

How do I use publicly available biodiversity data?

Get Started with...

Licenses

Darwin core and OBIS-ENV-DATA format

A quick guide to QGIS

Geoservers



First, make sure you are actually allowed to use the data for your project.

You may not be able to use certain datasets commercially, for example, or may need to approach the owner to ask them if you can use it for your project. This may also take some time, as you will need to wait for them to reply to you.

Most datasets will require that you at least acknowledge the data owner.

Check the site you are downloading the datasets from for this information, like copyright license, use restriction information, recommended citations, etc.



The latest standards indicate that datasets should define their copyright licence permissions through set copyright licences. This make it easier for the user, as you wont need to go onto every organisations Terms and Conditions page, or contact them to ask if you may use it.

Common copyright licenses include

Creative commons licenses: https://creativecommons.org/licenses/

Open Government License: http://www.nationalarchives.gov.uk/doc/open-

government-licence/version/3/





CC BY:



Others can use and build on the resource but must credit the owners of that resource

CC BY NC



Others can use and build on the resource for noncommercial purposes but must credit the owners of that resource

Open

Government Licence

Others can use and build on the resource but must credit the owners of that resource



A general guide citing a resource

Check the page you are downloading a data set or other resource from to see if it already has a recommended citation

Johns D (2011): Continuous Plankton Recorder Dataset (SAHFOS) - Phytoplankton. v4.1. Sir Alister Hardy Foundation for Ocean Science (SAHFOS). Dataset/Occurrence. https://doi.org/10.7487/ysaz8e

Citation

Johns D, Broughton D, Team D (2020). The CPR Survey. Version 1.2. Marine Biological Association. Sampling event dataset https://doi.org/10.17031/1629 accessed via GBIF.org on 2021-03-07.



If not, this is a general guide for how to lay out a citation

The name of the resource creator

The year the resource was made (or if this is not available, the year you accessed it)

The resource's title

The publisher of the resource or website name.

Journals and books

volume and page number.

Online sources

The date you accessed the resource(viewed a website, downloaded a dataset etc.)

DOI or website URL





DwC (Darwin Core) is a <u>framework</u> for standardising the way biodiversity data is complied to make data interoperable between different datasets, platforms and software.

DwC-A (Darwin Core Archive) is the <u>format of the files</u> you will download that contain the data. This is downloaded as a ZIP file containing data and metadata, that link types of information together through identification numbers. One of the files is the central 'core' file, and the other files contain information that links to it.

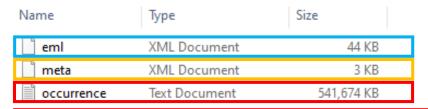
For more detailed information please see

https://www.gbif.org/darwin-core





In the ZIP file...



The 'Core' File:

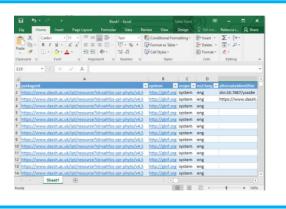
A CSV (Comma Separated Value) file- open this in Excel or Notes. This contains your species data and metadata like who collected the samples, the date, the location, etc. This can be a Taxon core, an Occurrence core like this example or the more recent Event core.

https://www.gbif.org/darwin-core for more explanation.

												ca Irvine	Q Shi
_	ile Home	Insert Page	Layout	Formu			_		,	want to do			X 241
ď	Calibri	v 11	* A A	= ;		General		Conditiona	Formatting			- <u>2</u> T-	
Pa	- Ba -			E :	= = =	- 👺 - 9	6 , 👜	Format as 1	able -	2 Delet	е т 🐺	- 0 -	
Pa.	ste 🦠 B I	U - 🖽 -	<u> </u>	6 ⊞ 3	<i>₩</i> -	*.0 .00 00 +0		Cell Styles		Form	at - 🧶	-	
Cli	aboard 15	Font		A A	anment	5 Numbe		Sty	les	Cells		Editing	
A	T 1	× .	f _e id										
a	A	В		С	D	Ε	F	G	н	1 1	,	K	
	id	type	mo							basisOfRer			no cat
2	001MD-1-112	Collection			The Marin			SAHFOS	CPR	HumanOb 0			
3	001MD-11-121	Collection		2:00.2	The Marin	Continuo	50	SAHFOS	CPR	HumanOb 0	01MD-1	1 001MD-	-11 00
4	001MD-11-128	Collection		2:04.2	The Marin	Continuo	50	SAHFOS	CPR	HumanOb 0	01MD-1	1 001MD-	11 00
5	001MD-1-113	Collection	- 2	1:15.8	The Marin	Continuo	50	SAHFOS	CPR	HumanOb 0	01MD-1-	001MD-	1- 00
6	001MD-1-117	Collection	- 1	0:45.0	The Marin	Continuo	50	SAHFOS	CPR	HumanOb (01MD-1-	001MD	1-:00
7	001MD-1-121	Collection	- 2	3:24.0	The Marin	Continuo	50	SAHFOS	CPR	HumanOb 0	01MD-1-	001MD	1- 00
8	001MD-1-124	Collection	- 2	3:31.4	The Marin	Continuo	50	SAHFOS	CPR	HumanOb 0	01MD-1-	001MD-	1-:00
9	001MD-1-198	Collection	- 2	1:03.7	The Marin	Continuo	50	SAHFOS	CPR	HumanOb 0	01MD-1-	001MD-	1- 00
10	001MD-13-103	Collection	- 4	8:02.2	The Marin	Continuo	50	SAHFOS	CPR	HumanOb 0	01MD-1	3 001MD-	-13 00
11	001MD-13-112	Collection		8:16.2	The Marin	Continuo	50	SAHFOS	CPR	HumanOb (01MD-1	3 001MD-	13 00
	001MD-13-117	Collection			The Marin			SAHFOS	CPR	HumanOb 0			
	001MD-13-166	Collection			The Marin			SAHFOS	CPR	HumanOb (
	001MD-13-198	Collection			The Marin			SAHFOS	CPR	HumanOb (
	001MD-15-103	Collection			The Marin			SAHFOS	CPR	HumanOb 0			
16	001MD-15-112	Collection	- 2	4:57.7	The Marin	Continuo	50	SAHFOS	CPR	HumanOb (01MD-1	5 001MD-	15 00:
	000	arrence She	et1	(+)				1 4					

eml:

This contains information about the dataset, and more generic metadata like it's DOI, copyright license, etc.



meta:

This is a 'descriptor' file that contains links to the Darwin Core term definitions that you will see in your core file. This defines the terms on the core file and its structure for the user.

