

# Introduction to database management with open source tools

Guillaume Larocque  
research professional,

Quebec Center for Biodiversity Science

<http://qcbs.ca/wiki/opendb>



CENTRE DE LA SCIENCE DE LA BIODIVERSITÉ DU QUÉBEC  
QUEBEC CENTRE FOR BIODIVERSITY SCIENCE

<http://registration.qcbs.ca/pay>



# Introduction to database management with open source tools

Guillaume Larocque  
research professional,

Quebec Center for Biodiversity Science

<http://qcbs.ca/wiki/opendb>



CENTRE DE LA SCIENCE DE LA BIODIVERSITÉ DU QUÉBEC  
QUEBEC CENTRE FOR BIODIVERSITY SCIENCE

<http://registration.qcbs.ca/pay>



# Objectives

- Understand what databases do
- Concepts of database management
- Intro to the SQL language
- Brief intro to LibreOffice Base
- Connecting PostgreSQL with other tools
- Where to find more info?

# Getting started...

## What is a database?

- An organized collection of data.



## Options for database management



## Benefits of open-source software

- Free today, free tomorrow.
- Collaborative development.
- Collaborative help system.
- Links more easily with other open source software.
- Disadvantages: not always good with user interface. No support from company.



# what is a database?

- An organized collection of data.

## What is a relational database?

- Data is stored in formatted tables.
- Tables are linked together with "keys".



ID	Name	Age
1	John	25
2	Jane	30
3	Bob	22

ID	Address
1	123 Main St
2	456 Oak St
3	789 Pine St

ID	Phone
1	555-1234
2	555-5678
3	555-9012

## What is a DBMS?

- Database Management System.
- Software.
- Controls the creation, maintenance, and use of a database.
- Allows concurrent access by several users and applications.
- Does not necessarily come with a user interface.
- **RDBMS** - Relational.

## What is SQL?

- Structured query language.
- Often pronounced 'sequel'.
- Programming language designed for RDBMS.
- Standards-based.
- Used by most RDBMS systems.
- Variations between implementations.

# what is a relational database?

- Data is stored in formatted tables.
- Tables are linked together with "keys".

QCBS\_students

Student id	Firstname	Lastname	University Institution affiliation	Project title in	Supervisor
1	Tyler	Reynold	1	Spittle: Simulations of infectious disease systems...	24
2	Thibault	Jean-L	2	Molecular interactions of adenoviral mycoplasma f...	42
3	Kivler	Clairinda	1	Adaptation as a spatiotemporal result of natural s...	5
4	Maheshwari	Arvind	4	AMGL	21
4	Maheshwari	Arvind	7	Functional responses of insect vegetation to insects...	2
7	Montgo	Lina Beaulé	1	Microbiological Variation of the Red-bellied Vole (P...	32
8	Magnus	Dan	1	AMGL	29
9	Malick	Bergson	4	W	35
10	Loren	Richard Mark	1	Questioning changes at the northern limit of t...	24
11	Sharma	Brenda	1	Phishing for established sources of knowledge...	6
12	Dominic	Chambers	1	Modeling tree abundance in eastern North America...	25
13	Gray	Christos	1	Phylocladogram of invasive species	4
14	Paul	Edward	1	managing invasive species	4
15	Tammy	Choi	1	Phylogenetic diversity of edge communities	63
16	Maria-Julia	Felix	1	The ecological and developmental genetic basis of...	19
17	Suzanne	Gagnon	2	Phylogenetic relationships and biogeography of Co...	19
18	Melissa	Gillett	7	AMGL	23
19	Natalie	James	1	Modeling the effects of climate change on the di...	25
20	Uma	Jones	1	The role of biotic and abiotic factors in exotic s...	36
21	Joseph	Hector de Gomez	7	AMGL	33
22	Melissa	Lacey	1	Oral development with house eggs in the scypha...	19
23	Christine	Molloy	2	Impacto des atividades humanas nos rios da bacia h...	29
24	Georgina	Ortiz	1	The Ecological Causes and Consequences of the Mex...	37

Universities

University Institution affiliation id	name	PROJECT_ID
1	NOVA University	4425
2	Université de Montréal	4427
3	Université de Québec à Montréal	4426
4	Université de Sherbrooke	4421
5	Université de la Gaspésie	4424
6	Université de la Côte-Nord	4423
7	Université de la Nouvelle-Écosse	4422
8	Université de la Nouvelle-Écosse	4423
9	Agriculture & Agri-Food Canada	4520
10	Science Canada/Le Jardin	4522
11	Ministry of Development, Studies, and Technical...	4523
12	York University	4526
13	Université d'agriculture pour le développement (U...	0
14	Université de la Nouvelle-Écosse	3025
15	Institut de recherche pour le développement (IRD)	3727
16	INRA	0
17	University of British Columbia	4529
18	University of Manitoba	4423
19	Université de Québec à Trois-Rivières	4425

Supervisors

member id	Firstname	Lastname	University Institution affiliation	research in
1	Boudin	Benoit	2	My second program is currently limited on how ...
2	Monica	Paul	7	in Web-based management, natural and conserva...
3	Jean-Marc	Reynold	3	NGLL
3	Andrew	Hardy	1	Danah suggested that inclusion proceed very slow...
4	Wesley	Leung	1	in Ecological restoration in Ecology of d...
7	Clara	Lawson	7	The Research Laboratory on Invasive Plants (RLIP)...
8	William	Le-Pyong	7	NGLL
8	Dominique	Frankel	2	NGLL
10	John	Johnson	1	My second program focuses on the evolution of d...
11	Bernad	Anges	2	NGLL
12	Griffin	Neil	1	major research themes in Adaptive net...
13	Clara	Benoist	1	Research in the Desert: its carries around quark...
13	Christophe	Berthou	5	NGLL
15	Nicolas	Breton	2	NGLL
17	Joseph	Breton	2	in Web-based management, natural and conserva...
18	Luca	Brogini	2	NGLL
18	Anne	Bronson	2	in Molecular systematics of legume and the flow...
20	Christopher	Budd	1	in Web-based management in Ecology of d...

# QCBS\_students

student_id ▲	Firstname	Lastname	University_Institution_affiliation	Project_title_en	Supervisor
1	Tyler	Bonnell	1	Spatial simulations of infectious disease: environ...	24
2	Youssef	Ismail	2	Molecular interactions of arbuscular mycorrhizal f...	42
3	Kiyoko	Gotanda	1	Adaptation as a spatiotemporal mosaic of natural a...	5
4	Marie-Eve	André	4	NULL	21
6	Marianne	Bachand	7	Functional response of boreal vegetation to overab...	2
7	Rodrigo	Lima Barata	1	Morphological Variation of the Red-backed Vole (My...	52
8	Magnus	Bein	1	NULL	29
9	Patrick	Bergeron	4	0	35
10	Laura	Boisvert-Marsh	1	Spatiotemporal changes at the northern limit of tr...	28
11	Johanna	Bradie	1	Predicting the establishment success of non-indige...	6
12	Dominic	Chambers	1	Modeling tree abundance in eastern North America i...	28
13	Corey	Chivers	1	Predicting spread of invasive species	6
14	Paul	Edwards	1	managing invasive species	6
15	Tammy	Elliot	1	Phylobetadiversity of sedge communities	65
16	Marie-Julie	Favé	1	The ecological and developmental genetic basis of ...	10
17	Edeline	Gagnon	2	Phylogenetic relationships and biogeography of Cae...	19
18	Melissa	Girard	7	NULL	33
19	Natalie	James	1	Modelling the effects of climate change on the dis...	28
20	Lisa	Jones	1	The role of biotic and abiotic factors in exotic s...	56
21	Joseph	Moisan de Serres	7	NULL	33
22	Maryna	Lesoway	1	Direct development with nurse eggs in the calyptra...	10
23	Chantale	Moisan	2	Impacts des activités humaines sur <i>Arethusa bul...	75
24	Georgina	O'Farrill	1	The Ecological Causes and Consequences of the Move...	37

