AquaMaps - behind the scene (Part II)

Eli E. Agbayani
Josephine D. Rius-Barile
FishBase Project/INCOFISH WP1
WorldFish Center

TCOM Meeting
Basel, Switzerland
8-12 September 2008

- Description of tables (HCAF, HSPEN, HSPEC)
- How the tables relate
- Rules on calculating the
  - environmental envelopes
  - probabilities
- Bottlenecks in AquaMaps
- Future of AquaMaps

HCAF (Half-degree Cells Authority File)

Contains information on:

a) various cell names (codes) that are used by this and other databases

b) statistical cell properties such (center, limits, and area);

c) membership in relevant areas(FAO areas, EEZs or LMEs);

d) physical attributes (depth, salinity or temperature);

e) biological properties (e.g. pimary production).

Data gathered from:

Sea Around Us Project

CSIRO

Kansas Geological Survey

Compiled by:

Kristin Kaschner & Jonathan Ready

Download Metadata
-local-

HSPEN (Species Environmental Envelope)

- Contains information used for describing the environmental tolerance and preference of a species:
  - distribution using FAO areas and bounding box
  - range of values per environmental parameter (min., preferred min., preferred max., max.)

Download Metadata

-local-

HSPEC (Half-degree Species Assignment)

- Contains the assignment of a species to a half-degree cell and the corresponding probability of occurrence of the species in a given cell.

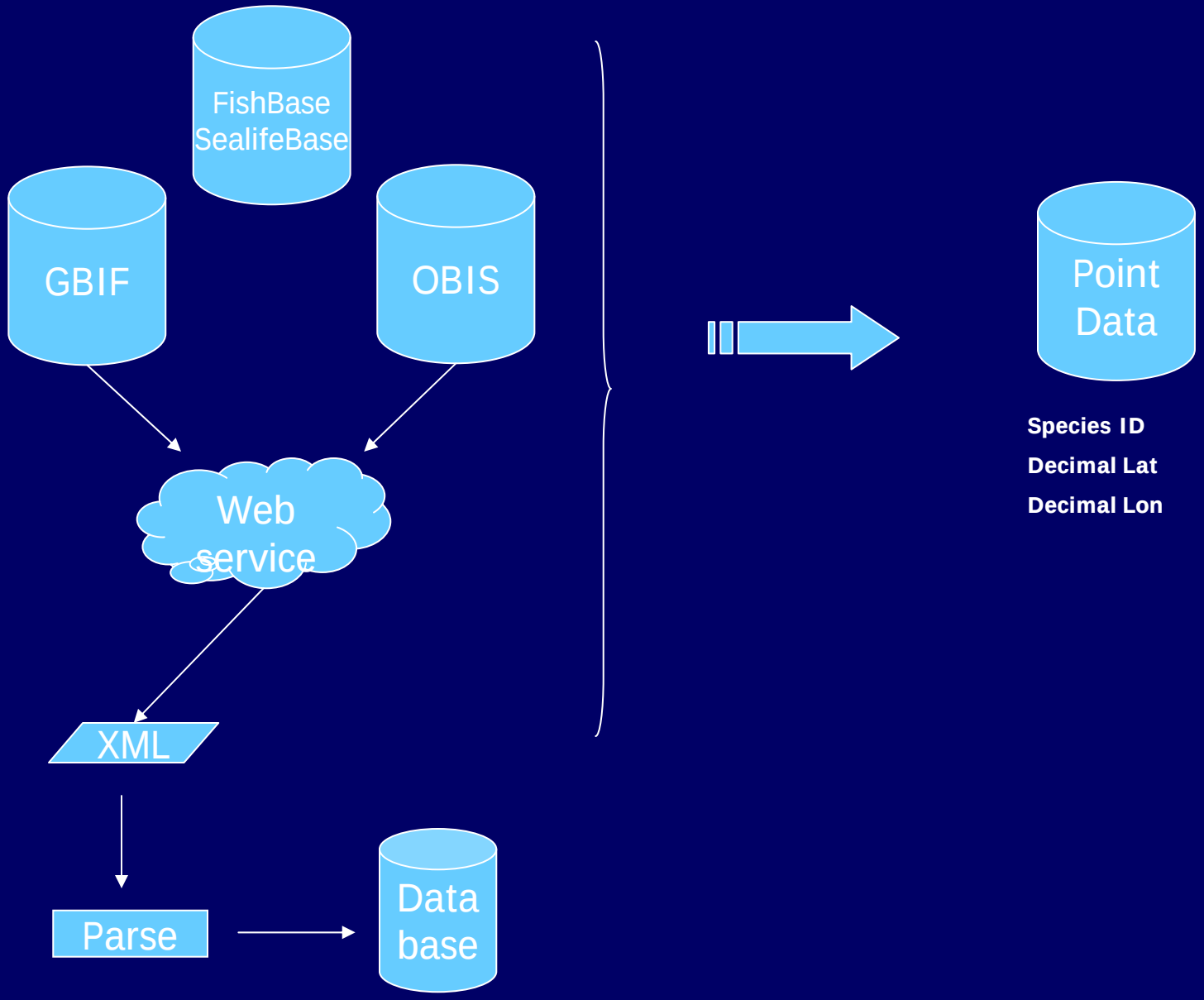
- Overall probability is the multiplicative equation of each of the environmental parameters (SST, salinity, prim. prod., sea ice concentration, distance to land)

