*The **International Ocean Institute (IOI)** founded in 1972 by Elisabeth Mann Borgese, is a world leading independent, non governmental non profit organisation conducting training & capacity building in Ocean Governance with the aim of creating knowledgeable future leaders.*

'Advancing Sustainable Coastal Planning through Effective Use of Open-Access Biodiversity Information Systems'

Where getting names right always matters!

bob.branton@gmail.com

GitHub.com
OceanExpert.org

Outline ...
- Present situation ...
  - Ecosystem change
  - Information infrastructure
  - Challenges
  - Getting names right
- Using data science ...
  - Data Preparation
  - Data Visualisations
  - Control panel / Dashboard
- Recap
  - Software components
  - Benefits
  - More Challenges
- Conclusion
  - What's next
  - Collaboration
- Demos

Ecosystem Change

Information Infrastructure
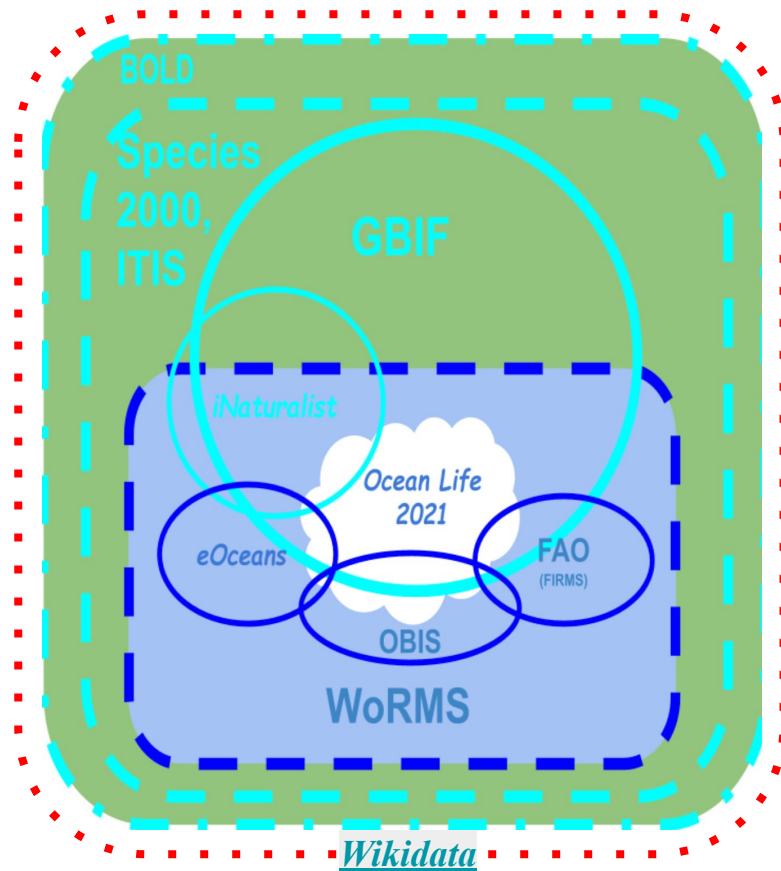
Invasive Species

"Removing green crab proves effective in controlling invasive species."

HALIFAX, 2016 … trap invented by a local fisherman was used to catch more than two million green crab [_Carcinus maenas_ Linnaeus, 1758] from one estuary over a few years. Since then, eelgrass [_Zostera L._] in that estuary has recovered by about 34 per cent and soft-shell clam [_Mya arenaria_ Linnaeus] populations are on the rise ...

Atlantic CTV News

BOLD

Species 2000, ITIS

GBIF

iNaturalist

Ocean Life 2021

eOceans

FAO (FIRMS)

OBIS

WoRMS

Wikidata

What's in a name? Wikidata, World_Register_of_Marine_Species#History, Taxonomy_(biology)#Taxonomic_descriptions , Taxonomy_mnemonic , Oceanlife-2021/domains…

Challenges

Scientific names change over time*

- 55 of 73 Ocean Life 1859 scientificNames required updating …

Data are not all in one place ..