For any column header terms you are unsure about in the occurrence file, check these links to find out what they are referring to.

They link to https://dwc.tdwg.org/terms/

				no Der			Debps						Secra Name	Q Share
Table Name: Table1 © Resize Table Properties	1-1 farmers 1 St. Convert to		load See	Capant Rad	0 =0				to Salton		Ten I	2222		
A1		V 5.	metadata											
A A			D				0			- 1				
I SHOULD NOT	- Distriction	DESCRIPTION OF	MI - 12	100		eTrans		- Inthesion	· man	market I	-			
2 amilumi	UTF-B	W 1	in .		2.10		es/Ant/Serms/Securities	occurrence.tx	4 0	- 1			Team .	
2 lamiand	UTF-8	No. 3	n		170	gc/7m,tdwg.c	sg/dwc/terms/Occurrence	occurrence, to	d 0	- 2	hete//purl.or	a/do/tames/	modified	
4 lent.xini	UTF-E	10 1	n:		1 10	gillin titling i	rs/des/terms/Occurrence	occurrence.to	4 0	1	http://put.or	g/dc/terms/	redictation of	
5 emi.xml	UTF-8	10 3	in .		2 her	gc//es.tohog.o	og/dws/terms/Occurrence	occurrence to	0 0	- 4	http://purl.or	p/do/terms/	bibliographs	cOtatoo
8 and and	Utti-8	38 3	n		1.10	or Houndard of	ng/det/lermi/Otturrence	occurrence.to	4 6		heter//w.hibes	Limpleworks		Oversity
7 and and	UTF-E	\t 1	in .		1 10	pr//rs.hthing.t	rs/Awc/terms/Occurrence	occurrence.to	4 0	- 4	hetp://rs.bbws	Ling/dwr/h		
6 and and	UTF-B	te 1	in .		2 10	griffer, tollway, o	raz/dwa/terms/Occurrence	occurrence.tx	4 0	7	http://ru.hdwo	Lors/dec/to		
\$ entant	UTF-E	10 1	n		2 700	gr//rs, tohog, o	op/dws/terms/Occurrence	occurrence.to	4 0		http://m.hthu	Loral theo're	my/beabCf	Second
10 emland	UTF-B	M Y	in.		1 10	gr.J/Ins. hthory, n	rs/Mes/terms/Occurrence	occurrence.tx	e e		hetes///es.tables	Lora/dwr/h	erms/materia	elSempl
11 ami ami	UTF-8	34 3	in .		2 141	groffins, belowing a	es/des/terms/Occumence	occurrence.br	4 5	30	http://es.takes	Lors/dwic/te	emu/occurre	Charma
12 amlund	UTF-B	12 1	in .		1 10	#27%.58wg.0	rg/dwc/terms/Occurrence	occurrence.to	4 0	11	hete//rs.tdw	Lorg/Sec/b	emiliation	Numbe
13 ent.ent	UTF-8	10 3	n ·		110	gr.J.frs. tollwag.o	rs/fes/terms/Scorrence	occurrence.tx	1 0	1.7	heta://es.tabes	corp/dws/h		0
14 emi.ani	UTF-E	10 1	in.				rg/dat/terms/Coomence		4 0	11	http://es.table	con/dec/s	mo/semple	nghoto
15 emland	UTF-8	\n \	0		1.76	gr.J.Frs. hthog. s	ng/dws/terms/Occurrence	occurrence.to	4 0	34	http://rs.tdws	Long/dwx/te	erms/sample	direvel
ts and and	UTF-E	10 1	je.		110	pr//vs.tdwg.t	ng/dec/terms/Constrance	occurrence.to	4 6	15	nets///s.nee	Long/dwc/h	error, sample	disesse
17 entant	UTF-8	34 3	in .		1 10	gc//rs.tdwg.o	rg/dws/terms/Occurrence	occurrence.tx	4 0	16	http://n.hthu	Lorg/dwc/tr	erms/sample	niiMod
10 ent.xmi	UTF-B	10 1	n		1 10	gr.//vs.tdwg.c	rop/dws/terms/Coourrence	occurrence.to	1 0	17	http://m.tdw	Local discrib	minutes and Ti	ine
18 ent.uni	UTF-B	M (1	in i		100	gr.//vs.tchag.c	rs/live/termy/Documence	occurrence.to	d 0	25	hetp://rs.tohus	Long/dwy/te	mms/year	
20 amilamii	UTF-B	14 1	in .		2 10	puller tilling i	rg/dat/terms/Occurrence	occurrence.tr	4 8	19	http://rs.tahes	Lorg/dwic/to	dancer Jerma	
21 ami, ami	UTF-B	34 3	n .		170	w//rs.ndwg.a	eg/dwc/terms/Occurrence	occurrence, to	1 0	.20	hetgr//rs.tdw	corp/dwc/to	erms/day	
22 emt.ami	UTF-III	10 1	n		1 10	gi//rs.htteg.i	rs/des/terms/Occurrence	occurrence, to	4 0	21	http://m.hibo	Lors/Shirt/h	emi/held%	amber.
23 eml.xml	UTF-8	10 3	n		1 her	gc//m.htheg.t	rsg/dest/terms/Occumence	occurrence to	d 0	22	hete//rs.hthu	Lorg/dws/to	emy/eventh	charie
te ami, ami	Utr-8	10 1	n		1.10	gr/linkting o	ng/dast/termiy/tecurrence	occurrence.to	a 6	23	http://ww.tship	Linglidwijts		imbert
	Sheet1	(1)												
finally	-										1901 10			A 100%

https://www.gbif.org/darwin-core and https://obis.org/manual/darwincore/ for more explanation.





