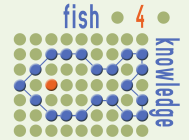
Fish4Knowledge Project Meeting

December 2011 - Catania



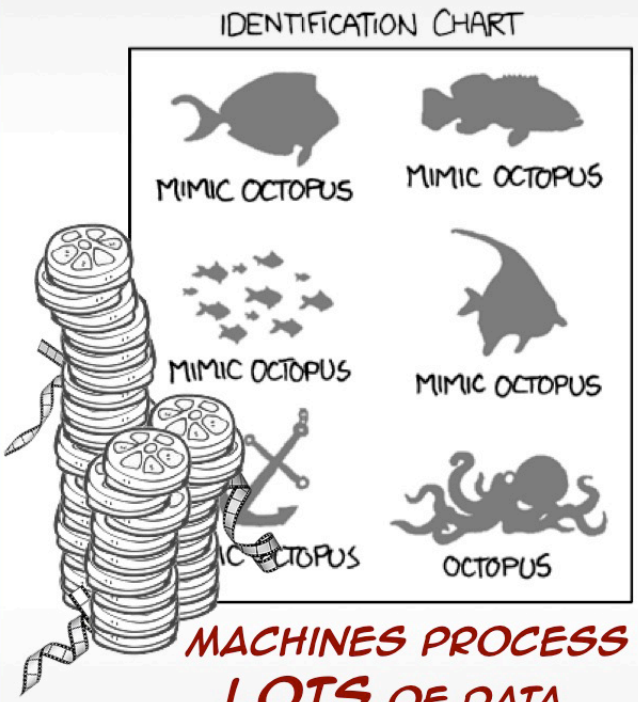
REPORT ON WP2

Interactive User Query Interface

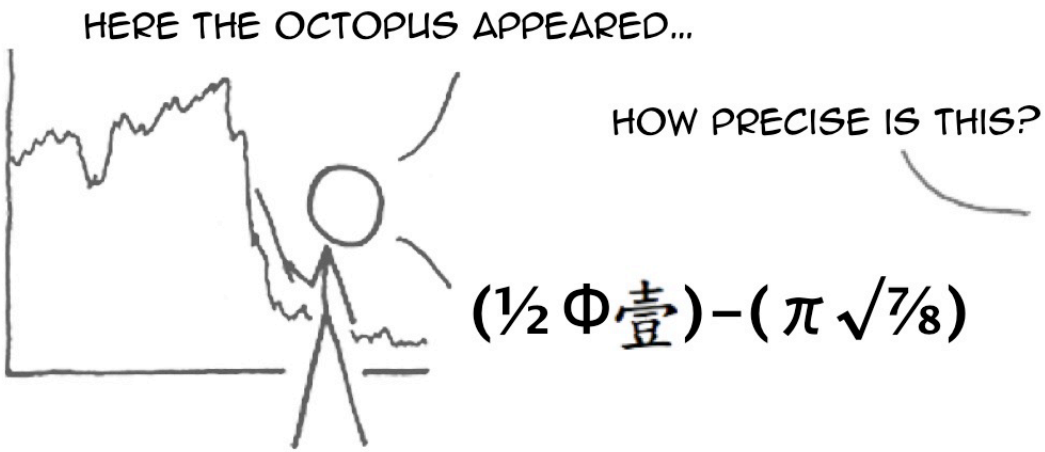
User Information Needs,
User Scenarios and
Implementation Plan

USER INTERFACE FOR COMPREHENSIVE DATA ANALYSIS

FOR INSTANCE, GIVE VIDEO ANALYSIS DATA TO BIOLOGISTS....



MACHINES PROCESS **LOTS** OF DATA...



HUMANS NEED TO UNDERSTAND. AND TRUST.