- 63 of 78m OBIS obs (~80%) and 13 of 40m iNat Research Grade obs (~33%) not on GBIF
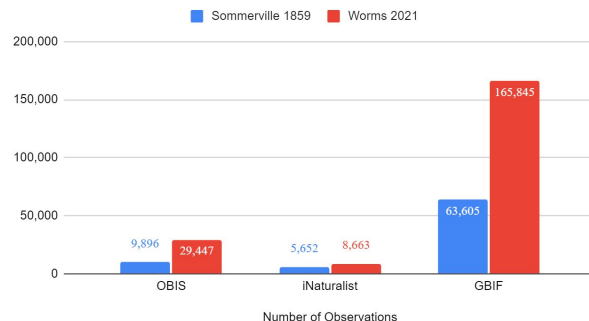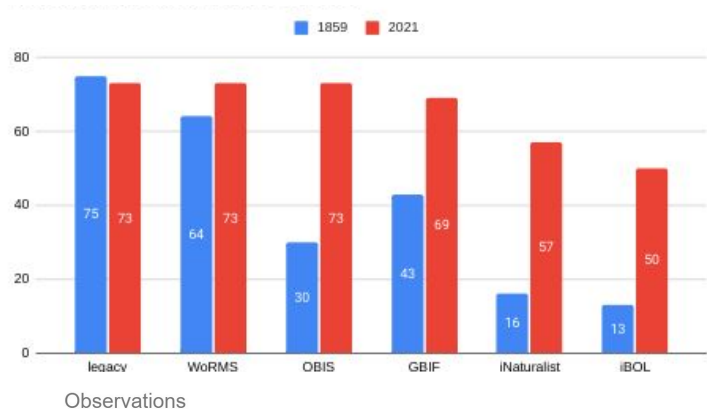
WoRMS/OBIS.authority not = GBIF.authority

- 26 of 68 GBIF.authorities contained abbreviations, 14 of 68 GBIF.authorities did not include year . The GBIF search window rejected WoRMS/OBIS.scientificName that contained dates

Inconsistent approach between portals regarding data queries, downloads & citations

OBIS and GBIF downloads often **too large** for Google Sheets

Getting Names Right

Using World Registry of Marine Species (WORMS) accepted names greatly improve query results ...

Observations

	1859	2021
legacy	75	73
WoRMS	64	73
OBIS	30	73
GBIF	43	69
iNaturalist	16	57
iBOL	13	50

Number of Observations

	Sommerville 1859	Worms 2021
OBIS	9,896	29,447
iNaturalist	5,652	8,663
GBIF	63,605	165,845

See Oceanlife-2021/worms

Using Data Science ...

*ocean.si.edu/ocean-life/5-invasive-species-you-should-know

Species

Carcinus maenas [European Green Crab]*
Caulerpa taxifolia [Killer Algae]*
Codium fragile [Dead Man's Fingers]
Dreissena polymorpha [Zebra Mussel]*
Mnemiopsis leidyi [Sea Walnut]*
Pterois volitans [Lion Fish]
Rapana venosa [Veined Rapa Whelk]*

* https://docs.ropensci.org/spocc/

Providers

ala - Atlas of Living Australia [AUS]*
bison - Biodiversity Information Serving Our Nation [USA]*
eBird - Cornell Lab of Ornithology {Global]
gbif - Global Biodiversity Information System [Global]*
idigbio - Integrated Digitized Biocollections [USA]*
inat - iNaturalist [Global]*
obis - Ocean Biodiversity Information System [Global]*
Vertnet - Distributed databases with a backbone

*https://www.marineregions.org/

FAO*
Boundaries of the FAO Fishing Areas

See:https://www.marineregions.org/sources.php#fao

EEZ*
Marineregions: the intersect of the Exclusive Economic Zones and IHO ocean areas

See:https://www.marineregions.org/sources.php#ihoeez

Primary data products ...

Summary ...

	prov	species	year	date	OCCS	DUPS	longitude	latitude	eez	fao
	<chr>	<chr>	<int>	<date>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>	<chr>
8687	UNIQUE	Carcinus maenas	2018	2018-09-28	1	3	-122.30957	37.86517	USA	99
11319	UNIQUE	Carcinus maenas	2017	2017-07-25	2	4	-127.76165	51.48631	CAN	67

See: species_style.r on GitHub

Expanded / hybercube ...

	species	prov	year	eez	fao	area
	<chr>	<chr>	<int>	<chr>	<int>	<chr>
1	Carcinus maenas	ala	1916	AUS	99	-38 145.1....
2	Carcinus maenas	ala	1916	AUS	99	-38 145.1....