Download Metadata

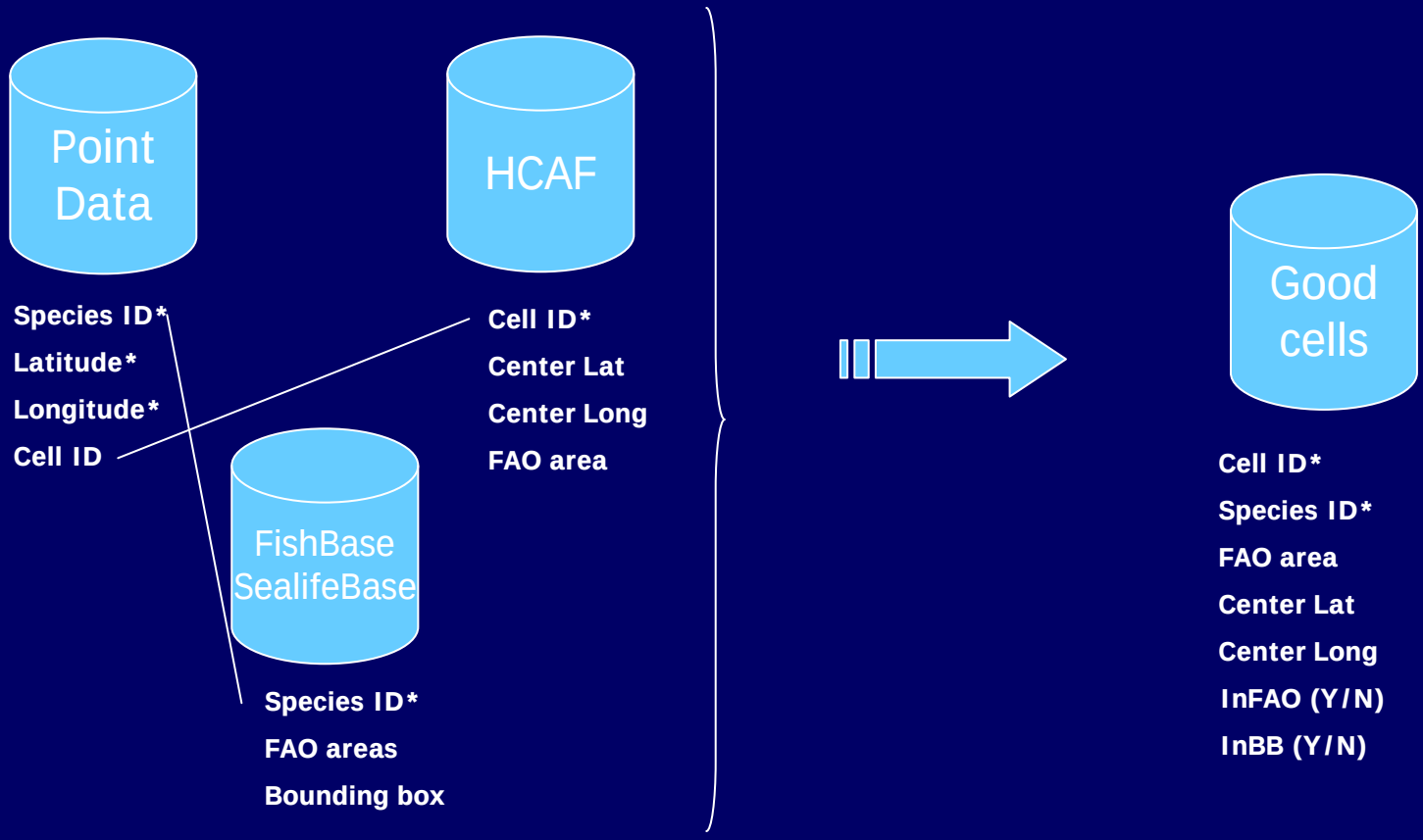
-local-

How the tables relate...