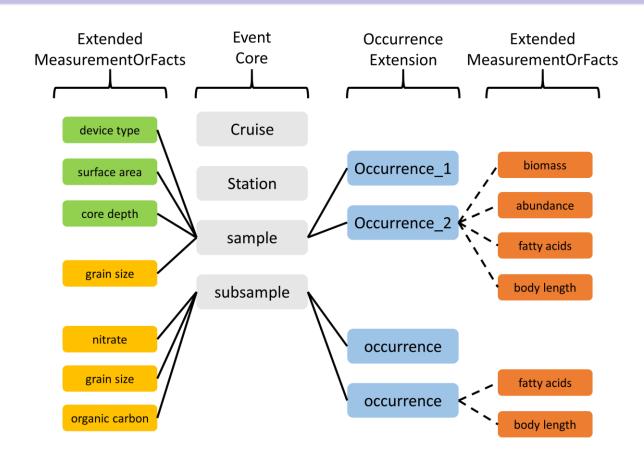
This builds upon the Darwin Core system by using an 'Event core', extra identifiers for measurements and linking to an Occurrence extension and ExtendedMeasurementOrFact extension.

For example, environmental data like geological, physical and chemical data and extra sampling information can be linked up with it's associated sample.

These can be used to link more measurements and information to a sampling event or species occurrence while minimising the amount of repeated information presented to the user.

For more information, please see

https://obis.org/manual/dataformat/



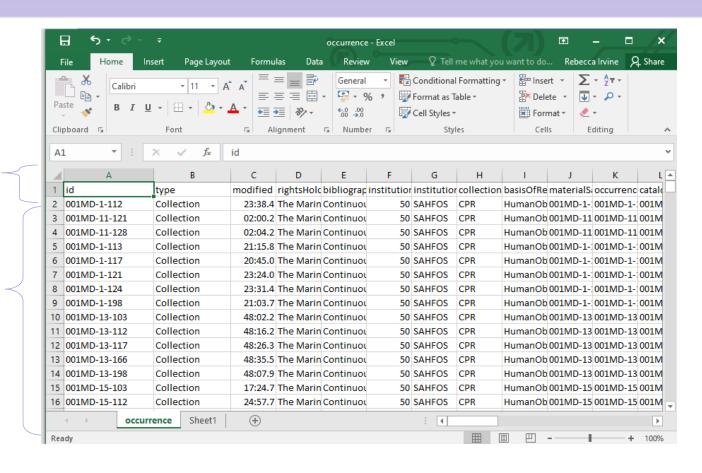




In the Core file...

Column headers tell you what information is in that particular column. If you are unsure about the definition of the header, check the information hyperlinked in the 'meta' file.

Each row also includes the associated metadata for that individual record. While this means that some information in the file may get repeated many times, it allows you to know what metadata applies to each specific record without confusion and you can easily filter out records you are interested in by date, taxa, etc.





Darwin Core

In the Core file...

Within the core file, there are columns containing various types if ID numbers- depending on the number, this may be unique to the survey the record was collected on (eventID), unique to each row (id), etc.

	Α	F	J	K	L	М	AB	AC	
1	<u>id</u>	institutionID	materialSampleID	occurrenceID	catalogNumber	<u>eventID</u>	taxonID	<u>scientificNameID</u>	
2	001MD-1-112	50	001MD-1-112	001MD-1-112	001MD-1-112	001MD-1	112	urn:lsid:marinespecies.org:taxname:370366	
3	001MD-11-121	50	001MD-11-121	001MD-11-121	001MD-11-121	001MD-11	121	urn:lsid:marinespecies.org:taxname:109951	
4	001MD-11-128	50	001MD-11-128	001MD-11-128	001MD-11-128	001MD-11	128	urn:lsid:marinespecies.org:taxname:156509	
5	001MD-1-113	50	001MD-1-113	001MD-1-113	001MD-1-113	001MD-1	113	urn:lsid:marinespecies.org:taxname:163030	
6	001MD-1-117	50	001MD-1-117	001MD-1-117	001MD-1-117	001MD-1	117	urn:lsid:marinespecies.org:taxname:149093	
7	001MD-1-121	50	001MD-1-121	001MD-1-121	001MD-1-121	001MD-1	121	urn:lsid:marinespecies.org:taxname:109951	
8	001MD-1-124	50	001MD-1-124	001MD-1-124	001MD-1-124	001MD-1	124	urn:lsid:marinespecies.org:taxname:109982	
9	001MD-1-198	50	001MD-1-198	001MD-1-198	001MD-1-198	001MD-1	198	urn:lsid:marinespecies.org:taxname:157256	
10	001MD-13-103	50	001MD-13-103	001MD-13-103	001MD-13-103	001MD-13	103	urn:lsid:marinespecies.org:taxname:148912	
11	001MD-13-112		001MD-13-112	001MD-13-112	001MD-13-112	001MD-13		urn-Isid-marinesnecies org-taxname-370366	
Ready	occurrence	①					1		+ 150%

Other ID numbers, like institutionID, are used to identify something on a controlled vocabulary list.





Controlled vocabularies are used to keep certain terms consistent and defined between datasets. These controlled vocabularies are defined in a few dedicated websites, for example:

Projections: https://epsg.org/home.html

General vocabulary lists: https://vocab.seadatanet.org/search

http://vocab.ices.dk/

https://obis.org/manual/dataformat/

Organisations: https://edmo.seadatanet.org/search

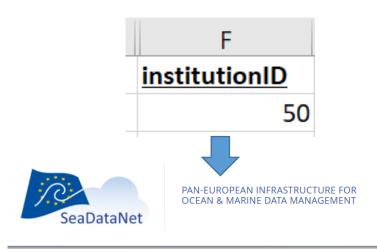
Marine species: http://www.marinespecies.org/aphia.php?p=search

http://www.marinespecies.org/msbias/aphia.php?p=search

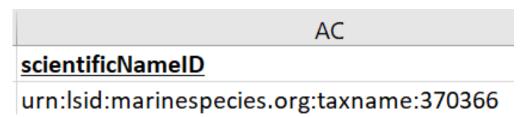
There are controlled vocabularies for all sort of things- the method something was collected (sediment cores, transect counts...), the platform it was collected on (fishing boats, ROV's...), the name of the organisation collecting the data (Marine Biological Association, University of Plymouth), the type of projection the coordinate system is in... even individual research vessels can have their own ID numbers in a controlled vocabulary.