<b>university_insitution_affiliation_id</b>	<b>name</b>	<b>FRQNT_ID</b>
1	McGill University	44563
2	Université de Montréal	44607
3	Université du Québec à Montréal	44986
4	Université de Sherbrooke	44811
5	Université du Québec à Rimouski	44988
6	Concordia University	46092
7	Université Laval	44501
8	Bishop's University	45674
9	Agriculture & Agroalimentaire Canada	45080
10	Service canadien des forêts	45809
11	Ministère du Développement durable, de l'Environne...	48535
12	Yale University	45066
13	La recherche agronomique pour le développement (Ci...	0
14	Duke University	46724
15	Institut de recherche pour le développement (IRD)	47712
16	NA	0
17	University of British Columbia	45729
18	University of Manitoba	44933
19	Université du Québec à Trois-Rivière	44993



# Supervisors

member_id ▲	Firstname	Lastname	University_Institution_affiliation	research_en
1	Beatrix	Beisner	3	My research program is currently centred on three ...
2	Monique	Poulin	7	<li>Wetland management, restoration and conservati...
4	Jean-Pierre	Revéret	3	NULL
5	Andrew	Hendry	1	Darwin suggested that evolution proceeds very slow...
6	Brian	Leung	1	<li>Biological invasions</li> <li>Ecology of dise...
7	Claude	Lavoie	7	The Research Laboratory on Invasive Plants (RELIP)...
8	Philippe	Le Prestre	7	NULL
9	Bernadette	Pinel-Alloul	2	NULL
10	Ehab	Abouheif	1	My research program focuses on the evolution of de...
11	Bernard	Angers	2	NULL
12	Graham	Bell	1	Major Research Themes:  <b>Adaptive radiat...
13	Elena	Bennett	1	Research in the Bennett lab centers around questio...
15	Dominique	Berteaux	5	NULL
16	Jacques	Brisson	2	NULL
17	Jacques	Brodeur	2	<li>Plant-insect interactions and biological contr...
18	Luc	Brouillet	2	NULL
19	Anne	Bruneau	2	<li>Molecular systematics of legumes and the Rosa ...
20	Christopher	Buddle	1	<li>Arctic Biodiversity</li> <li>Canopy Arthropod...

# what is a relational database?

- Data is stored in formatted tables.
- Tables are linked together with "keys".

QCBS\_students

Student id	Firstname	Lastname	University Institution affiliation	Project title in	Supervisor
1	Tyler	Reynold	1	Spittle: Simulations of infectious disease systems...	24
2	Thibault	Jean-L	2	Molecular interactions of adenoviral mycoplasma f...	42
3	Kivler	Clairinda	1	Adaptation as a spatiotemporal result of natural s...	5
4	Maheshwari	Arvind	4	AMGL	21
4	Maheshwari	Arvind	7	Functional responses of insect vegetation to insects...	2
7	Montoya	Lina Beatriz	1	Microbiological Variation of the Red-bellied Vole (P...	32
8	Magnus	Dan	1	AMGL	29
9	Mallick	Bergsten	4	W	35
10	Laine	Reinhold Martin	1	Questioning changes at the northern limit of t...	24
11	Johnson	Brenda	1	Phenology for established success of nonindigen...	6
12	Cornejo	Charmaine	1	Modeling tree abundance in eastern North America...	25
13	Grigg	Christine	1	Phenology of spread of invasive species	4
14	Paul	Edward	1	Managing invasive species	4
15	Tanner	Cliff	1	Phylogenetic diversity of edge communities	63
16	Maria-Julia	Felix	1	The ecological and developmental genetic basis of...	19
17	Schmitt	Gilbert	2	Phylogenetic relationships and biogeography of C...	19
18	Melissa	Gilbert	7	AMGL	23
19	Nabali	James	1	Modeling the effects of climate change on the di...	25
20	Uma	John	1	The role of biotic and abiotic factors in exotic s...	36
21	Joseph	Robert de Smet	7	AMGL	33
22	Melissa	Lacey	1	Oral development with house eggs in the saproxy...	19
23	Christine	Melissa	2	Impacto des atividades humanas nos rios da Bacia h...	29
24	Georgina	Ortiz	1	The Ecological Causes and Consequences of the M...	37

Universities

University_Institution_Affiliation	name	PROJECT_ID
1	NOVA University	4425
2	Université de Montréal	4427
3	Université de Québec en Maritimes	4426
4	Université de Sherbrooke	4421
5	Université de la Gaspésie Université	4424
6	Université Laval	4423
7	Université Laval	4422
8	University of Guelph	4521
9	Agriculture & Agri-Food Canada	4520
10	Science Canada/Agri-Food	4522
11	Ministry of Development, Studies, and Technol...	4523
12	Agri-Food Canada	4524
13	Université d'agriculture pour le développement (U...	0
14	Université Laval	3005
15	Institut de recherche pour le développement (IRD)	3707
16	INRA	0
17	University of British Columbia	4529
18	University of Manitoba	4423
19	Université de Québec à Trois-Rivières	4425

Supervisors

member id	Firstname	Lastname	University Institution affiliation	research in
1	Benoit	Benoit	2	My second program is currently under review ...
2	Monica	Paul	7	in-Webster management, natural and conserva...
3	Jean-Marc	Reynold	3	NGLL
3	Andrew	Hardy	1	Danah suggested that inclusion proceed very slow...
4	Wesley	Leung	1	in-ecological resources in-ecology of d...
7	Clara	Lawson	7	The Research Laboratory on Invasive Plants (RLIP)...
8	William	Le-Pyost	7	NGLL
8	Dominique	Frankel	2	NGLL
10	John	Johnson	1	My second program focuses on the evolution of d...
11	Benoit	Anges	2	NGLL
12	Griffin	Itai	1	major research themes in-Adaptive net...
13	Clara	Benoit	1	Research in the Desert: its carries around quads...
13	Christian	Berthelet	5	NGLL
15	Nicolas	Benoit	2	NGLL
17	Joseph	Benoit	7	in-Webster management, natural and conserva...
18	Luc	Benoit	2	NGLL
18	Anne	Benoit	2	in-Molecular systematics of legume and the flow...
20	Christopher	Benoit	1	in-Webster management, natural and conserva...

# What is a DBMS?

- Database Management System.
- Software.
- Controls the creation, maintenance, and use of a database.
- Allows concurrent access by several users and applications.
- Does not necessarily come with a user interface.
- **RDBMS** - Relational.

# What is SQL?

- Structured query language.
- Often pronounced 'sequel'
- Programming language designed for RDBMS.
- Standards-based.
- Used by most RDBMS systems.
- Variations between implementations.

# Options for database management



## Microsoft Excel?



- Poor at working with multiple linked tables.
- Poor at working with very large datasets.
- No multi-user access.
- No strict formatting.
- No server access.
- Performing complex queries can be difficult.
- Data and 'reports' are not separated.



## Microsoft Access?



- MS Access is a front-end to a database 'server' called Jet/ACE. It can be used with other database servers such as PostgreSQL.
- Not an open platform.
- Multiple versions and document formats.
- Not frequently used on servers.
- Limits to the file sizes and concurrent users.
- Can make a good front-end for designing forms or generating reports.

## Full-fledged DBMS

- Open source
  - MySQL/MariaDB
  - PostgreSQL
  - SQLite
- Proprietary
  - Oracle
  - Microsoft SQL Server

# Microsoft Excel?



- Poor at working with multiple linked tables.
- Poor at working with very large datasets.
- No multi-user access.
- No strict formatting.
- No server access.
- Performing complex queries can be difficult.
- Data and 'reports' are not separated.

# Microsoft Access?



- MS Access is a front-end to a database 'server' called Jet/ACE. It can be used with other database servers such as PostgreSQL.
- Not an open platform.
- Multiple versions and document formats.
- Not frequently used on servers.
- Limits to the file sizes and concurrent users.
- Can make a good front-end for designing forms or generating reports.

# Full-fledged DBMS

## Open source

- MySQL/MariaDB
- PostgreSQL
- SQLite

## Proprietary

- Oracle
- Microsoft SQL Server



# Benefits of open-Source Software

- Free today, free tomorrow.
- Collaborative development.
- Collaborative help system.
- Links more easily with other open source software.
  
- Disadvantages: not always good with user interface. No support from company.