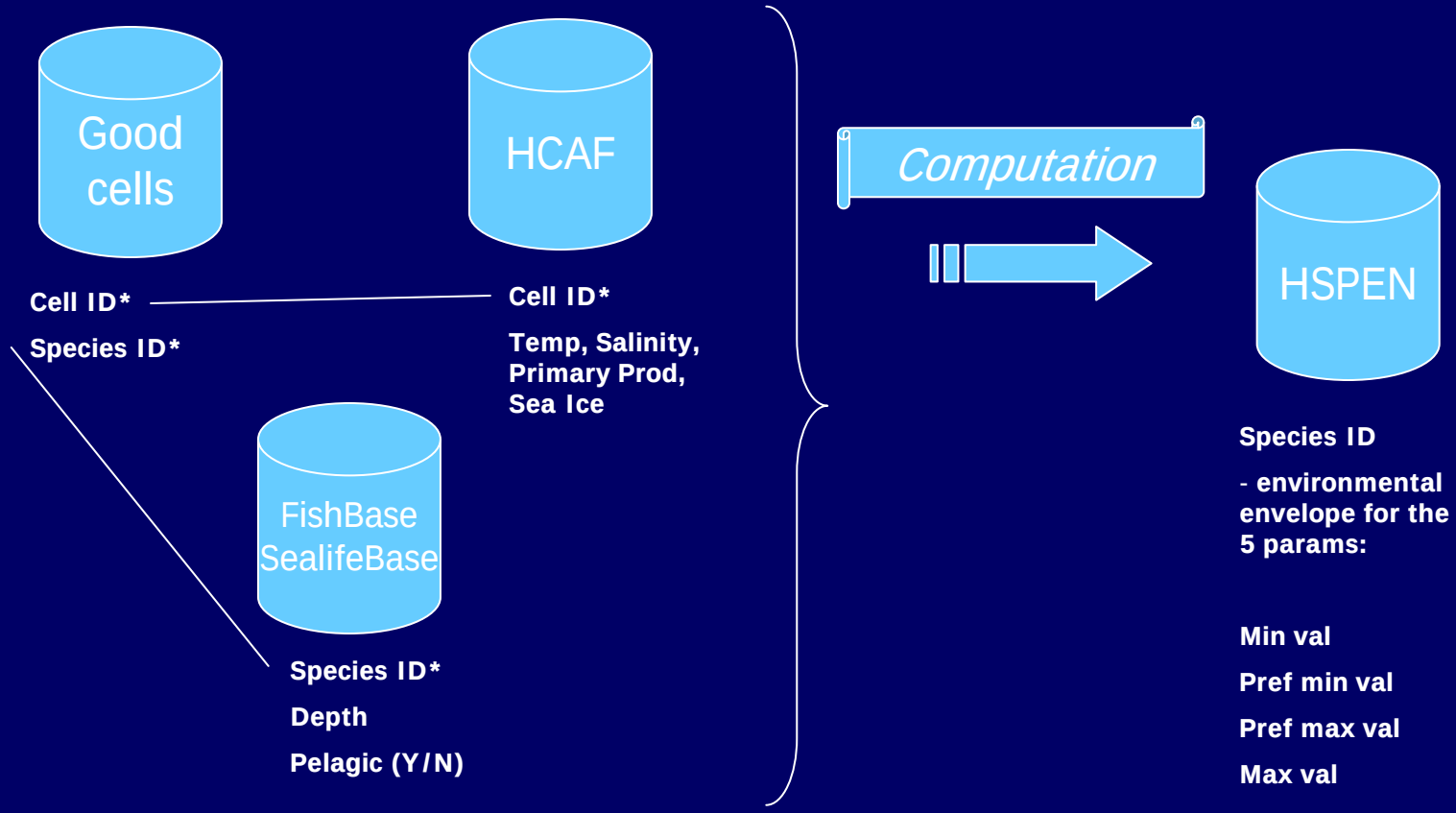
Point Data


FishBase
SealifeBase
GBIF
OBIS
Web service
XML
Parse
Data base
Point Data
Species ID
Decimal Lat
Decimal Lon