Controlled vocabularies

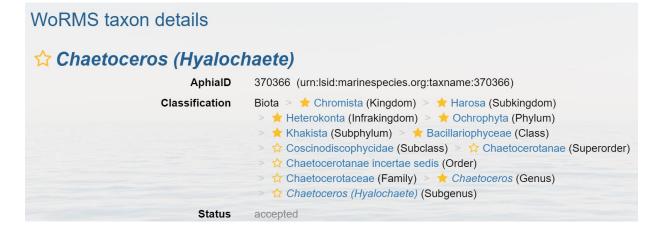
















A quick guide to QGIS

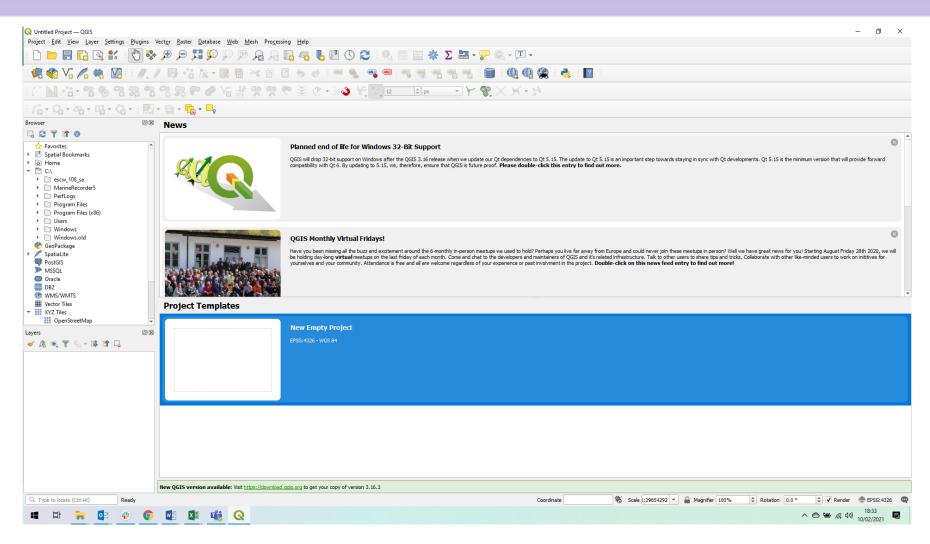
QGIS is a geographic information system. It is **free** and **open source**, making it accessible to anyone with the internet to download it and a computer that can run it. For this reason, it has been chosen for this tutorial, however there are other GIS software packages out there.

See https://www.qgis.org/en/site/forusers/index.html#trainingmaterial to download QGIS, view tutorials and learn more.



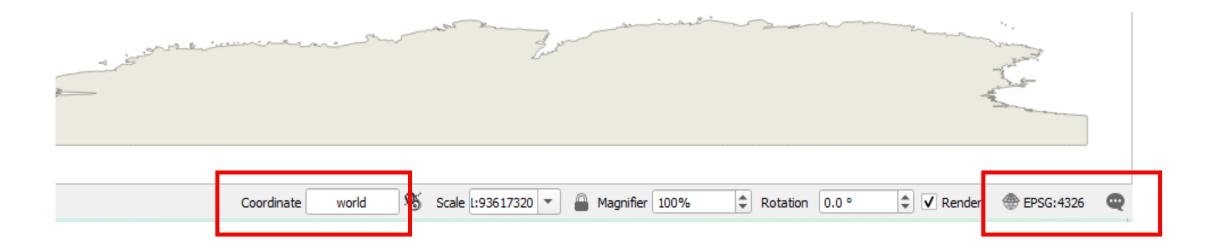


Open QGIS and start a new project



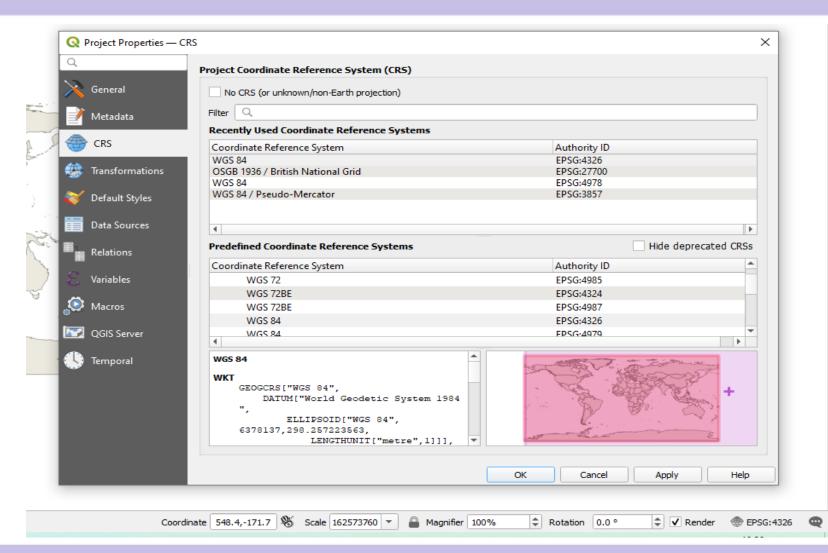


Type 'world' into the coordinate box at the bottom and hit enter. Check the coordinate reference system (CRS) in the bottom right where the half dome/ cone symbol is- WGS 84 EPSG: 4326 is often a good bet.