Many thanks to XKCD: xkcd.com/928/

THROUGH EMPIRICAL RESEARCH, WE STUDY INTERACTIVE MEANS TO ACCESS AND CONTROL:

- DOMAIN-ORIENTED INFORMATION
- UNCERTAINTY CONTAINED IN DATA
- UNDERLYING PROCESSES

CONTENT OF THE PRESENTATION

Objective 2.1

- User information needs: 20 questions

Objective 2.2

- User scenarios: interacting with big, complex and imperfect data sets

USER INFORMATION NEEDS

Results reported in Deliverable 2.1

USER INTERVIEWS

- Interviews with marine biology experts
  - Prof. Shao, Taiwan
  - Dr. Day, Caribbean
  - Prof. Stergiou Greece

- Established list of “20 questions” (à la Jim Gray) that represent the most important marine biologist queries

THE “20 QUESTIONS” 1/3

Q1	How many species appear and their abundance and body size in day and night including sunrise and sunset period.
Q2	How many species appear and their abundance and body size in certain period of time (day, week, month, season or year). Species composition change within one period.
Q3	Give the rank of above species, i.e., list them according to their abundance or dominance. How many percent are dominant (abundant), common, occasional and rare species.
Q4	Fish colour pattern change and fish behaviour in the night for diurnal fish and in daytime for nocturnal fishes.
Q5	Fish activity within one day (24 hours).
Q6	Feeding, predator-prey, territorial, reproduction (mating, spawning or nursing) or other social or interaction behaviour of various species.
Q7	Growth rate of certain species for a certain colony or group of observed fishes.
Q8	Population size change for certain species within a single period of time.
Q9	The relationship of above population size change or species composition change with environmental factors, such as turbidity, current velocity, water temperature, salinity, typhoon, surge or wave, pollution or other human impact or disturbance.

THE "20 QUESTIONS" 2/3

Q10	Immigration or emigration rate of one group of fish inside one monitoring station or one coral head.
Q11	Solitary, pairing or schooling behaviour of fishes.
Q12	Settle down time or recruitment season, body size and abundance for various fish.
Q13	In certain area or geographical region, how many species could be identified or recognized easily and how many species are difficult. The most important diagnostic character to distinguish some similar or sibling species.
Q14	Association among different fish species or fish-invertebrates.
Q15	Short term, mid-term or long term fish assemblage fluctuation at one monitoring station or comparison between experimental and control (MPA) station.
Q16	Comparison of the different study result between using diving observation or underwater real time video monitoring techniques. Or the advantage and disadvantage of using this new technique.
Q17	The difference of using different camera lens and different angle width.
Q18	Is it possible to do the same monitoring in the evening time.

THE "20 QUESTIONS" 3/3

Q19	How to clean the lens and solve the biofouling problem.
Q20	Hardware and information technique problem and the possible improvement based on current technology development and how much cost they are.
Q21	What is the average body size for species X? How many percent of fish are small, normal or big?
Q22	What is the number of fish in area X for indicative species related to pollution?
Q23	What is the distribution and number of fish for indicative species of factor X?
Q24	What is the analysis of factor X impact, using pattern of indicative data Y?
Q25	What are the areas and periods of time of species X migrations?
Q26	What are the areas and periods of time of species X SPAGS?
Q27	What are the SPAGS periods in area Y?

RESULTS OF "20 QUESTIONS" INTERVIEWS

Targeted biology topics of research:

- **Population dynamics**
  - Demographical measurements based on counts of fish
  - Middle to long-term studies, for various species and areas
- **Impact of events**
  - Correlation of Events and biology-specific measurements
  - Short to middle-term studies, for specific species and areas

Excluded biology topics of research:

- **Tropic systems, Reproduction, Migration**
  - Need more cameras than available
  - Need data that cannot be precisely extracted from video (fish size, specific behaviours, cryptic fish, night time data)

USER SCENARIOS

Details in Deliverable 2.2

CHARLES - THE STUDY OF FISH ABUNDANCE OVER TIME

Charles' actions:

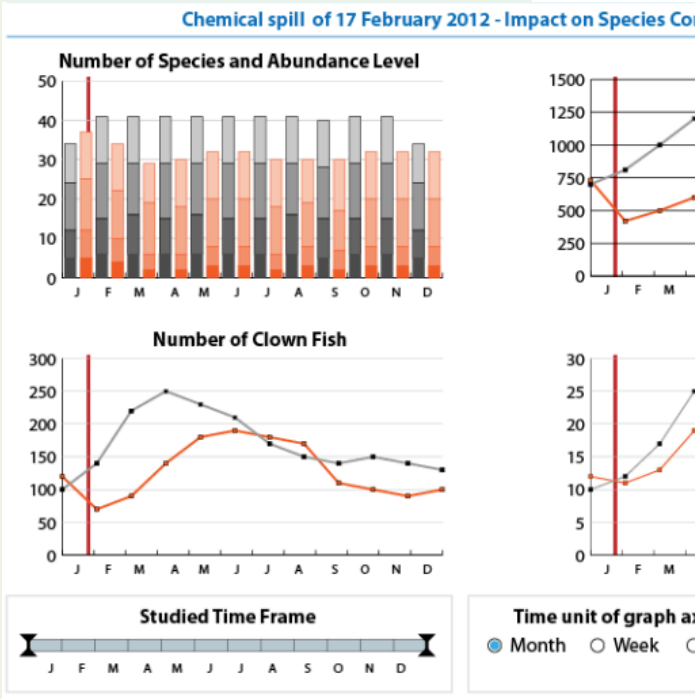
- Explore the count of fish over 3 years
- Spot irregularities in 2010
- Control the visualisation
- Exclude technical explanations
- Study abundance per area versus average abundance
- Find area that influenced the overall abundance
- Customise and share visualisation

1. CHARLES - THE STUDY OF FISH ABUNDANCE OVER TIME

- Explore the count of fish over 3 years

Charles wants the overall abundance, i.e. the total number of fish across all species, for each week over the last 3 years.

He uses the built-in measurement overall abundance and indicates the time unit (week) and the period of interest (January 2010 until December 2012). The F4K tool creates a graphic visualisation of the overall abundance.

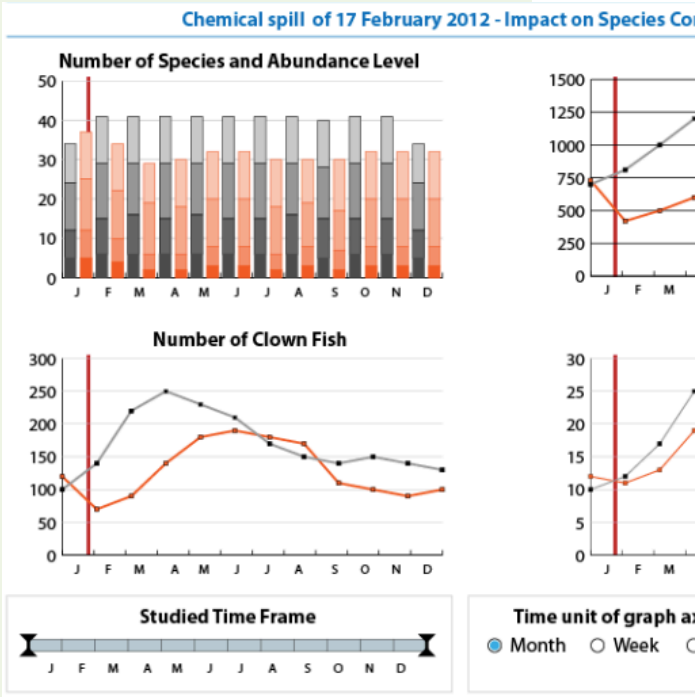


2. CHARLES - THE STUDY OF FISH ABUNDANCE OVER TIME

- Spot irregularities in 2010

Charles is a bit surprised by the overall abundance in 2010. He switches to a weekly evaluation of the overall abundance for 2010 only. He sees that the figures for the month of March are particularly surprising.

Charles modifies the timeframe of study and the time unit to get a daily evaluation of the overall abundance for the month of March 2010.