# MySQL/MariaDB and PostgreSQL

- Open source.
- Extremely powerful.
- Fast.
- Can handle very large datasets.
- Good development.
- Used by major companies (Facebook, Twitter, Google, etc.).
- SQL standard.
- MySQL - developed mostly by Oracle.
- Postgres and MariaDB - community development.
- Database servers with no front-ends.



PostgreSQL



MariaDB

# When Should I use a DBMS?

- Standardization of data is important
- Very large datasets
- Multiple users, multiple platforms
- Access on a server
- Interaction with other software or computer tools

## When I should not use a DBMS?

- Small datasets that don't need standardization
- Single users, simple needs
- Lack of technical expertise/time

### Examples of DBMS use

- QCBS student/prof database and website.
- Des nids chez vous website
- The internet (e.g. <http://quebio.ca/bam>)
- Collaborative databases (GBIF, GENBANK, etc.)

# Examples of DBMS use

- QCBS student/prof database and website.
- Des nids chez vous website
- The internet (e.g. <http://quebio.ca/bam>)
- Collaborative databases (GBIF, GENBANK, etc.)

# Database design principles

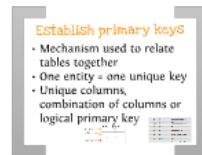


# Basic principles

- Standardization
- Coherence
- Expandability
- Flexibility

# Steps of database design

- Determine entities and tables
- Relationships/cardinality
- Establish primary keys
- Establish data types



# Determine entities and tables

- One table per type of entity: person, type of person, address, sample, quadrat, ...

Author ID	Author	Location	Address, phone, etc. (if any)	Project ID(s)	Year(s)
1	Alan	London		1	1964
2	Frank	London		2	1964
3	Frank	London		3	1964
4	Frank	London		4	1964
5	Frank	London		5	1964
6	Frank	London		6	1964
7	Frank	London		7	1964
8	Frank	London		8	1964
9	Frank	London		9	1964
10	Frank	London		10	1964
11	Frank	London		11	1964
12	Frank	London		12	1964
13	Frank	London		13	1964
14	Frank	London		14	1964
15	Frank	London		15	1964
16	Frank	London		16	1964
17	Frank	London		17	1964
18	Frank	London		18	1964
19	Frank	London		19	1964
20	Frank	London		20	1964

Project ID	Project	Location	Year(s)	Project ID(s)	Year(s)
1	Quadrat	London	1964	1	1964
2	Quadrat	London	1964	2	1964
3	Quadrat	London	1964	3	1964
4	Quadrat	London	1964	4	1964
5	Quadrat	London	1964	5	1964
6	Quadrat	London	1964	6	1964
7	Quadrat	London	1964	7	1964
8	Quadrat	London	1964	8	1964
9	Quadrat	London	1964	9	1964
10	Quadrat	London	1964	10	1964
11	Quadrat	London	1964	11	1964
12	Quadrat	London	1964	12	1964
13	Quadrat	London	1964	13	1964
14	Quadrat	London	1964	14	1964
15	Quadrat	London	1964	15	1964
16	Quadrat	London	1964	16	1964
17	Quadrat	London	1964	17	1964
18	Quadrat	London	1964	18	1964
19	Quadrat	London	1964	19	1964
20	Quadrat	London	1964	20	1964

Project ID	Project	Year(s)
1	Quadrat	1964
2	Quadrat	1964
3	Quadrat	1964
4	Quadrat	1964
5	Quadrat	1964
6	Quadrat	1964
7	Quadrat	1964
8	Quadrat	1964
9	Quadrat	1964
10	Quadrat	1964
11	Quadrat	1964
12	Quadrat	1964
13	Quadrat	1964
14	Quadrat	1964
15	Quadrat	1964
16	Quadrat	1964
17	Quadrat	1964
18	Quadrat	1964
19	Quadrat	1964
20	Quadrat	1964

Project ID	Project
1	Quadrat
2	Quadrat
3	Quadrat
4	Quadrat
5	Quadrat
6	Quadrat
7	Quadrat
8	Quadrat
9	Quadrat
10	Quadrat
11	Quadrat
12	Quadrat
13	Quadrat
14	Quadrat
15	Quadrat
16	Quadrat
17	Quadrat
18	Quadrat
19	Quadrat
20	Quadrat



student_id ▲	Firstname	Lastname	University_Institution_affiliation	Project_title_en	Supervisor
1	Tyler	Bonnell	1	Spatial simulations of infectious disease: environ...	24
2	Youssef	Ismail	2	Molecular interactions of arbuscular mycorrhizal f...	42
3	Kiyoko	Gotanda	1	Adaptation as a spatiotemporal mosaic of natural a...	5
4	Marie-Eve	André	4	NULL	21
6	Marianne	Bachand	7	Functional response of boreal vegetation to overab...	2
7	Rodrigo	Lima Barata	1	Morphological Variation of the Red-backed Vole (My...	52
8	Magnus	Bein	1	NULL	29
9	Patrick	Bergeron	4	0	35
10	Laura	Boisvert-Marsh	1	Spatiotemporal changes at the northern limit of tr...	28
11	Johanna	Bradie	1	Predicting the establishment success of non-indige...	6
12	Dominic	Chambers	1	Modeling tree abundance in eastern North America i...	28
13	Corey	Chivers	1	Predicting spread of invasive species	6
14	Paul	Edwards	1	managing invasive species	6
15	Tammy	Elliot	1	Phylobetadiversity of sedge communities	65
16	Marie-Julie	Favé	1	The ecological and developmental genetic basis of ...	10
17	Edeline	Gagnon	2	Phylogenetic relationships and biogeography of Cae...	19
18	Melissa	Girard	7	NULL	33
19	Natalie	James	1	Modelling the effects of climate change on the dis...	28
20	Lisa	Jones	1	The role of biotic and abiotic factors in exotic s...	56
21	Joseph	Moisan de Serres	7	NULL	33
22	Maryna	Lesoway	1	Direct development with nurse eggs in the calyptra...	10
23	Chantale	Moisan	2	Impacts des activités humaines sur <i>Arethusa bul...	75
24	Georgina	O'Farrill	1	The Ecological Causes and Consequences of the Move...	37

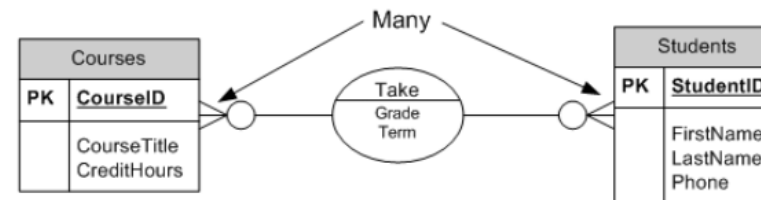
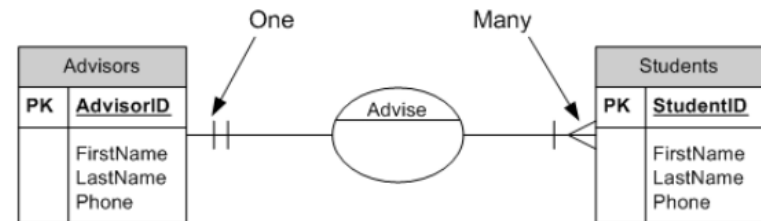
member_id ▲	Firstname	Lastname	University_Institution_affiliation	research_en
1	Beatrix	Beisner	3	My research program is currently centred on three ...
2	Monique	Poulin	7	<li>Wetland management, restoration and conservati...
4	Jean-Pierre	Revéret	3	NULL
5	Andrew	Hendry	1	Darwin suggested that evolution proceeds very slow...
6	Brian	Leung	1	<li>Biological invasions</li> <li>Ecology of dise...
7	Claude	Lavoie	7	The Research Laboratory on Invasive Plants (RELIP)...
8	Philippe	Le Prestre	7	NULL
9	Bernadette	Pinel-Alloul	2	NULL
10	Ehab	Abouheif	1	My research program focuses on the evolution of de...
11	Bernard	Angers	2	NULL
12	Graham	Bell	1	Major Research Themes:  <b>Adaptive radiat...
13	Elena	Bennett	1	Research in the Bennett lab centers around questio...
15	Dominique	Berteaux	5	NULL
16	Jacques	Brisson	2	NULL
17	Jacques	Brodeur	2	<li>Plant-insect interactions and biological contr...
18	Luc	Brouillet	2	NULL
19	Anne	Bruneau	2	<li>Molecular systematics of legumes and the Rosa ...
20	Christopher	Buddle	1	<li>Arctic Biodiversity</li> <li>Canopy Arthropod...