How to create Good Cells



HSPEN (Species Environmental Envelope)



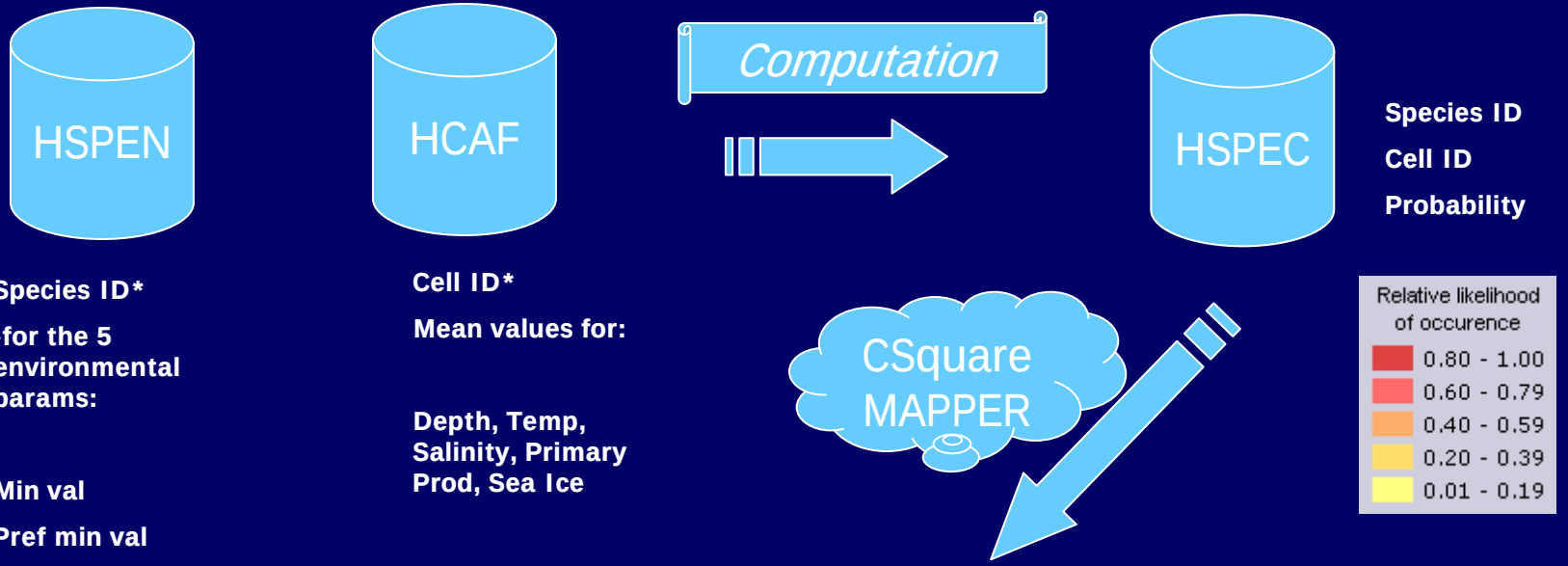
Envelope computation for each environmental parameter – HSPEN

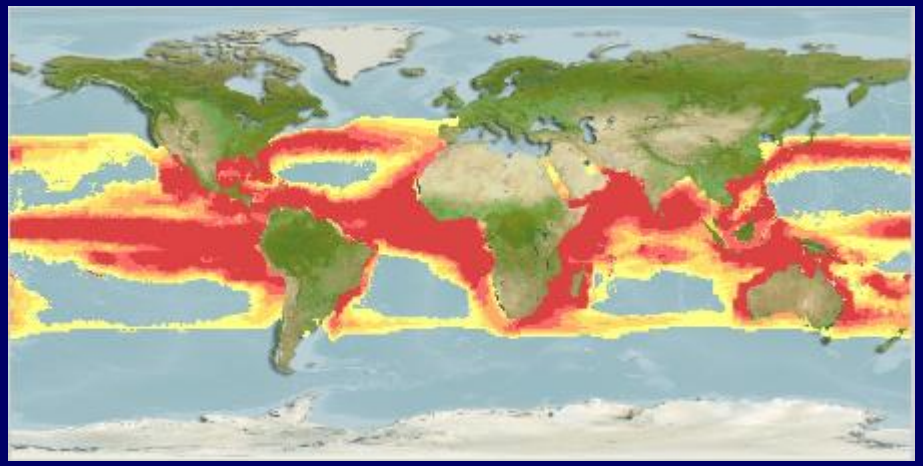
Step 1	Using the good cells, a. take the observed minimum value as the true min b. take the 10th percentile value as the preferred minimum c. take the 90th percentile value as the preferred maximum d. take the observed maximum value as the true maximum
Step 2	Get the extended min. value by: 25th percentile - 1.5 * interquartile range if observed min. > extended min., use extended min. value as the current min. Get the extended max. value by: 75th percentile + 1.5 * interquartile range if observed max < extended max., use extended max. value as the current max.