3. CHARLES - THE STUDY OF FISH ABUNDANCE OVER TIME

- Check the visualisation of overall abundance

Charles does not trust the consistency of the video processed and wants to check this. He starts an explorative view and sees information about:

- What videos were taken into account
- What software components analysed the videos and produced the recognition of fish
- What computer vision data were used for the visualisation
- What domain-specific metrics were calculated, which formula were used, how data were used.
- How data are represented in the visualisation, e.g., the type of diagram, how data are plotted in the diagram.

4. CHARLES - THE STUDY OF FISH ABUNDANCE OVER TIME

Checking the videos

Charles wants to check if a homogeneous set of cameras and videos were available at the beginning of 2010. He asks for more details about the videos that were taken into account. He looks for inconsistencies in the number of cameras or videos available from each location e.g., if cameras were in maintenance, if a significant number of videos are missing or bad quality.

Charles can select a custom set of cameras or videos to get a homogeneous set of videos, but the set is already consistent enough. Charles also wants to verify if the quality of videos is consistent. He inspects a few videos from 2010, and he even looks for video in 2011 to verify if the quality is similar, and if fishes are accurately identified.

5. CHARLES - THE STUDY OF FISH ABUNDANCE OVER TIME

- Control the video analysis components

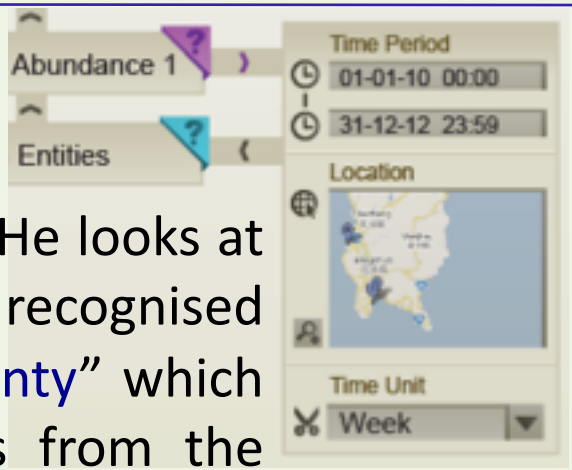
Charles checks the software components that analysed the videos, to see if they are responsible for the irregularities in the fish abundance in 2010.

He looks for differences in the versions of the components. The Fish Detection component had a major version change in September 2010. To check the influence of this, he launches an analysis of the video before September 2010 with the new version of the Fish Detection component.

When this is completed, Charles sees that the differences between the overall abundance in 2010 and the other years remain unchanged.

6. CHARLES - THE STUDY OF FISH ABUNDANCE OVER TIME

- Control the selected data set



Charles checks the data directly from the database. He looks at the table named “Fish” that contains one entry per recognised fish. He notices the column called “Detection certainty” which specifies the certainty threshold to select entries from the “Fish” table.

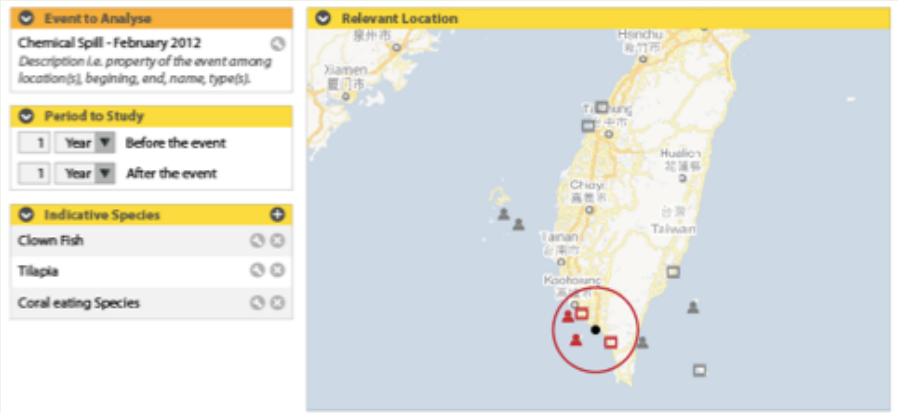
Charles decides to set a higher certainty threshold to see how it influences the results. The corresponding overall abundance has lower values, but there are no major changes in the trends and the results for 2010 are still different from the other years. He confirms this for a lower certainty threshold as well.