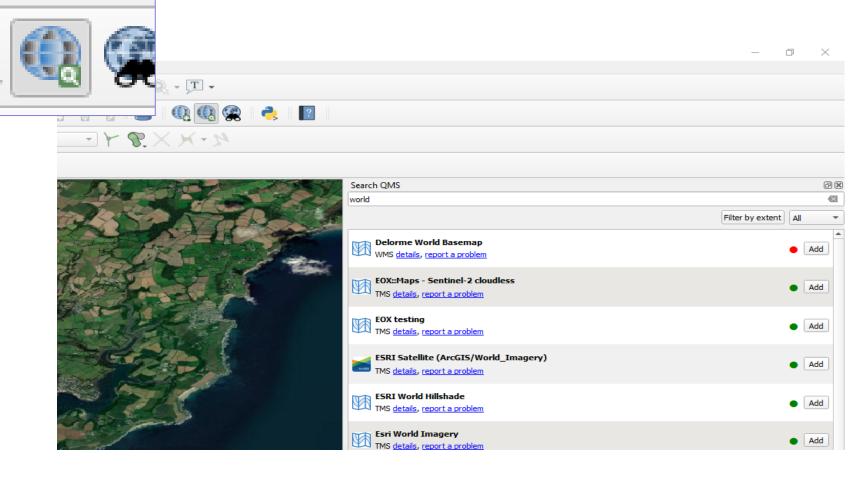


To change the coordinate reference system, click on the bottom right where it says what CRS is being used. It will bring up a window where you can search and apply for the correct one for your project.





To add more detailed map layers, click on the globe with a magnifying glass symbol in the toolbar at the top of the screen. You can use a keyword search to select different types of map, view details about the map layer like its CRS and select it for the map.

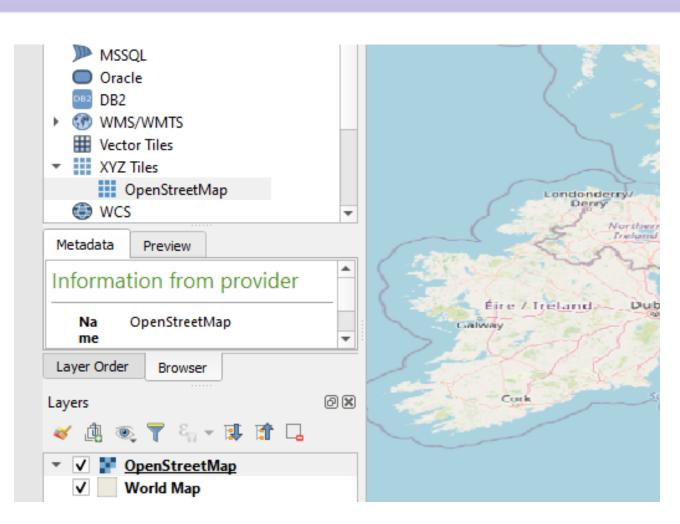




You can also select XYZ Tiles from the left hand options box and double click on OpenStreetMap

You can rearrange layers by clicking and dragging things in the 'layers' box into different orders.

Make sure the layers you want on top are at the top of the list.







There are a few different formats of files that you can download online for use in QGIS and other GIS software. The following slides will cover a few file formats that can be downloaded from the DASSH Geoserver as examples. Some of these formats will be offered by other data portals, too. Think about what format is easiest for you to use, and works best in your project.

Generally, if you are creating a map, the outcome of using these different formats will be similar, if not the same. If given a choice between different file formats just go with what is most easy for you and don't worry about trying a few types out to see what works best for your project.

This will cover...

Text and CSV files

Shape Files

GeoJSON

Importing data into QGIS



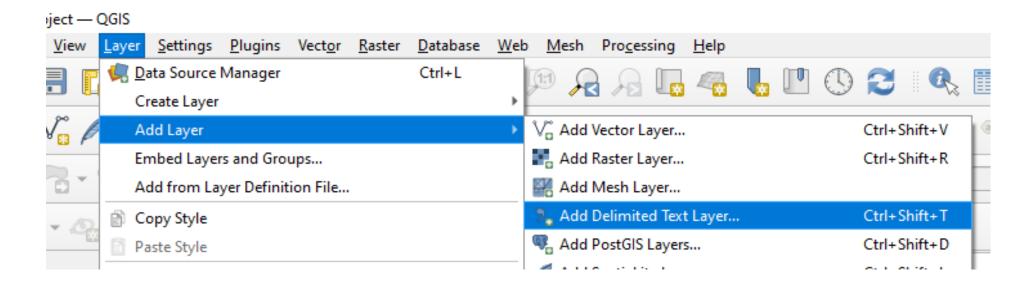
Please see the slides for 'Where can I find publicly available biodiversity data?' for a guide for how to access the DASSH Geoserver and download data.

Those slides also cover accessing other data portals offering biodiversity data. This is a generic tutorial that can apply to data from any portal that offers the file formats used in these examples.





CSV/ Text file



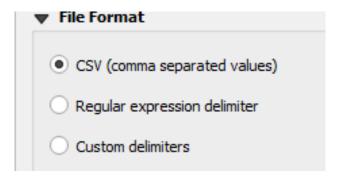
To add layers from a CSV or text file Select Layer> Add Layer> Delimited Text Layer





Select the input file by clicking the three dots by the File Name box and navigating to the CSV you wish to use. To note, you can also do this with your own text files of coordinates if they are laid out in columns, for example, if you have copied columns of coordinates from excel and pasted them into Notepad.

Use file format to tell QGIS how to read the file and separate out the data in the file format box. You can select CSV, or you can tell it what to look for in the document to separate out the columns properly



▼ File Format		
CSV (comma separated values)	Tab	Colon
Regular expression delimiter	Semicolon	✓ Comma
Custom delimiters	Quote "	





Check the file in excel so you know how the columns should be laid out, and don't be afraid to click through the different options and see how they look.

You can check at the Sample Data box at the bottom to see how QGIS is interpreting the document.