See: prepare_data.r on GitHub

Visualizations ....

	ala	bison	bold	gbif	idigbio	inat	obis	UNIQUE	DUPS	total
Carcinus maenas	236	1652	10	9988	127	6785	8408	26794	412	54412
Caulerpa taxifolia	501	248	0	1355	403	55	637	2432	767	6398
Codium fragile	454	1268	0	6006	895	2341	759	10345	1378	23446
Dreissena polymorpha	0	4717	20	9979	514	1284	769	15185	2098	34566
Mnemiopsis leidyi	0	87	0	1166	12	262	2314	3479	362	7682
Pterois volitans	515	1213	0	9999	282	1523	2602	15175	959	32268
Rapana venosa	0	0	1	272	10	191	178	627	25	1304
total	1706	9185	31	38765	2243	12441	15667	74037	6001	160076
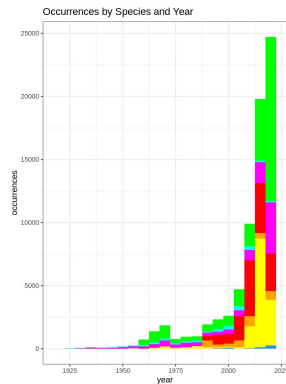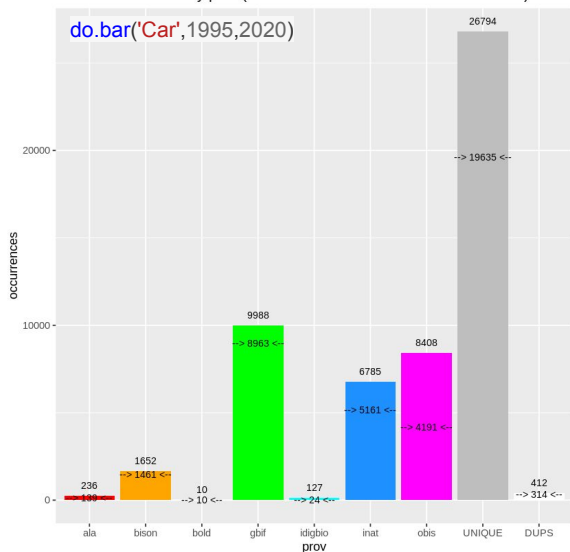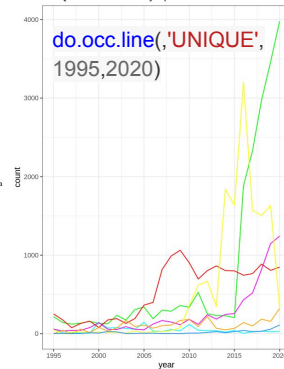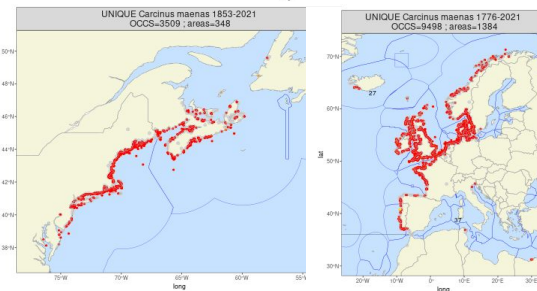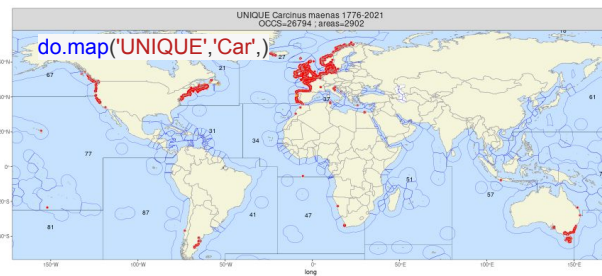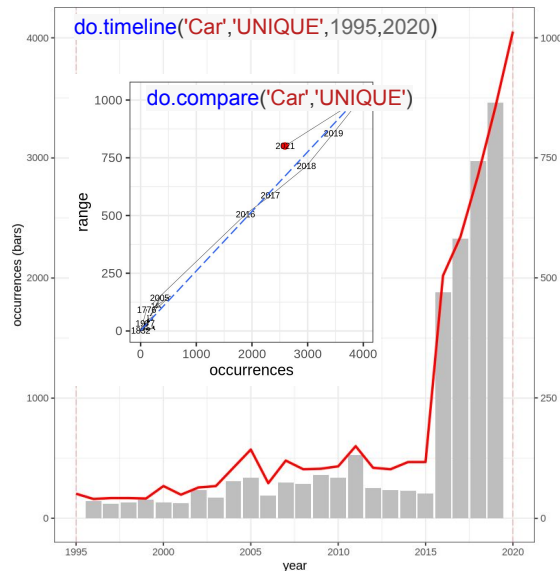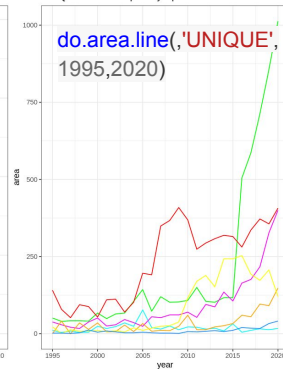
Dashboard / Control Panel
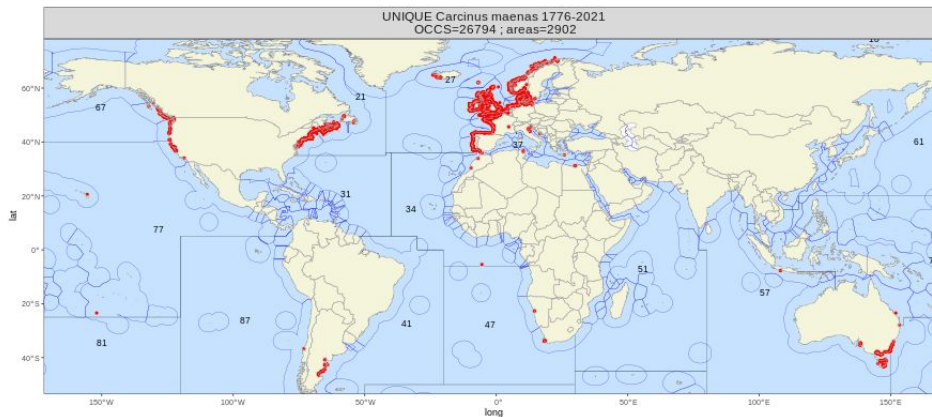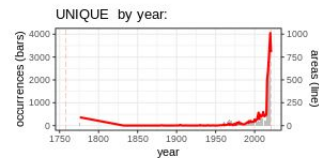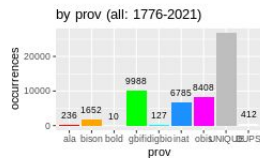
Live Access: https://rmbranto.shinyapps.io/invasives/