<b>university_insitution_affiliation_id</b>	<b>name</b>	<b>FRQNT_ID</b>
1	McGill University	44563
2	Université de Montréal	44607
3	Université du Québec à Montréal	44986
4	Université de Sherbrooke	44811
5	Université du Québec à Rimouski	44988
6	Concordia University	46092
7	Université Laval	44501
8	Bishop's University	45674
9	Agriculture & Agroalimentaire Canada	45080
10	Service canadien des forêts	45809
11	Ministère du Développement durable, de l'Environne...	48535
12	Yale University	45066
13	La recherche agronomique pour le développement (Ci...	0
14	Duke University	46724
15	Institut de recherche pour le développement (IRD)	47712
16	NA	0
17	University of British Columbia	45729
18	University of Manitoba	44933
19	Université du Québec à Trois-Rivière	44993

<b>keyword_id</b>	<b>keyword</b>
1	biodiversity
2	ecology
3	ecosystem
4	ecosystem services
5	population ecology
6	genetics
7	ecosystem functioning
8	forest ecology
9	phylogenetics
10	population genetics
11	geographic information systems
12	generalized linear models
13	limnology
14	forestry
15	bayesian analysis
16	spatial statistics

# Establish relationships/ cardinality

- One to one
- One to many
- Many to many



## Storing one to many relationships

### Option 1 - Multiple columns

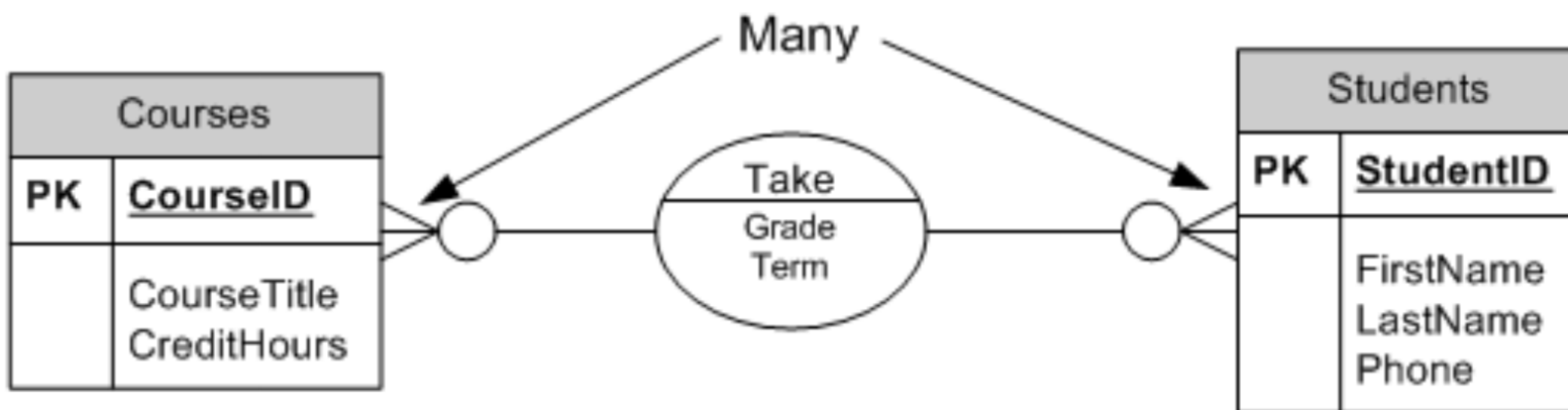
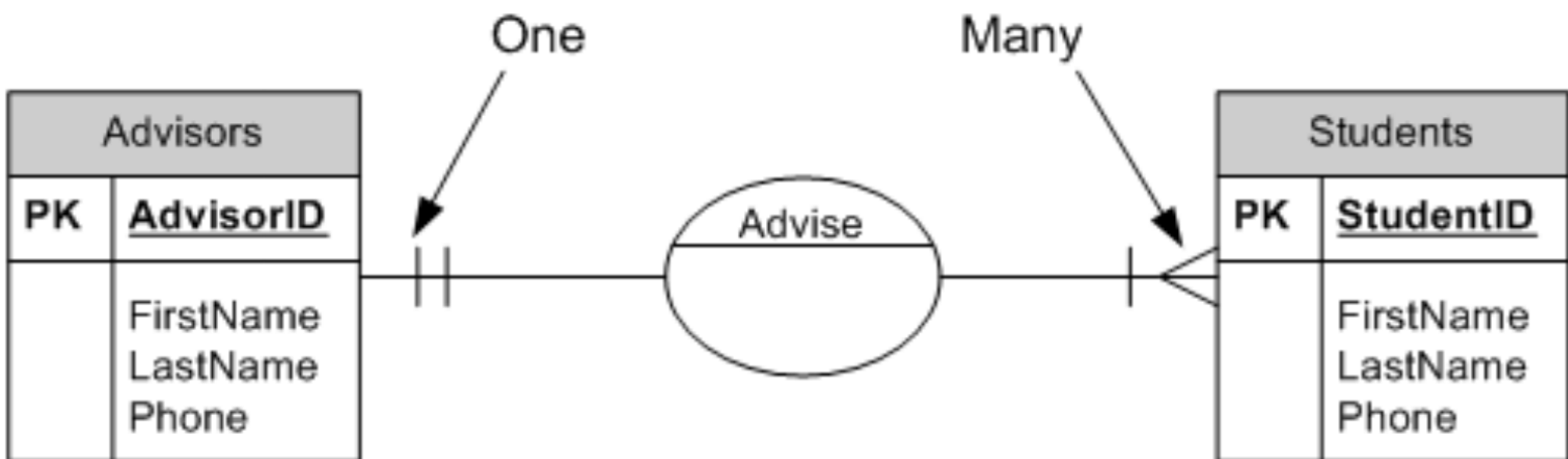
AdvisorID	StudentID	AdvisorID	StudentID
1	1	2	2
1	2	3	3

### Option 2 - Lists

AdvisorID	StudentID
1	1,2
2	2,3

### Option 3 - Lookup table

AdvisorID	StudentID
1	1
1	2
2	2
2	3



# Storing one to many relationships

## Option 1 - Multiple columns

Supervisor	Students1	Students2	Student3
John Smith	Vicky Côté	Alain Lambert	
Alan Gartner	Marc Sauvé		

## Option 2 - Lists

Supervisor	Students
John Smith	Vicky Côté, Alain Lambert
Alan Gartner	Marc Sauvé

## Option 3 - Lookup table

Supervisor	Student
John Smith	Vicky Côté
John Smith	Alain Lambert
Alan Gartner	Marc Sauvé

# Establish primary keys

- Mechanism used to relate tables together
- One entity = one unique key
- Unique columns, combination of columns or logical primary key

Supervisor_id	Supervisor_name
54	John Smith
56	Alan Gartner

Student_id	Student_name
23	Vicky Cook
15	Alain Lambert
12	Marc Sawyer

Supervisor_id	Student_id
54	23
54	15
56	12

Primary keys → (point to Supervisor\_id and Student\_id in the first two tables)

No Key → (points to the first row of the third table)

