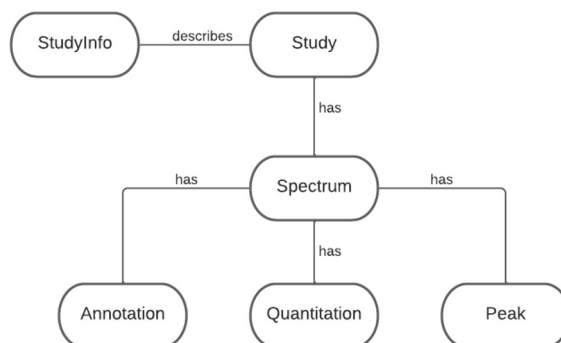


Lab 11

We continue working on our Metabolomics Database. As a reminder, this is the conceptual schema of the database.



This time, we'll focus on the **Annotation** data. Files "spectra.tsv" (note the extension) contains annotations for some of the spectra in each study. Not all spectra have corresponding annotations!

In addition to these files, we'll use the PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) to retrieve extra data for each compound (molecular weight and SMILES). On the PubChem website, you can type any molecule name and see what information is available for it.

Task 1

Download the script "read_annotations.py" from Canvas, and execute it for each file "spectra.tsv" as follows:

```
> python3 read_annotations.py FILENAME.tsv
```

This script will read all annotations from the file, and then use REST API to search compound names in PubChem and retrieve their molecular weight and SMILES if available. The script will create file "spectra.pubchem.csv" with all the information needed for the database.

If interested, you can take a look at the line 13 in the script, which contains the URL pattern to retrieve the data from PubChem.

WARNING: there is 1 second delay between each URL request. Don't remove it! Usually, websites don't like receiving too many requests from the same IP address (which is similar to the DDOS attack), so it's a good practice to have a short delay between requests.

Task 2

Create a new table **annotations** in your database "metabolomics.db" from Lab 9. This table should contain columns: signal, name, monoisotopic_mass, matching_score, formula, cas, molecular_weight, and smiles. Import the annotation data from all four studies.

Course: BINF-6211 LAB 11

Name: Drashti Mehta, SID: 801262877

Structure	Data	Constraints	Indexes	Triggers	DDL
Grid view	Form view				
<div><div><div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div><div></div></div><div><div></div></div></div></div></div>					

Task 3 (15pts)

Now we need to link the tables **annotations** and **spectra**. To link the two tables, you need to add a foreign-key column into one of the tables. Think which one, and what values this foreign-key column should contain.

Report the SQL statement to update the foreign key values.

Alter table annotations add foreign key (signal) REFERENCES Spectrum (name)

Task 4 (20pts)

Finally, let's make sure that the link between **annotations** and **spectra** works. In order to do that, try to calculate the number of annotated spectra in each study.

As I mentioned earlier, some of the spectra in the table **spectra** will have corresponding annotations in the table **annotations**, and other spectra will have no annotations at all. You need to count the number of spectra **with** annotations for each study.

Report the following:

A. SQL query to calculate the number of annotated spectra in each study.

Ans:

1. select Spectrum.spectrum_id, annotations.signal, annotations.name from Spectrum join annotations on Spectrum.name = annotations.signal