FID	survey_code	survey_name	abstract	project_code	metadata_url	originator	dataset_owner
combined_data.fid4647908f_178114efe5c1c	b DASSHDT00000337	2011 Marine Biological	A SHARC and MBA members run timed species surv	DASSHSE00000046	https://portal	45: Marine Biological Associ	45: Marine Biological Asso
combined_data.fid4647908f_178114efe5c1c	a DASSHDT00000337	2011 Marine Biological	A SHARC and MBA members run timed species surv	DASSHSE00000046	https://portal	45: Marine Biological Associ	45: Marine Biological Asso
combined_data.fid4647908f_178114efe5c1c	9 DASSHDT00000237	1998-2003 Stentiford et	A collection of Non-Native Species sightings from G		https://portal	GD Stentiford	GD Stentiford
combined_data.fid4647908f_178114efe5c1c	8 DASSHDT00000070	2014 Marine Biological	Timed searches undertaken by the Marine Biological		https://portal	45: Marine Biological Associ	45: Marine Biological Asso



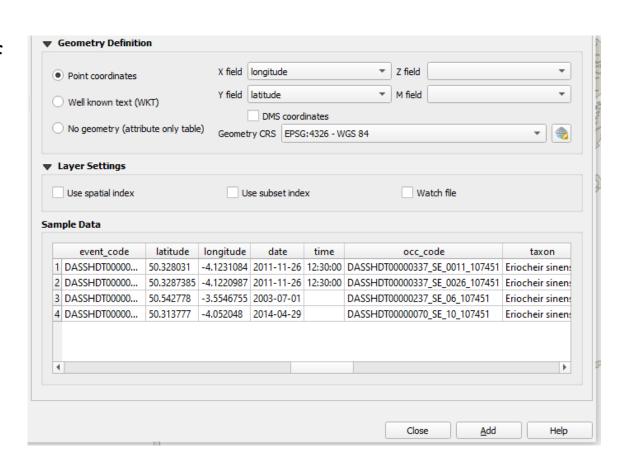


The latitude and longitudinal data should be in separate columns in the Sample Data box if you have selected the correct options in File Format.

Select the latitudinal and longitudinal columns in the Geometry Definition drop down boxes.

QGIS may already populate this for you, but double check it is correct.

When you are done, click Add and close the box.

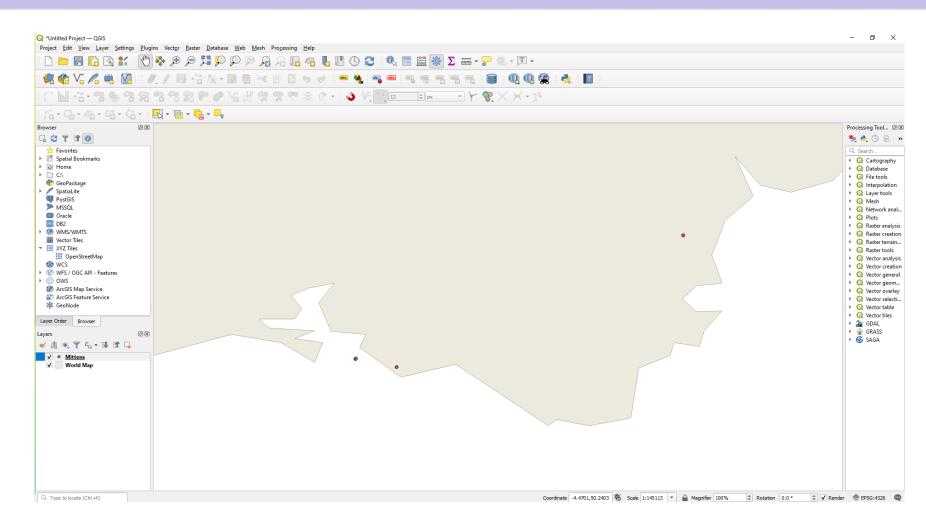


Checking your map layer



Look in the layers box in the bottom left. Right click on the layer you just created, and select the option 'Zoom to Layer'. Check that the points are were you expect them to be from the map on the DASSH records search. If not, you may have got the latitude and longitude the wrong way around and should remake the layer.

You can rearrange layers by clicking and dragging things in the 'layers' box into different orders.

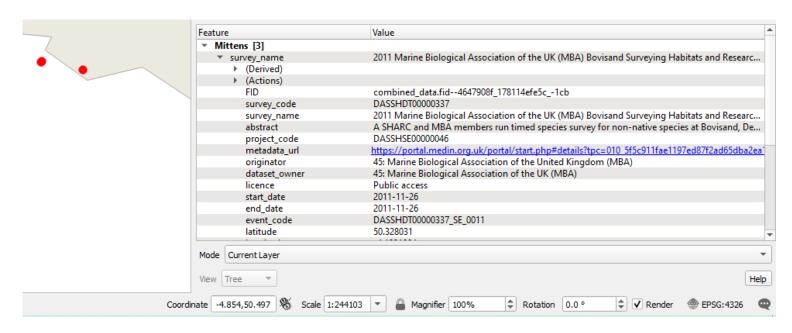


Checking your map layer





You can use the Identify Feature tool in the top toolbar, then click or click and drag to draw boxes around points of interest. Information about the points will appear in the Identify Results box. Make sure you have clicked on the layer you want to check first in the Layers box, to select it out.

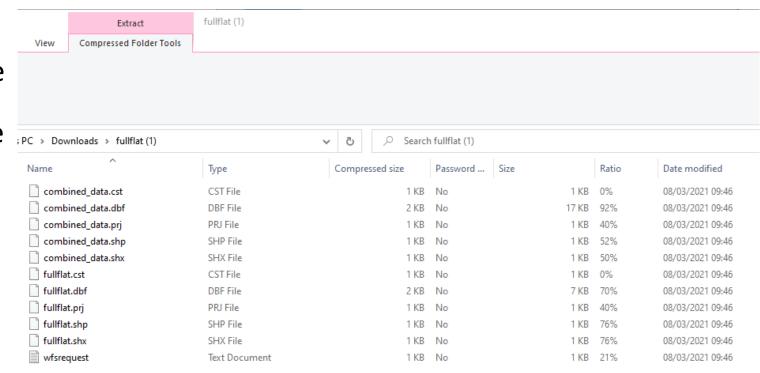






If you download a Shape File, you will download a ZIP file. Extract the contents and save them somewhere sensible under a name you will be able to find later.