Recap ...

Software Components:
- R: A highly programmable free computer environment under GNU General Public License;
- Jupyter: Combine software code, computational output & explanatory text into a single document;
- Spocc - (SPecies OCCurrence) query & collect species occurrence data from many sources;
- ggplot - a general scheme for creating graphs and maps;
- ShinyApp: build interactive web apps straight from R;
- GitHub; where the world builds software;

Features and benefits:
- Queries species occurrence data from available providers using list of WoRMS accepted names
- Generates synthetic UNIQUE / DUPlicate data objects from available provider data
- Generates provider specific Taxon IDs and Keys and so formulate the deep links into live provider portals
- Filter and zoom maps and graphs using list of Providers, date range, FAO fishing Area and EEZs

Challenges:
- Spocc queries limited to 9999 occurrence records per species and provider
- bison and inigbio providers do not support deep linking by species and or taxon key/id
- using EEZ and FAO shapefiles for Pacific centred mapping

Last slide

Conclusion

Continue preparing for IOI 2022 ...
- More testing and refinement
- Add provider API links to dashboard
- Experiment with single species control panel
- Creation and use of defined area polygon queries
- Optional eez/fao and/or

Seek collaborations …
- marine plants, trawl bycatch species lists
- Create vignette for OpenSci
- Create wikipedia content