Module ID	Firstname	Lastname	University_Institution_ID	Project_Title	Supervisor
1	Tyler	Borrell	1	Optical simulations of electron device structures	54
2	Yusef	Shariq	2	Molecular interactions of electron transport in ...	62
3	Kyran	Stewart	1	Adaptation as a spatiotemporal process of spatial a...	5
4	Shahira	Alshay	4	MEMS	21
5	Heidi	Stewart	7	Functional response of forest vegetation to insect...	2
7	George	Lima Barros	1	Morphological Variation of the Red-backed Vireo (Vireo ...	52
8	Magnus	Ryan	1	MEMS	20
9	Patrick	Shannon	4	MEMS	30
10	Liam	Burrows	1	Spatiotemporal changes of the northern limit of V...	20
11	John	Finlay	1	Understanding the establishment success of introduced...	5
12	Dennis	Chapman	1	Modeling the abundance of western North American...	30
13	Corey	Oliver	1	Modeling spread of invasive species	5
14	Paul	Kawachi	1	Nonlinear invasive species	4
15	Tiffany	Shah	1	Impacts of land-use change on conservation	30
16	Maria-Julia	Faria	1	The ecological and developmental genetic basis of ...	10
17	Esther	Griffin	2	Phylogenetic relationships and biogeography of Oen...	10
18	Isabelle	Chen	7	MEMS	10
19	Nikola	Jones	1	Modeling the effects of climate change on the di...	20
20	Lisa	Jones	1	The role of biotic and abiotic factors in water l...	30
21	Joseph	Victor de Sa	7	MEMS	10
22	Mayra	Levy	1	Direct development with ruse eggs in the catfish...	10
23	Chantal	Moham	2	Genetic diversity in human populations in the ...	10
24	Georgina	Chen	1	The Biological Causes and Consequences of the Green...	10



# Primary Key

Primary  
keys



Supervisor_id	Supervisor_name
54	John Smith
58	Alan <u>Gartner</u>

Student_id	Student_name
23	Vicky <u>Côté</u>
15	<u>Alain</u> Lambert
12	Marc <u>Sauvé</u>

Supervisor_id	Student_id
54	23
54	15
58	12

No Key



student_id ▲	Firstname	Lastname	University_Institution_affiliation	Project_title_en	Supervisor
1	Tyler	Bonnell	1	Spatial simulations of infectious disease: environ...	24
2	Youssef	Ismail	2	Molecular interactions of arbuscular mycorrhizal f...	42
3	Kiyoko	Gotanda	1	Adaptation as a spatiotemporal mosaic of natural a...	5
4	Marie-Eve	André	4	NULL	21
6	Marianne	Bachand	7	Functional response of boreal vegetation to overab...	2
7	Rodrigo	Lima Barata	1	Morphological Variation of the Red-backed Vole (My...	52
8	Magnus	Bein	1	NULL	29
9	Patrick	Bergeron	4	0	35
10	Laura	Boisvert-Marsh	1	Spatiotemporal changes at the northern limit of tr...	28
11	Johanna	Bradie	1	Predicting the establishment success of non-indige...	6
12	Dominic	Chambers	1	Modeling tree abundance in eastern North America i...	28
13	Corey	Chivers	1	Predicting spread of invasive species	6
14	Paul	Edwards	1	managing invasive species	6
15	Tammy	Elliot	1	Phylobetadiversity of sedge communities	65
16	Marie-Julie	Favé	1	The ecological and developmental genetic basis of ...	10
17	Edeline	Gagnon	2	Phylogenetic relationships and biogeography of Cae...	19
18	Melissa	Girard	7	NULL	33
19	Natalie	James	1	Modelling the effects of climate change on the dis...	28
20	Lisa	Jones	1	The role of biotic and abiotic factors in exotic s...	56
21	Joseph	Moisan de Serres	7	NULL	33
22	Maryna	Lesoway	1	Direct development with nurse eggs in the calyptra...	10
23	Chantale	Moisan	2	Impacts des activités humaines sur <i>Arethusa bul...	75
24	Georgina	O'Farrill	1	The Ecological Causes and Consequences of the Move...	37

# Establish data types

- Has to allow all possible current and future values.