Step 3	Set a minimum for the preferred range by: If preferred max. – preferred min. < set minimum range then current pref. min. = MidPrefRange – .5 * set min. range current pref. max. = MidPrefRange + .5 * set min. range	where: MidPrefRange = preferred max. + preferred min. / 2 Set min. range varies depending on parameter: SST = 1 Salinity = 1 Primary Production = 2 Sea ice concentration = N.A. Distance to land = 2

Step 4	Set a minimum difference between the current min. and current preferred min. by:
	
	if current pref. min. – current min. < set min. differrence then
	current min. = current pref. min. - set min. differrence
	
	Set a minimum difference between the current max. and current preferred max.
	
	if current max. – current pref. max. < set min. differrence then
	current max. = current pref. max. + set min. differrence
		Set min. difference varies depending on parameter:
		SST = .5
		Salinity = .5
		Primary Production = 1
		Sea ice concentration = N.A.
		Distance to land = 1

HSPEC (Half-degree Species Assignment)





Probability computation - HSPEC

1. For each species, compare its envelope values (HSPEN) for each parameter with that of each cell in the HCAF table

2. Compute individual probabilities for each parameter as:

   0: if cell value < envelope minimum

   0: if cell value > envelpe maximum

   1: if cell value is within envelope preferred values

   Cell value – env. Min / pref. min – cell value:

   if cell value between envelope min. and preferred min.

   Env. Max. – cell value / cell value – pref. max :

   if cell value between envelope max. and preferred max.

3. Compute overall probability:

   Depth x SST x Primary Production x Sea ice concentration x Salinity

Creation of Richness Maps

cell_id	spec_id	probability
01	s_01	0.5
02	s_01	0.5
03	s_01	1
04	s_01	1
05	s_01	1
06	s_01	0.2

cell_id	spec_id	probability
01	s_02	1
02	s_02	1
03	s_02	1
04	s_02	0.2
05	s_02	0.2