Demos

Dashboard

GitHub

Hyperslide

Wikidata

Appendices: synthetic data objects

species.style

A data.frame: 7 × 5

Names	cNames	sColors	wikidata	fImages
<chr>	<chr>	<chr>	<chr>	<chr>
Carcinus maenas	European Green Crab	green	Q27779	https://upload.wikimedia.org/wikipedia/commons/1/17/Carcinus_maenas.jpg
Caulerpa taxifolia	Killer Algae	cyan	Q310961	https://upload.wikimedia.org/wikipedia/commons/e/e7/CaulerpaTaxifolia.jpg
Codium fragile	Dead Mans Fingers	magenta	Q2712208	https://upload.wikimedia.org/wikipedia/commons/e/ed/Codiumfragile.jpg
Dreissena polymorpha	Zebra Mussel	red	Q752130	https://upload.wikimedia.org/wikipedia/commons/thumb/a/a9/Dreissena_polymorpha3.jpg/220px-Dreissena_polymorpha3.jpg
Mnemiopsis leidyi	Sea Walnut	orange	Q133630	https://upload.wikimedia.org/wikipedia/commons/thumb/1/1e/Sea_walnut%2C_Boston_Aquarium.jpg/220px-Sea_walnut%2C_Boston_Aquarium.jpg
Pterois volitans	Lion Fish	yellow	Q824672	https://upload.wikimedia.org/wikipedia/commons/thumb/b/bf/Pterois_volitans_Manado-e_edit.jpg/220px-Pterois_volitans_Manado-e_edit.jpg
Rapana venosa	Veined Rapa Whelk	dodgerblue	Q139053	https://upload.wikimedia.org/wikipedia/commons/thumb/f/fb/Rapana_Black_Sea_2008_G1.jpg/220px-Rapana_Black_Sea_2008_G1.jpg

head(prov.keys)

A data.frame: 6 × 7

	species	ala	bison	gbif	idigbio	inat	obis
	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
1	Carcinus maenas	284294	98734	5178595	5178595	52523	107381
2	Caulerpa taxifolia	83052	6974	2643172	2643172	50919	144476
3	Codium fragile	84082	6897	5272096	5272096	67555	370562
4	Dreissena polymorpha	NA	81339	2287072	2287072	116340	181566
5	Mnemiopsis leidyi	NA	53917	2501248	2501248	180788	106401
6	Pterois volitans	189453	166883	2334438	2334438	47280	159559

head(df.prov[df.prov$DUPS>2&df.prov$species=='Carcinus maenas',])

A data.frame: 6 × 10

	prov	species	year	date	OCCS	DUPS	longitude	latitude	eez	fao
	<chr>	<chr>	<int>	<date>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>	<chr>
8687	UNIQUE	Carcinus maenas	2018	2018-09-28	1	3	-122.30957	37.86517	USA	99
11319	UNIQUE	Carcinus maenas	2017	2017-07-25	2	4	-127.76165	51.48631	CAN	67
11320	DUPS	Carcinus maenas	2017	2017-07-25	6	3	-127.76165	51.48631	CAN	67
23906	UNIQUE	Carcinus maenas	2016	2016-08-18	1	3	-60.33113	46.80721	CAN	99
27236	UNIQUE	Carcinus maenas	2017	2017-10-13	1	3	-63.93700	44.51300	CAN	21
30240	UNIQUE	Carcinus maenas	2017	2017-05-31	1	3	-66.89060	45.03820	CAN	21

head(df.exp)

A data.frame: 6 × 6

	species	prov	year	eez	fao	area
	<chr>	<chr>	<int>	<chr>	<int>	<chr>
1	Carcinus maenas	ala	1916	AUS	99	-38 145.1....
2	Carcinus maenas	ala	1916	AUS	99	-38 145.1....
3	Carcinus maenas	ala	1955	AUS	99	-37.6 149.8..
4	Carcinus maenas	ala	1957	GBR	99	50.3 -4.1....
5	Carcinus maenas	ala	1958	AUS	57	-38.4 144.9..
6	Carcinus maenas	ala	1958	AUS	57	-38.7 145.7..

prov.style

A data.frame: 9 × 4

prov	id	color	order
<chr>	<chr>	<chr>	<dbl>
UNIQUE	UNI	grey	8
DUPS	DUP	white	9
ala	ala	red	1
bison	bis	orange	2
bold	bol	yellow	3
gbif	gbi	green	4
idigbio	idi	cyan	5
inat	ina	dodgerblue	6
obis	obi	magenta	7