## String types

**CHAR**( ) A fixed section from 0 to 255 characters long.

**VARCHAR**( ) A variable section from 0 to 255 characters long.

**TEXT** A string with a maximum length of 65535 characters.

## Numeric types

**TINYINT**( ) -128 to 127 normal 0 to 255 UNSIGNED.

**SMALLINT**( ) -32768 to 32767 normal 0 to 65535 UNSIGNED.

**MEDIUMINT**( ) -8388608 to 8388607 normal 0 to 16777215 UNSIGNED.

**INT**( ) -2147483648 to 2147483647 normal 0 to 4294967295 UNSIGNED.

**BIGINT**( ) -9223372036854775808 to 9223372036854775807 normal 0 to 18446744073709551615 UNSIGNED.

**FLOAT** A small number with a floating decimal point.

**DOUBLE**( , ) A large number with a floating decimal point.

**DECIMAL**( , ) A DOUBLE stored as a string , allowing for a fixed decimal point.

## Date-time

**DATE** YYYY-MM-DD.

**TIME** HH:MM:SS.

**TIMESTAMP** YYYY-MM-DD HH:MM:SS

## NULL

Used for unknown attributes

# current and futur

## String types

**CHAR( )** A fixed section from 0 to 255 characters long.

**VARCHAR( )** A variable section from 0 to 255 characters long.

**TEXT** A string with a maximum length of 65535 characters.

## Date-time

**DATE** YYYY MM DD

# Numeric types

**TINYINT**( ) -128 to 127 normal 0 to 255 UNSIGNED.

**SMALLINT**( ) -32768 to 32767 normal 0 to 65535 UNSIGNED.

**MEDIUMINT**( ) -8388608 to 8388607 normal 0 to 16777215 UNSIGNED.

**INT**( ) -2147483648 to 2147483647 normal 0 to 4294967295 UNSIGNED.

**BIGINT**( ) -9223372036854775808 to 9223372036854775807 normal  
0 to 18446744073709551615 UNSIGNED.

**FLOAT** A small number with a floating decimal point.

**DOUBLE**( , ) A large number with a floating decimal point.

**DECIMAL**( , ) A DOUBLE stored as a string , allowing for a fixed decimal point.

## NULL

Used for unknown attributes

# Date-time

**DATE** YYYY-MM-DD.

**TIME** HH:MM:SS.

**TIMESTAMP** YYYY-MM-DD HH:MM:SS

# NULL

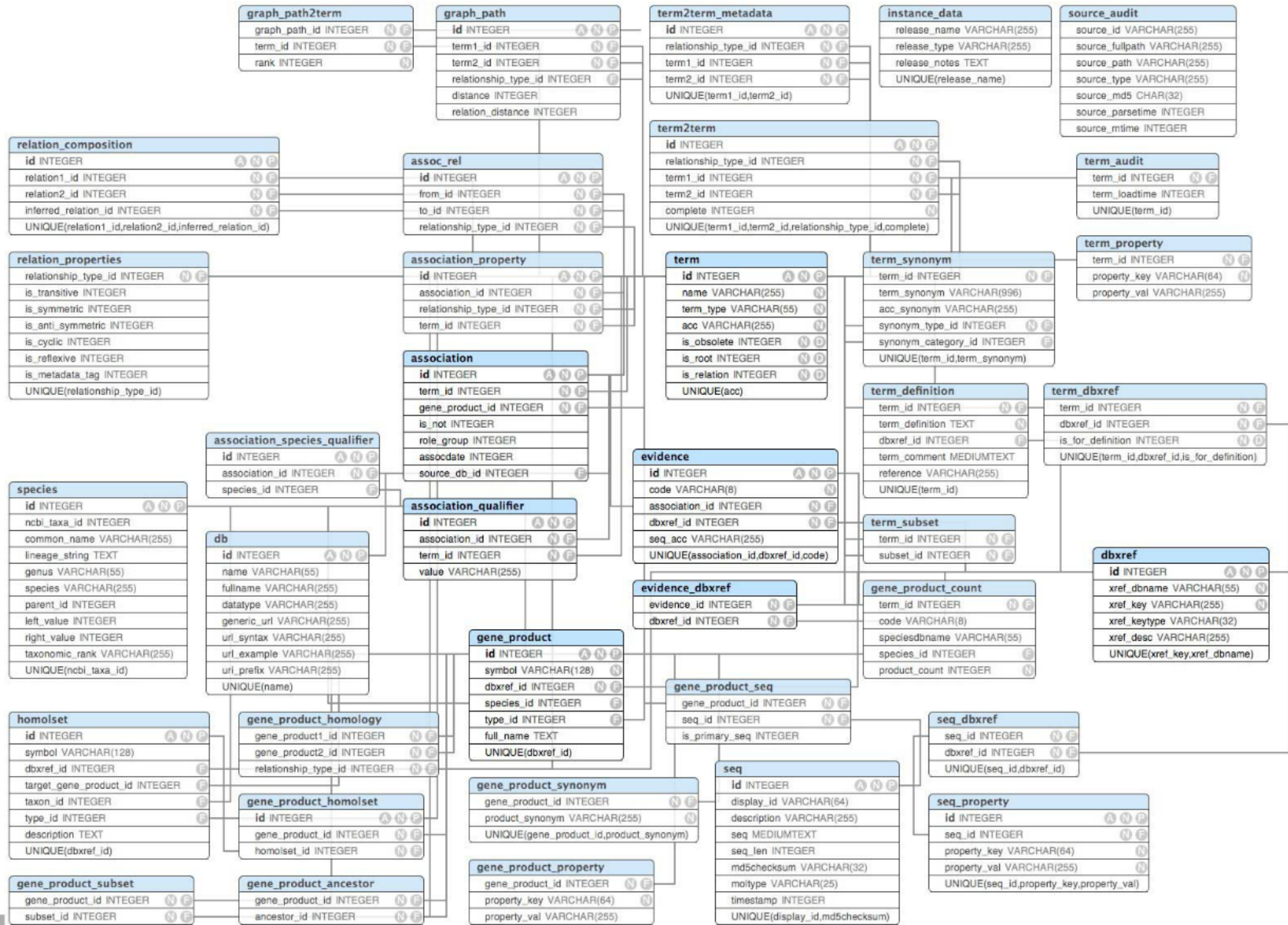
Used for unknown attributes

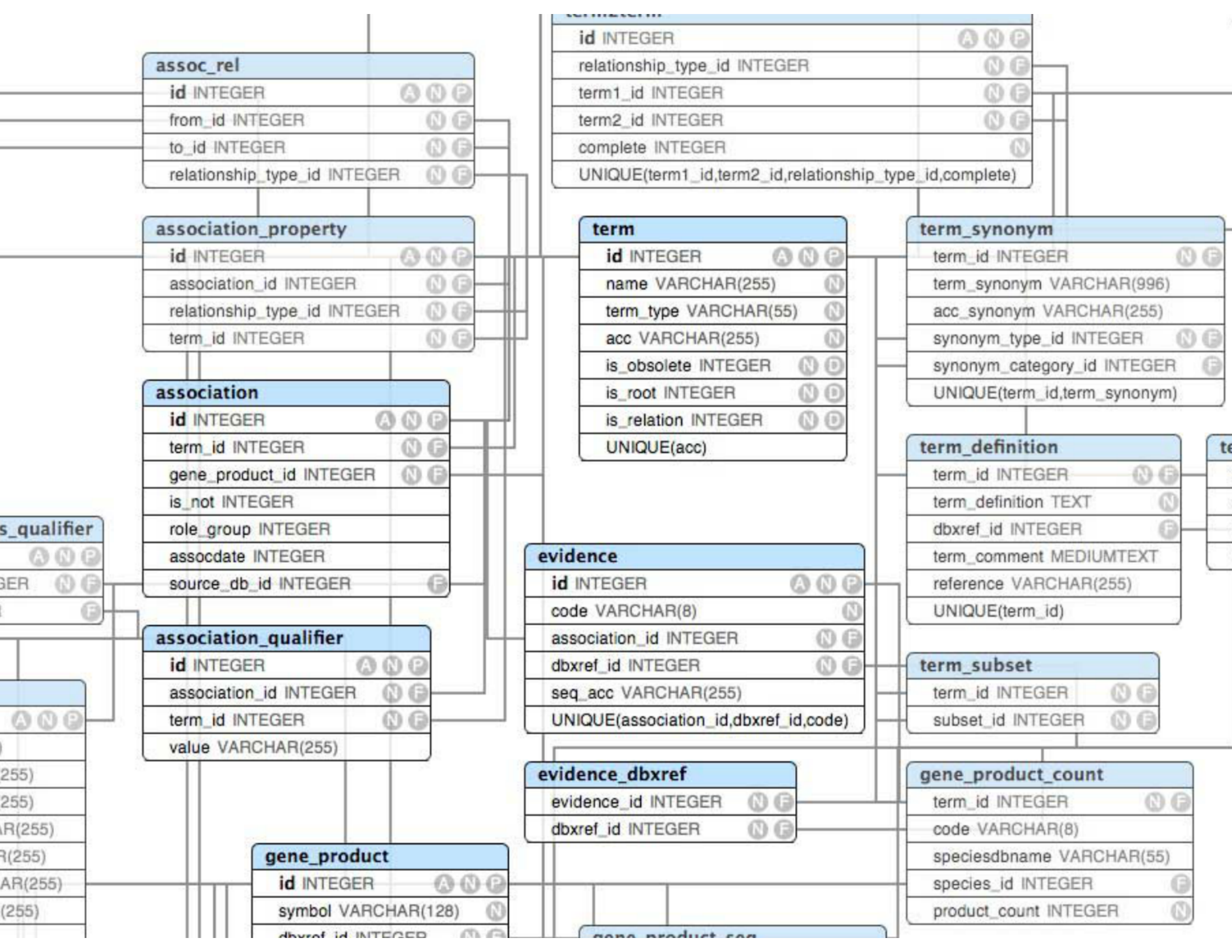
# Rules of thumb

- One attribute per column. Cannot be subdivided.
- Unique way to identify each row.
- No repeating info.
- Simple but meaningful table and column names. Avoid words such as 'name', 'text', 'count', 'long', etc.



# Entity-relationship diagram





# Basics of the PostgreSQL language

## User interface

- Command line client
- LibreOffice BASE/ MS Access
- PgAdmin III
- Phppgadmin

## Entering data...

- Create forms (Access, LibreOffice, PHP)
- Edit directly in user interface (e.g. LibreOffice Base)
- INSERT command

## Importing data...

- From LibreOffice/Access (csv/ods)
- From the PostgreSQL command line (CREATE TABLE COPY...)

## Basic commands

```
CREATE TABLE table_name (column_name data_type);  
INSERT INTO table_name (column_name) VALUES (value);  
UPDATE table_name SET column_name = value;  
DELETE FROM table_name;  
DROP TABLE table_name;
```

## Operators

```
SELECT column_name FROM table_name WHERE condition;  
JOIN ... GROUP BY column_name HAVING condition;  
ORDER BY column_name
```

## Basics

- Table, column names separated by commas, not commands
- Strings are separated by "
- No spaces, capital letters or special characters \$&?%!...
- Names can be enclosed with "" when spaces or special characters are present or for reserved words.
- Not case-sensitive. But always use good case.

## USER interface

- Command line client
- LibreOffice BASE/ MS Access
- PgAdmin III
- Phppgadmin

# Basics

- Table, column names separated by commas, not commands
- Strings are separated by "
- No spaces, capital letters or special characters \$&?!...
- Names can be enclosed with "" when spaces or special characters are present or for reserved words.
- Not case-sensitive. But always use good case.

## Entering data...

- Create forms (Access, LibreOffice, PHP)
- Edit directly in user interface (e.g. Libreoffice Base)
- INSERT command

## Importing data...

- From Libreoffice/Access (csv,xls,ods)
- From the PostgreSQL command line (CREATE TABLE, COPY...)

