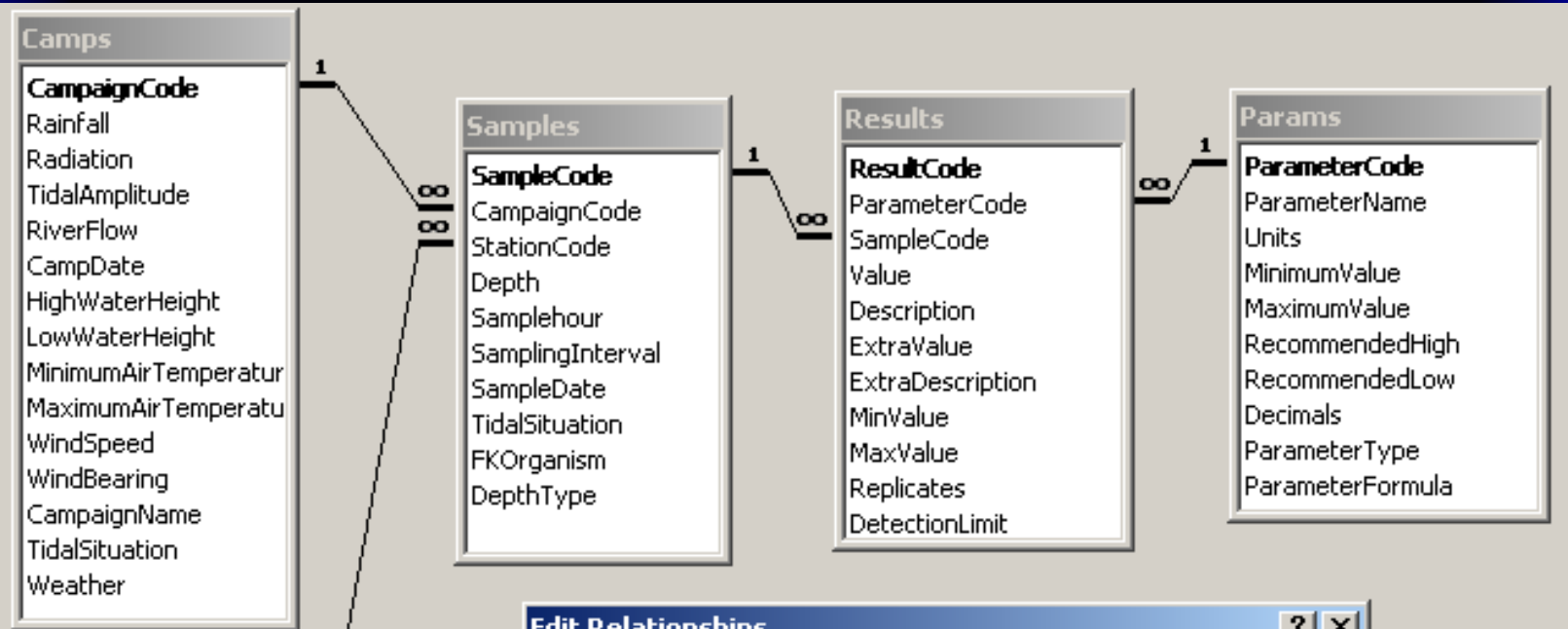
Date & Time Info	CTD Data	Other Data
<input type="checkbox"/> Date (MM/DD/YYYY)	<input type="checkbox"/> Discrete Chlorophyll	<input type="checkbox"/> Delta Outflow Index
<input type="checkbox"/> Julian Date (YYYYDDD)	<input type="checkbox"/> Chlorophyll a/a+PHA ratio	Nutrients:
<input type="checkbox"/> Days since 01/01/1990	<input type="checkbox"/> Fluorescence	<input type="checkbox"/> Nitrite
<input type="checkbox"/> Decimal Date	<input type="checkbox"/> Calculated Chlorophyll	<input type="checkbox"/> Nitrate + Nitrite
<input type="checkbox"/> Time of Day	<input type="checkbox"/> Discrete Oxygen	<input type="checkbox"/> Ammonium
	<input type="checkbox"/> Oxygen electrode output	<input type="checkbox"/> Phosphate
	<input type="checkbox"/> Oxygen Saturation %	<input type="checkbox"/> Silicate
Location Info	<input type="checkbox"/> Calculated Oxygen	
<input type="checkbox"/> Station Number	<input type="checkbox"/> Discrete SPM	
<input type="checkbox"/> Distance from Station 36	<input type="checkbox"/> Optical Backscatter	
<input type="checkbox"/> Depth	<input type="checkbox"/> Calculated SPM	
	<input type="checkbox"/> Measured Excoeff	
	<input type="checkbox"/> Calculated Excoeff	
	<input type="checkbox"/> Salinity	
	<input type="checkbox"/> Temperature	
	<input type="checkbox"/> Sigma-t	
<input type="button" value="→ Continue to step 2 →"/>		

Layout of BarcaWin2000 tables (abridged)

Types of relationships and primary keys

Table	Primary key field(s)	Relation to table	Relation type	Comments
Stations	Station code	Samples	One to many	Primary key
Campaigns	Campaign code	Samples	One to many	Primary key
Samples	Sample code	Results	One to many	Primary key
		Stations	Many to one	Foreign key
		Campaigns	Many to one	Foreign key
Parameters	Parameter code	Results	One to many	Primary key
Results	Sample code and	Samples	Many to one	Composite
	Parameter code	Parameters	Many to one	primary key

BarcaWin2000



Stations
StationCode
CurrentName
Descriptor
Latitude
Longitude
MilitaryLatitude
MilitaryLongitude
UseForIsolines
Depth
TidalDelay

Edit Relationships [?] [X]

Table/Query: Params Related Table/Query: Results

ParameterCode ParameterCode

Enforce Referential Integrity

Cascade Update Related Fields

Cascade Delete Related Records

Relationship Type: One-To-Many

Buttons: OK, Cancel, Join Type.., Create New..