cell_id	spec_id	probability
02	s_03	0.7
03	s_03	0.6

cell_id	spec_id	probability
02	s_04	1
03	s_04	0.2
04	s_04	0.6

cell_id	spec_id	probability
01	s_05	0.9
06	s_05	0.5
07	s_05	1

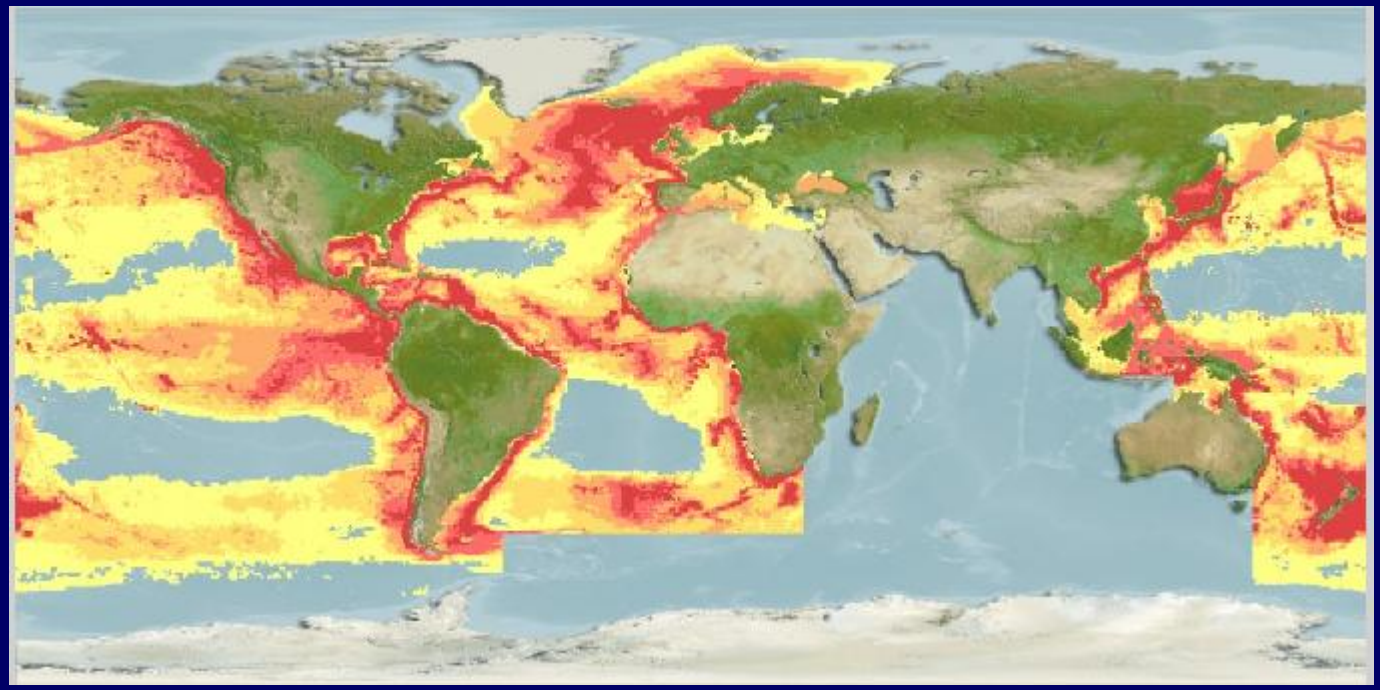
cell_id	count of species
03	4
02	4
04	3
01	3
06	2
05	2
07	1

```
SELECT map_data.cell_id,
Count(map_data.spec_id) AS [count of species]
FROM map_data
GROUP BY map_data.cell_id
ORDER BY Count(map_data.spec_id) DESC;
```

cell_id	count of species	ratio
03	4	1
02	4	1
04	3	0.75
01	3	0.75
06	2	0.5
05	2	0.5
07	1	0.25

```
SELECT map_data.cell_id,
Count(map_data.spec_id) AS [count of species],
Count([map_data].[spec_id])/4 AS ratio
FROM map_data
GROUP BY map_data.cell_id
ORDER BY Count(map_data.spec_id) DESC;
```

Delphinus delphis
Common dolphin



85,431 half degree cells – native range

111,645 half degree cells – suitable range

Bottlenecks in AquaMaps

MAINTENANCE
Size of the map data (HSPEC) ➡

Re-generate all data files used in the web-interface every time a species is included or excluded.

Maintenance of different checklists. ➡

Re-generate maps every time the algorithm changes.

WEB INTERFACE
On-demand richness maps (family, country, ecosystem, bounding box, etc.) ➡

AquaMaps deliverables in the course of 3 years:

(1) AquaMaps for 50,000 species being served through improved e-infrastructures;

(2) E-infrastructures allowing for real-time cross-species map generation and analysis (major computational challenge because of the high number of relational records to be processed);