# Basic Commands

**\l** list tables in database  
**\d+** describe table and columns  
**CREATE** create database or table  
**INSERT** insert line(s) into table  
**SELECT** make queries  
**UPDATE** modify values  
**DELETE** delete lines  
**ALTER** add columns or modify format  
**DROP** delete table or database

# Operators

**SELECT** columns **FROM** tables **WHERE** conditions  
**JOIN ... GROUP BY** columns **HAVING** condition  
**ORDER BY** columns

# Basic Commands

- `\l` list tables in database
- `\d+` describe table and columns
- CREATE** create database or table
- INSERT** insert line(s) into table
- SELECT** make queries
- UPDATE** modify values
- DELETE** delete lines
- ALTER** add columns or modify format
- DROP** delete table or database

# Operators



**UPDATE** modify values

**DELETE** delete lines

**ALTER** add columns or modify format

**DROP** delete table or database

# Operators

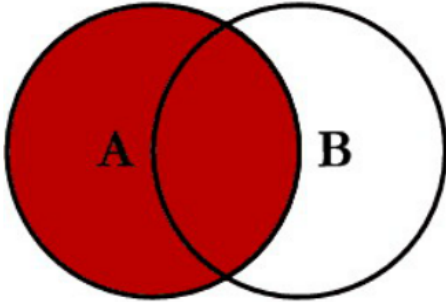
**SELECT** columns **FROM** tables **WHERE** conditions

**JOIN ... GROUP BY** columns **HAVING** condition

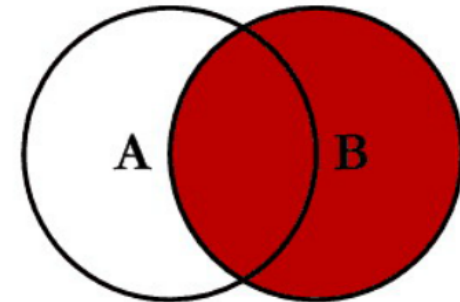
**ORDER BY** columns



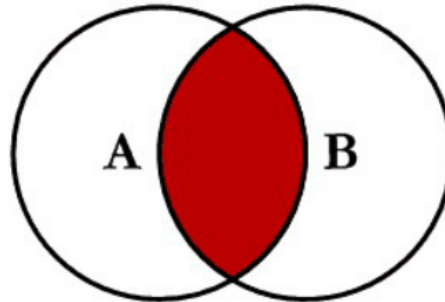
# SQL JOINS



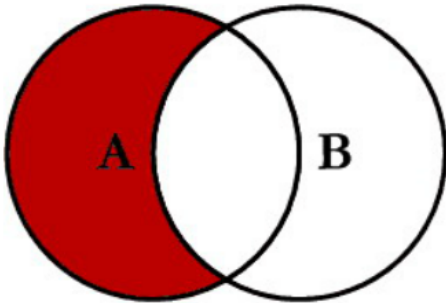
```
SELECT <select_list>  
FROM TableA A  
LEFT JOIN TableB B  
ON A.Key = B.Key
```



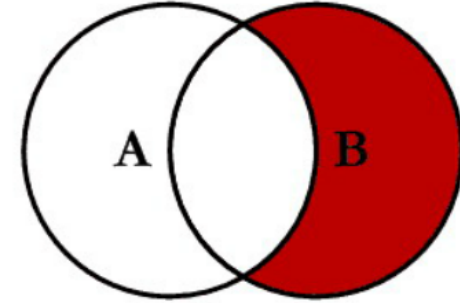
```
SELECT <select_list>  
FROM TableA A  
RIGHT JOIN TableB B  
ON A.Key = B.Key
```



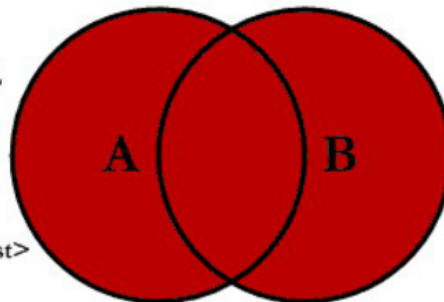
```
SELECT <select_list>  
FROM TableA A  
INNER JOIN TableB B  
ON A.Key = B.Key
```



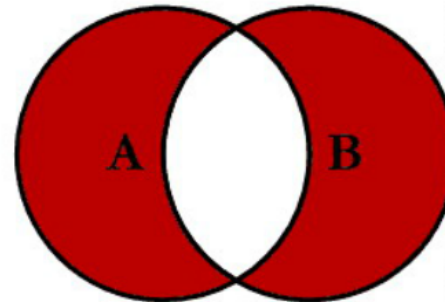
```
SELECT <select_list>  
FROM TableA A  
LEFT JOIN TableB B  
ON A.Key = B.Key  
WHERE B.Key IS NULL
```



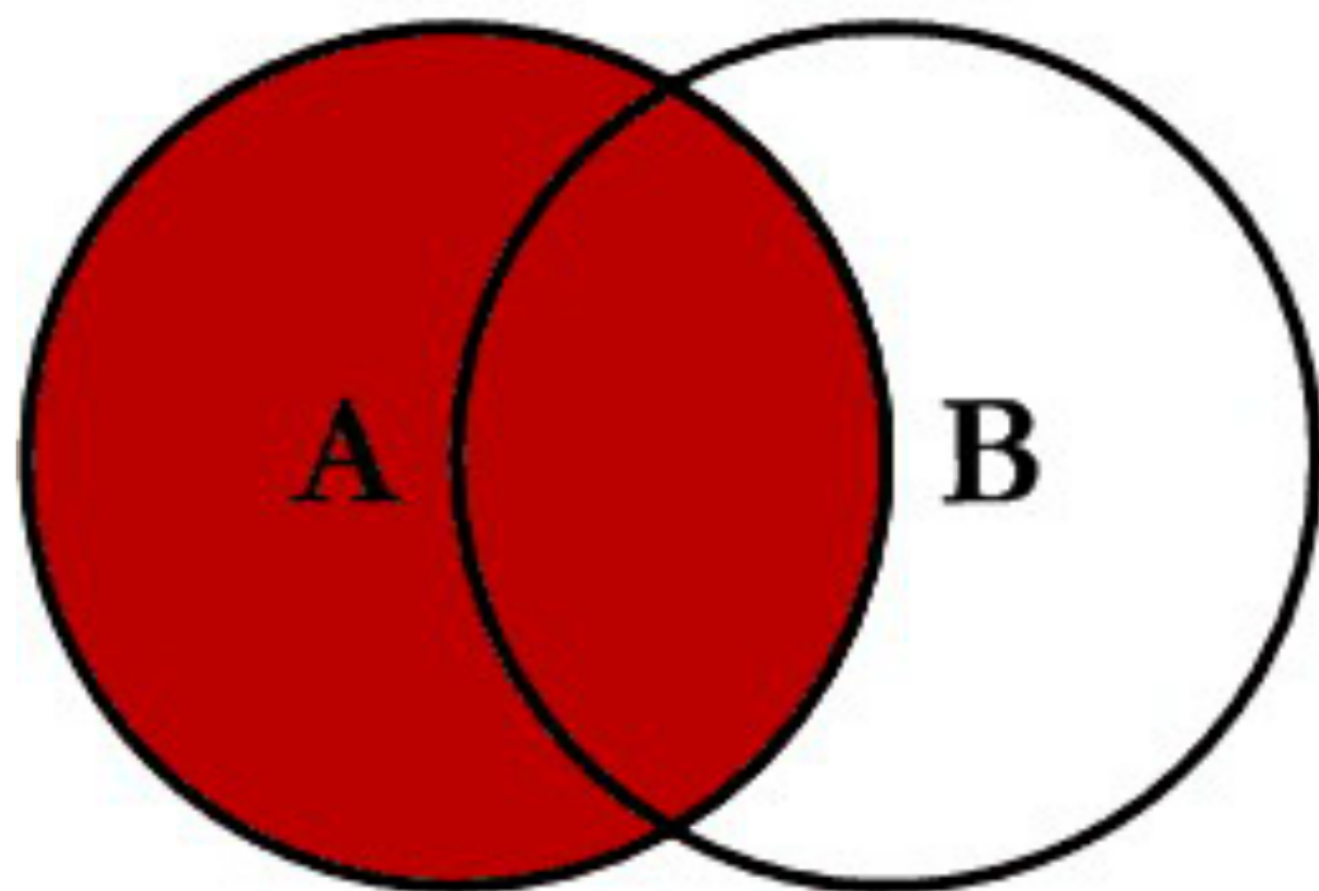
```
SELECT <select_list>  
FROM TableA A  
RIGHT JOIN TableB B  
ON A.Key = B.Key  
WHERE A.Key IS NULL
```