In the DASSH Geoserver, the files you download will be called **combined_data** and **fullflat**, depending on which of our databases each particular dataset is stored in.





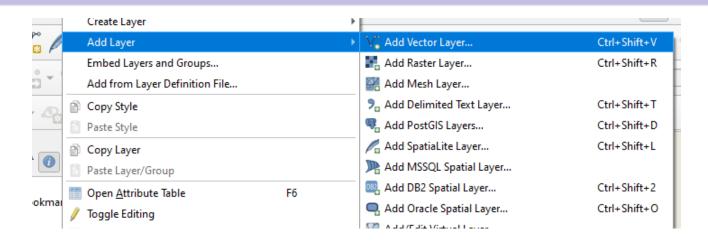


Select

Layer> Add Layer> Add Vector Layer

Use the button with three dots in the Source box to navigate to your download. Select combined_data.shp.

(if the map layer doesn't look right later, retry this step and select fullflat.shp).



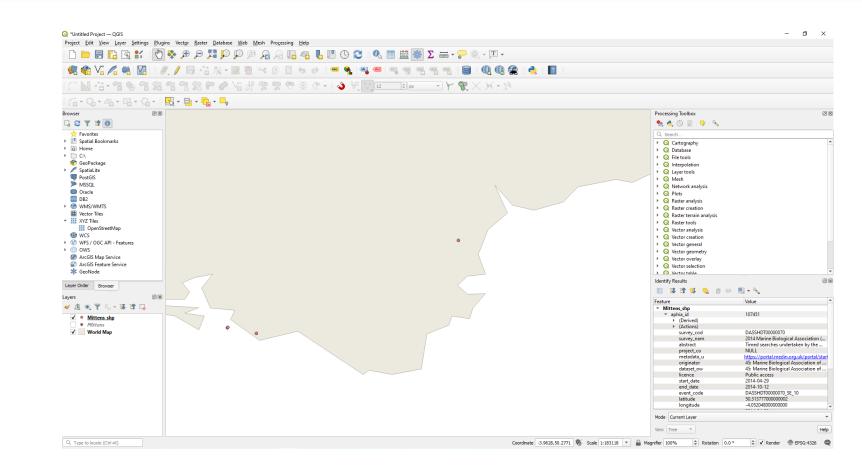
Vame	Date modified	Type	Size
combined_data.cst	08/03/2021 12:54	CST File	1 KB
combined_data.dbf	08/03/2021 12:54	DBF File	17 KB
combined_data.prj	08/03/2021 12:54	PRJ File	1 KB
combined_data.shp	08/03/2021 12:54	SHP File	1 KB
combined_data.shx	08/03/2021 12:54	SHX File	1 KB
fullflat.cst	08/03/2021 12:54	CST File	1 KB
fullflat.dbf	08/03/2021 12:54	DBF File	7 KB
📄 fullflat.prj	08/03/2021 12:54	PRJ File	1 KB
fullflat.shp	08/03/2021 12:54	SHP File	1 KB
fullflat.shx	08/03/2021 12:54	SHX File	1 KB
wfsrequest	08/03/2021 12:54	Text Document	1 KB





Leave the automatic settings on and click Add.

As before, zoom to the layer and check the points are where you expect them to be.





GeoJSON

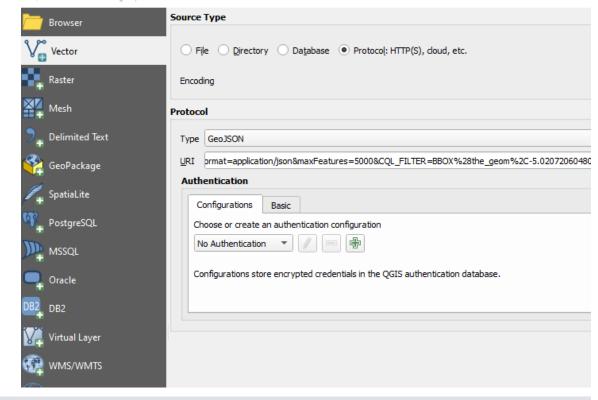
Select Layer> Add Layer> Add Vector Layer

Select the **Protocol** option in **Source Type**.

Select **GEOJSON** from the options in the **Protocol** box.

Copy and paste the **URI** from the web page that appears after you select to download a GeoJSON option (For example from the DASSH Geoserver), and paste it into the URI box in the Protocol section.

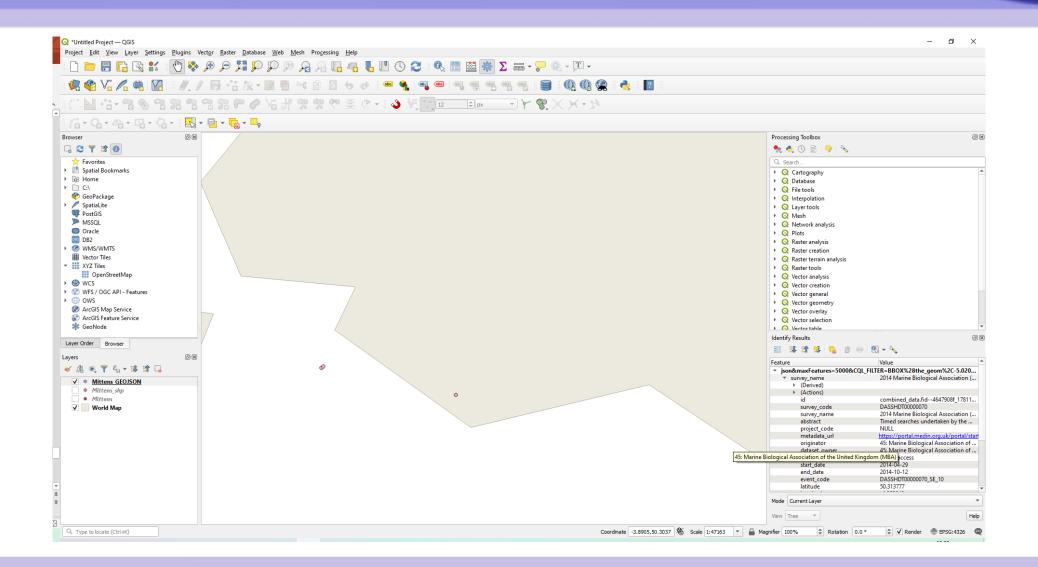
Select Add.