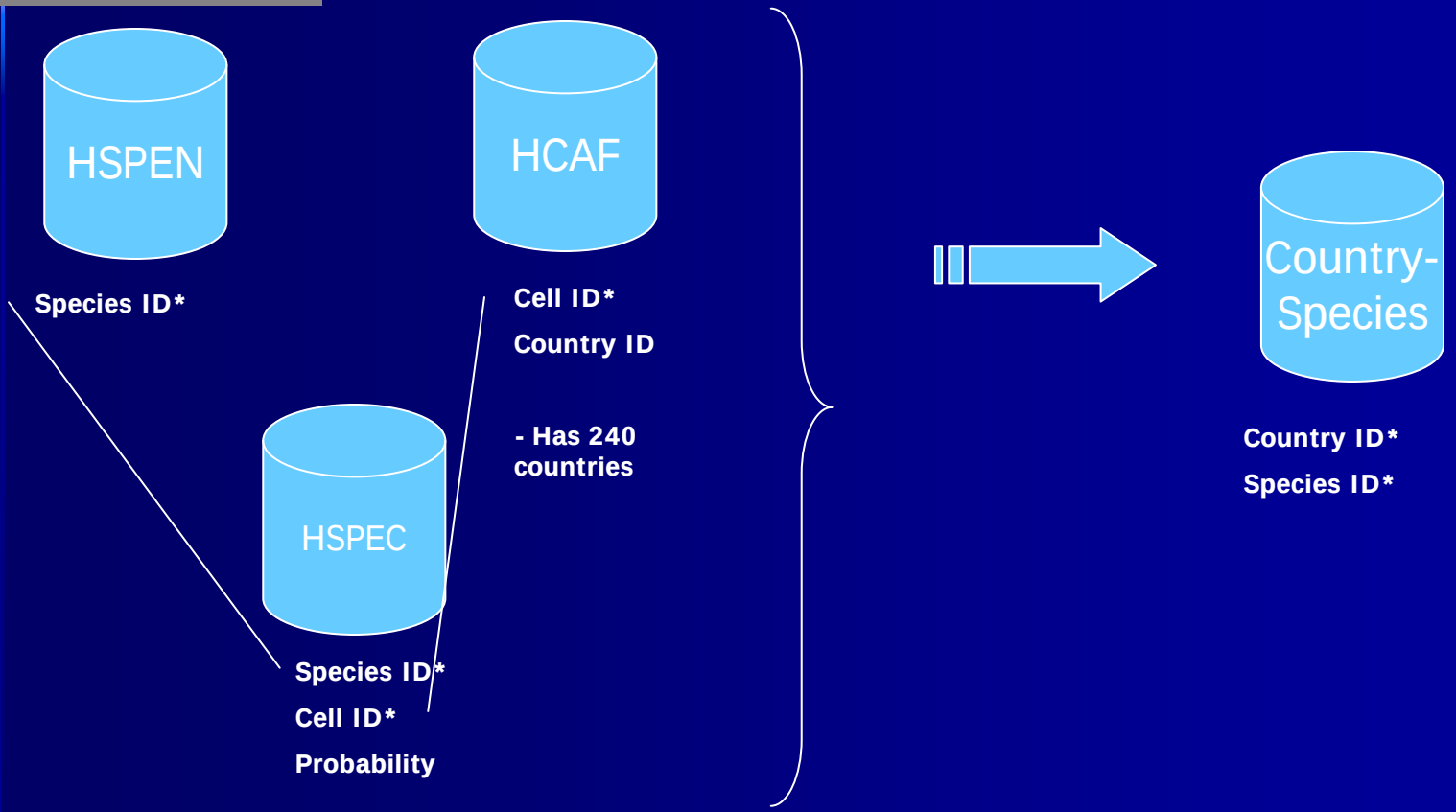
(3) Provision of predictive maps for future distributions under different climate scenarios;

(4) AquaMaps-based applications for mobile phones (What species are most likely around me? Guide me to my favorite species!)

country_species : Table

country_id	spec_id
01	s_01
01	s_02
02	s_01
02	s_02
03	s_01
03	s_03
04	s_03
05	s_03

List of species found on a country



AquaMaps Platform

Front-end interface	PHP	5.0.4 – 5.2.6
Back-end database	MySQL	5.0.18 – 5.0.51
Web server	Apache	1.3.19 – 2.2.8
Operating System	Windows	server, XP – any version

Coding style:
- Not object oriented, more on procedural
- at times especially on computation it is linear programming

GROUP	# OF SPECIES
Fish	6,956
Invertebrates	957
Marine mammals	115
TOTAL	8,028

HSPEC native range = 56,468,301

HSPEC suitable range = 114,989,360

Estimate for 50,000 species:

HSPEC native range = 350,000,000

HSPEC suitable range = 715,000,000

CSquare Mapper

http://www.marine.csiro.au/csquares/
Mapper written in Perl
CSIRO Marine and Atmospheric Research (CMAR)
Tony Rees as contact person

Sample CSquare codes

7501:469:3	half-degree cell
7501:469	1-degree cell
7501:4	5-degree cell
7501	10-degree cell

- Invoke the mapper by sending a <form> submit
http://www.marine.csiro.au/csquares/activemap_test.html

AquaMaps Platform

Front-end interface	PHP	5.0.4 – 5.2.6
Back-end database	MySQL	5.0.18 – 5.0.51
Web server	Apache	1.3.19 – 2.2.8
Operating System	Windows	server, XP – any version

Coding style:
- Not object oriented, more on procedural
- at times especially on computation it is linear programming

Thank You