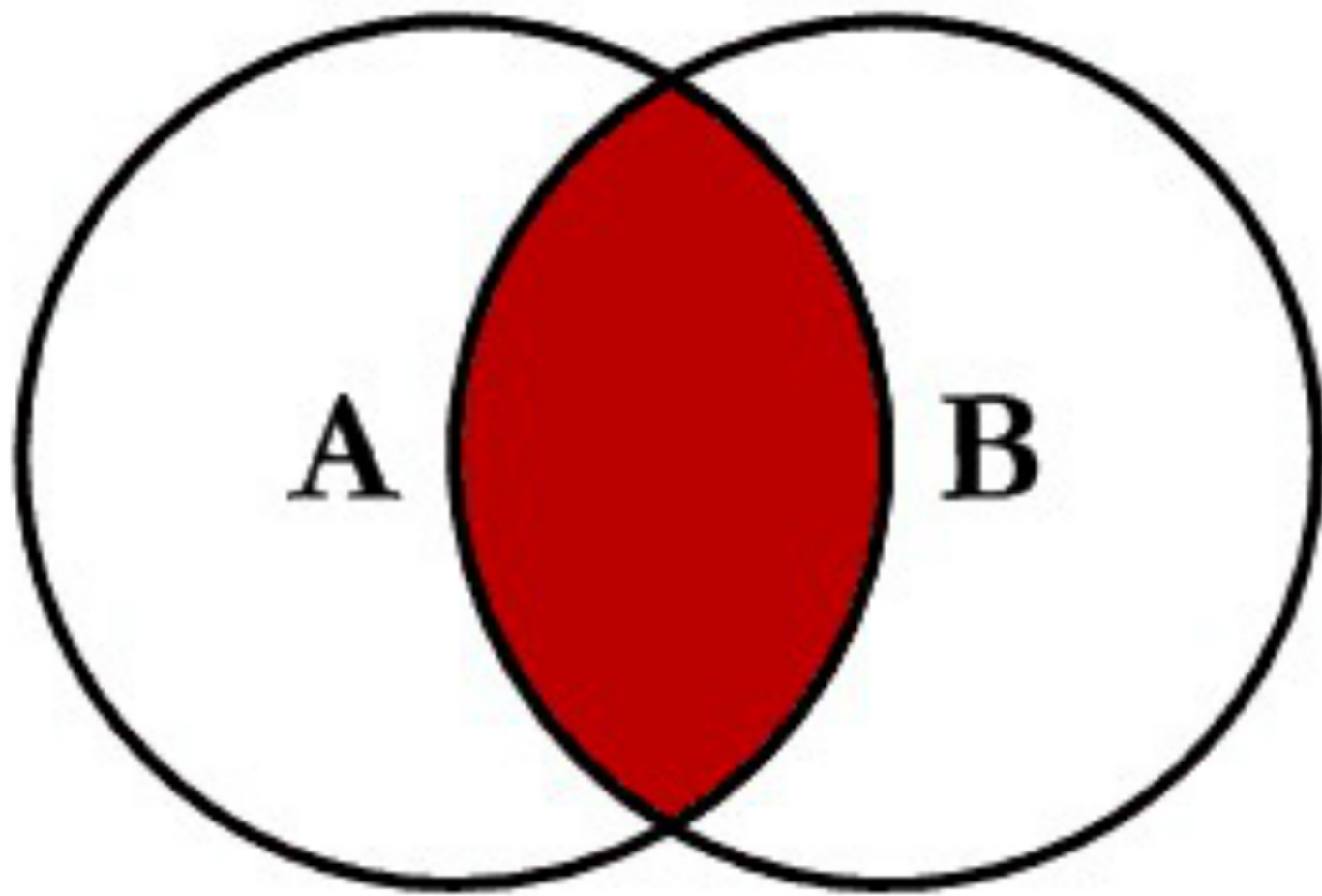
```
SELECT <select_list>  
FROM TableA A  
FULL OUTER JOIN TableB B  
ON A.Key = B.Key
```



```
SELECT <select_list>  
FROM TableA A  
FULL OUTER JOIN TableB B  
ON A.Key = B.Key  
WHERE A.Key IS NULL  
OR B.Key IS NULL
```

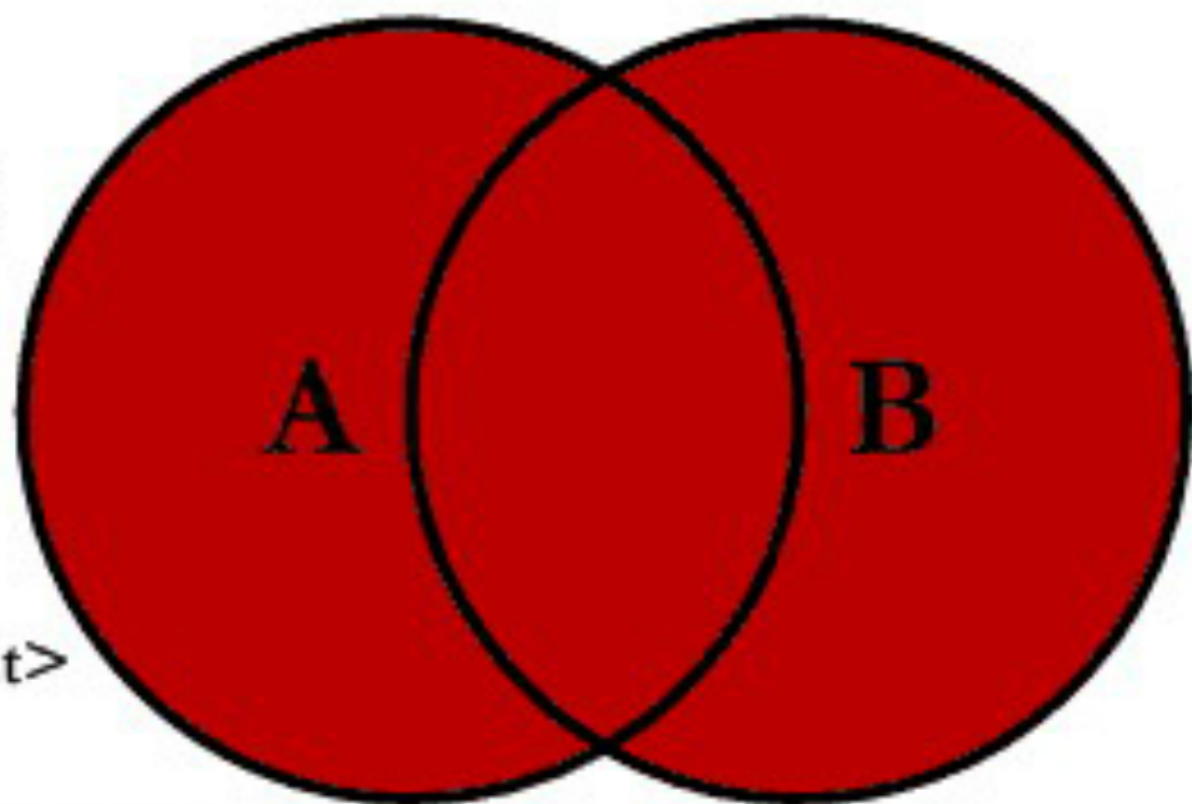


```
SELECT <select_list>  
FROM TableA A  
LEFT JOIN TableB B  
ON A.Key = B.Key
```



```
SELECT <select_list>  
FROM TableA A  
INNER JOIN TableB B  
ON A.Key = B.Key
```

```
SELECT <select_list>  
FROM TableA A  
LEFT JOIN TableB B  
ON A.Key = B.Key  
WHERE B.Key IS NULL
```



```
SELECT <select_list>  
FROM TableA A  
FULL OUTER JOIN TableB B  
ON A.Key = B.Key
```

# Introduction to database management with open source tools

Guillaume Larocque

research professional,

Quebec Center for Biodiversity Science

<http://qcbs.ca/wiki/opendb>



CENTRE DE LA SCIENCE DE LA BIODIVERSITÉ DU QUÉBEC  
QUEBEC CENTRE FOR BIODIVERSITY SCIENCE

<http://registration.qcbs.ca/pay>