Course: BINF-6211 LAB 11
Name: Drashti Mehta, SID: 801262877

Grid view Form view		
<div> </div> Total rows loaded: 345		
spectrum_id	signal	name
1 7d79ba6e-bdc4-40a5-ab69-6108beb1862f	#1 73.0000@5.7190 MS1+	d-Mannose, 2,3,4,5,6-pentakis-O-(trimethylsilyl)-, o-methyloxyme, (1Z)-
2 ec66d437-ef4a-4a54-9fce-d7aee946b1a6	#2 73.0000@5.8930 MS1+	D-Xylopyranose, 4TMS derivative
3 3133cbf1-febb-4f46-a64d-52e3314fa05b	#3 73.0000@5.7554 MS1+	d-Mannose, 2,3,4,5,6-pentakis-O-(trimethylsilyl)-, o-methyloxyme, (1Z)-
4 86a6a2bb-e13f-4bf4-82aa-1836108c03b7	#4 73.0000@5.8068 MS1+	L-Tyrosine, 3TMS derivative
5 a58cb49c-492b-4440-ad31-6a7e6ce54758	#5 73.0000@2.8264 MS1+	Lactic Acid, 2TMS derivative
6 5a978f24-35d3-4c3f-bebf-fc85923b9274	#7 73.0000@5.6847 MS1+	d-Galactose, 2,3,4,5,6-pentakis-O-(trimethylsilyl)-, o-methyloxyme, (1Z)-
7 fa768853-0bdf-4568-9073-95bfccb6b309	#8 73.0000@5.7510 MS1+	d-Galactose, 2,3,4,5,6-pentakis-O-(trimethylsilyl)-, o-methyloxyme, (1Z)-
8 a087245e-d40c-4066-b01b-11a054a69f75	#10 73.0000@3.1849 MS1+	Oxalic acid, 2TMS derivative

B. The number of annotated spectra in each of the four studies.

1. select Spectrum.spectrum_id, annotations.signal, annotations.name from Spectrum join annotations on Spectrum.name = annotations.signal where Spectrum.study_id = "ST001270"

Grid view Form view		
<div> </div> Total rows loaded: 58		
spectrum_id	signal	name
1 7d79ba6e-bdc4-40a5-ab69-6108beb1862f	#1 73.0000@5.7190 MS1+	d-Mannose, 2,3,4,5,6-pentakis-O-(trimethylsilyl)-, o-methyloxyme, (1Z)-
2 ec66d437-ef4a-4a54-9fce-d7aee946b1a6	#2 73.0000@5.8930 MS1+	D-Xylopyranose, 4TMS derivative
3 3133cbf1-febb-4f46-a64d-52e3314fa05b	#3 73.0000@5.7554 MS1+	d-Mannose, 2,3,4,5,6-pentakis-O-(trimethylsilyl)-, o-methyloxyme, (1Z)-
4 86a6a2bb-e13f-4bf4-82aa-1836108c03b7	#4 73.0000@5.8068 MS1+	L-Tyrosine, 3TMS derivative
5 a58cb49c-492b-4440-ad31-6a7e6ce54758	#5 73.0000@2.8264 MS1+	Lactic Acid, 2TMS derivative
6 5a978f24-35d3-4c3f-bebf-fc85923b9274	#7 73.0000@5.6847 MS1+	d-Galactose, 2,3,4,5,6-pentakis-O-(trimethylsilyl)-, o-methyloxyme, (1Z)-
7 fa768853-0bdf-4568-9073-95bfccb6b309	#8 73.0000@5.7510 MS1+	d-Galactose, 2,3,4,5,6-pentakis-O-(trimethylsilyl)-, o-methyloxyme, (1Z)-
8 a087245e-d40c-4066-b01b-11a054a69f75	#10 73.0000@3.1849 MS1+	Oxalic acid, 2TMS derivative










2. select Spectrum.spectrum_id, annotations.signal, annotations.name from Spectrum join annotations on Spectrum.name = annotations.signal where Spectrum.study_id = "ST001369"

Grid view Form view		
<div> </div> Total rows loaded: 88		
spectrum_id	signal	name
1 978703a0-7527-44e0-a1e8-8b56e87de571	#1 73.1000@17.3530 MS1+	D-(+)-Talose, pentakis(trimethylsilyl) ether, methyloxime (syn)
2 6328d54c-85f8-419f-9ec7-2d4a4466fc0c	#2 147.1000@6.8253 MS1+	Lactic Acid, 2TMS derivative
3 ac43119c-fa01-4a8f-9020-3028c8c92d8e	#3 73.1000@17.5532 MS1+	D-(+)-Talose, pentakis(trimethylsilyl) ether, methyloxime (anti)
4 57757043-82d0-4854-b223-8404a3cf41f1	#4 147.1000@9.4589 MS1+	Urea, 2TMS derivative
5 7a79ea3a-fbc4-42dc-be8c-ef744ceadb0	#5 147.1000@9.4527 MS1+	Urea, 2TMS derivative
6 65d3edb2-f5e1-4574-bb06-ea7781e60de8	#8 73.1000@18.1348 MS1+	3-.alpha.-Mannobiose, octakis(trimethylsilyl) ether (isomer 2)
7 b512c795-ac42-4459-854a-9b58f82f704d	#9 73.1000@16.8965 MS1+	1,5-Anhydroglucitol, 4TMS derivative