GeoJSON







A Geoserver is an open-source server that can be used to share geospatial data.

You may be able to connect directly to a Geoserver with URL links made available on data portals.



Connect to a Geoserver

On EMODnet, for example

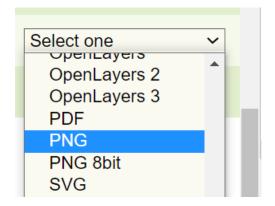






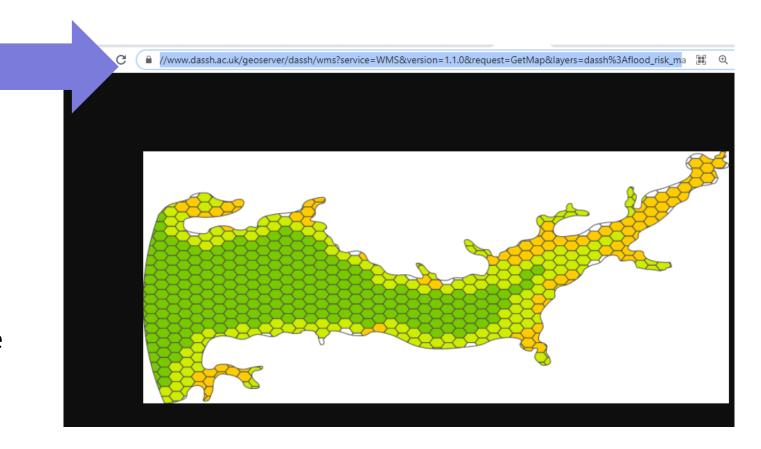


And the DASSH Geoserver



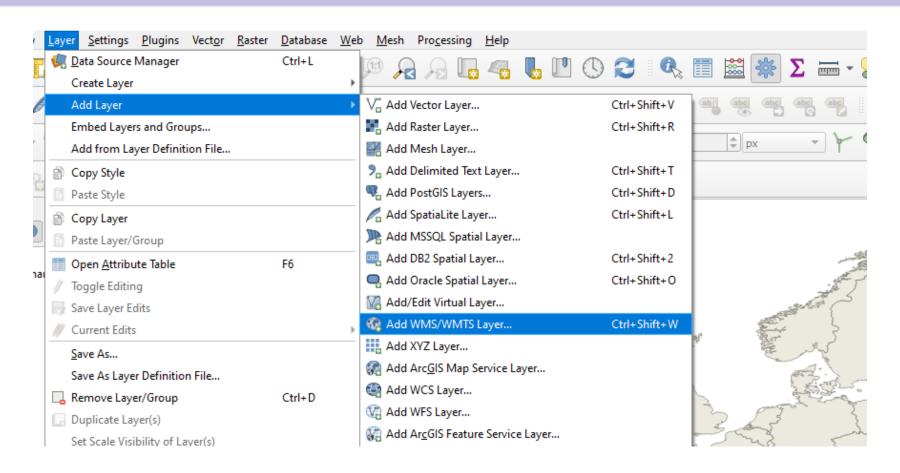
SVG PNG JPEG GIF AtomPub

These are formats that when selected from the DASSH Geoserver, will open a browser window. Copy the browser URL for a Geoserver link.





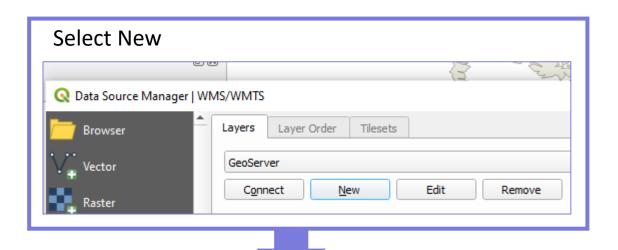




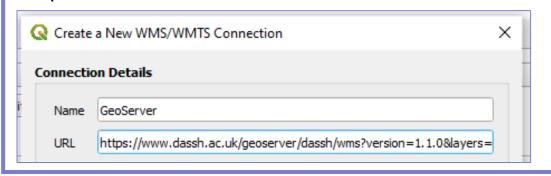
Select Layer> Add Layer> Add WMS/ WMTS Layer...

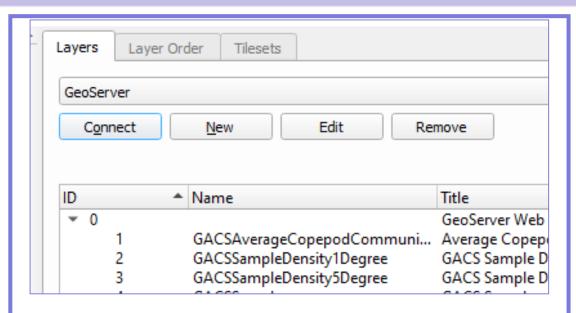
Connect to a Geoserver





Name the connection and insert the link you copied into the URL box. Click Ok.





Click Connect. The available layers on the Geoserver will appear below. Scroll and select the layer you want to add, select the Image encoding- PNG is a good option if are unsure. You can name the layer in the 'Layer Name' Box. Select Add when you are ready. Add new layers one by one, don't select a few layers and try to add them all at once.