7. CHARLES - THE STUDY OF FISH ABUNDANCE OVER TIME

- Find area that influenced the results

Charles found no technical reason to explain the differences that appear for the year 2010, but now he knows exactly what is represented in the visualisation, and he trusts the findings of the system.

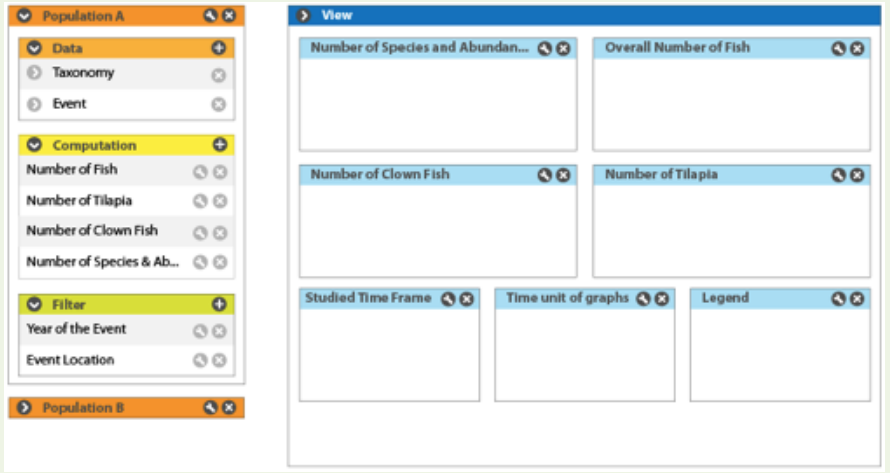
Charles starts thinking about biological reasons that influenced the population dynamics in 2010. He wonders if the changes in 2010 occurred for the whole island of Taiwan, or if they were localised in a specific area. He calculates the overall abundance for various areas. He finds that major irregularities occurred in the eastern part of Taiwan and this influenced the national overall abundance.



8. CHARLES - THE STUDY OF FISH ABUNDANCE OVER TIME

Customise and share visualisation

Charles wants to compare the overall abundance in the eastern area in 2010, with the average overall abundance in the same area during 2011 and 2012. He checks the calculations of the overall abundance. He adds the evaluation of the average overall abundance during 2011 and 2012 by defining a custom data set to use. He customises the visualisation to get a representation that allows him to best compare the overall abundance in 2010 and the average overall abundance in 2011-2012.



CHARLES - THE STUDY OF FISH ABUNDANCE OVER TIME

Charles' actions:

- Explore the count of fish over 3 years
- Spot irregularities in 2010
- Control the visualisation
- Exclude technical explanations
- Study abundance per area versus average abundance
- Find area that influenced the overall abundance
- Customise and share visualisation

ERICA - THE STUDY OF ENVIRONMENTAL CONDITIONS

Erica's actions:

- Explore Charles' visualisation
- Research and define environmental events to study
- Define species to study w.r.t. events to study
- Explore and compare several visualisations
- Find correlations with 3 events in a row
- Create visualisation and measurements to illustrate findings
- Share the visualisation with Charles

WRAP UP: MAIN DESIGN PROBLEM

- Marine biologists need to trust their data
- In F4K, data is generated by algorithms they do not understand
- Data generated by algorithms are imperfect

➢ The goal for the system UI design is to **empower marine biologists to assess whether data supports scientifically valid findings**

YEAR 2

Objective 2.2 - Create initial component-based prototypes to establish their usefulness.

- Create interface mockups and UI prototypes based on scenarios
- Evaluate these with users

Objective 2.3 - Create advanced prototypes that fulfill specific high-level user information needs.

- Coordinate with other WPs to create working end-to-end scenario

YEAR 3

Objective 2.4 - Integrate user interface components into end-to-end system.