Course: BINF-6211 LAB 11

Name: Drashti Mehta, SID: 801262877

3. `select Spectrum.spectrum_id, annotations.signal, annotations.name from Spectrum join annotations on Spectrum.name = annotations.signal where Spectrum.study_id = "ST001399"`

Grid view		Form view
		
     		 Total rows loaded: 95
spectrum_id	signal	name
1 9b195bcf-1390-4d90-b777-57c05e279484	#1 117.0000@11.8706 MS1+	Palmitic Acid, TMS derivative
2 be7132b8-c3a4-4743-9744-f7c2c7fb9859	#3 117.0000@12.7869 MS1+	Stearic acid, TMS derivative
3 a0570957-0aec-4948-beb7-8a258f2f5c77	#4 86.0000@4.8156 MS1+	Methoxyamine, 2TMS derivative
4 b245a1fd-53d4-407a-8b16-ddf3a8e2d9a9	#5 86.0000@4.9989 MS1+	Methoxyamine, 2TMS derivative
5 964d3c61-9559-443e-96bd-e75047627dcb	#6 86.0000@4.7255 MS1+	Methoxyamine, 2TMS derivative
6 8acee29c-0858-48b2-b341-bff5f9245154	#7 117.0000@12.6801 MS1+	Oleic Acid, (Z)-, TMS derivative
7 1ad54f0b-8643-4dde-bf16-e22ef6a8d673	#8 147.0000@11.0523 MS1+	d-Galactose, 2,3,4,5,6-pentakis-O-(trimethylsilyl)-, o-methyloxyme, (1E)-

4. `select Spectrum.spectrum_id, annotations.signal, annotations.name from Spectrum join annotations on Spectrum.name = annotations.signal where Spectrum.study_id = "ST001613"`

Grid view		Form view
		
     		 Total rows loaded: 104
spectrum_id	signal	name
1 4a7154f5-dc83-4736-aacf-f5470f10a9b3	#1 100.0000@7.5690 MS1+	Silane, isocyanatotrimethyl-
2 7b71c1a3-daf8-44ca-84ba-04b874a18a39	#2 147.0000@6.7321 MS1+	tert-Butylpentamethyldisiloxane
3 d7e82022-f46e-4805-a346-57ef0bb70365	#3 147.0000@6.7281 MS1+	Methylmalonic acid, 2TBDMS derivative
4 f4ebd9da-d165-472c-a68d-d41734d6e58b	#4 147.0000@6.5606 MS1+	Disiloxane, 1,3-bis(1,1-dimethylethyl)-1,1,3,3-tetramethyl-
5 9f7c3321-30fa-4bc1-a802-ba385c436ba8	#5 147.0000@6.7046 MS1+	Disiloxane, 1,3-bis(1,1-dimethylethyl)-1,1,3,3-tetramethyl-
6 86a7fbff-b472-4bcf-b74b-922020cd92d0	#6 147.0000@8.1972 MS1+	Bis(tert-butyldimethylsilyl) carbonate
7 d27d0275-68bb-40be-810f-39d1c901df22	#8 146.0000@7.2466 MS1+	Bis(tert-butyldimethylsilyl)amine
8 50bb3246-4f63-4457-80ec-20544aa383de	#9 147.0000@6.6184 MS1+	1-Pentamethyldisilyloxycyclohexane