

# Enhancing the Quality and Trust of Citizen Science Data



**Abdul Alabri**  
**eResearch Lab**  
**School of ITEE, UQ**

# Citizen Science

- **Citizen Scientist:** refers to a volunteer who collects and/or processes data to contribute to scientific research.
- e.g. astronomy, bird watching, water and air quality, reef watching and endangered species monitoring.
- Growing rapidly because
  - Internet, Social networking
  - Increased awareness – climate change
  - Availability of technical tools
  - Free labour, skills, computational power
  - Funding tied to projects that encourage community participation



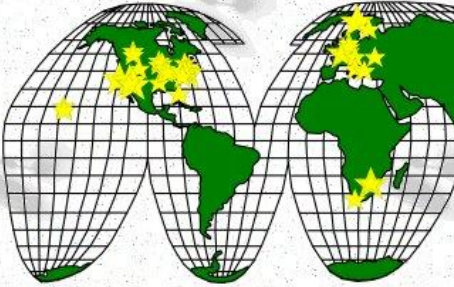
# Examples

- The Internet Bird Collection
  - Non-profit project
  - Providing information about the world's avifauna.
  - Collect video, audio and photos of birds
  - Audiovisual library of the world's birds free of charge.
  - Online community – social network
- The NatureMapping Foundation
  - Non-profit project
  - Monitoring biodiversity
  - Free nature biodiversity database to all
  - Contact through the web, schools and universities



# Examples cont...

*Center for Backyard Astrophysics*



**A global network of small telescopes to photometry of cataclysmic variables**

Welcome!  
cbastro.org

Waterwatch Australia home page

Australian Government

Waterwatch home | About Waterwatch | Get involved | Monitoring water quality | Events | Publications

**Waterwatch Australia**  
Waterwatch is a national community water quality monitoring network that encourages all Australians to monitor their waterways. [More about Waterwatch](#)

**How to get involved**  
Anybody that cares about the local environment can get involved in Waterwatch. [More about getting involved](#)

**Monitoring water quality**  
Monitoring water quality gives us a picture of catchment health and can assist with the rehabilitation of waterways. [More about monitoring water quality](#)

**Grants and funding**  
Waterwatch Australia does not provide funding for water monitoring or conservation activities. Funding is available through Australian Government initiatives:

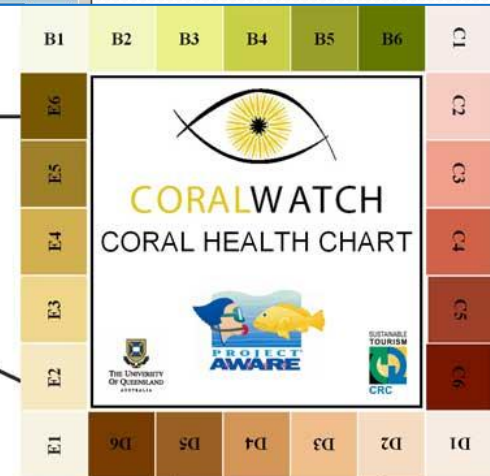
- [Australian Government's Community Water Grants](#)
- [Australian Government Envirofund](#)
- [Australian Government Water Fund](#)
- [Environment and heritage grants and funding](#)
- [GrantsLINK](#)

Waterwatch home | About Waterwatch | Get involved | Monitoring water quality | Events | Publications | Contact Us | Accessibility | Disclaimer | Privacy | © Commonwealth of Australia  
Last updated: Wednesday, 10-Sep-2008 15:42:26 EST  
Waterwatch Australia: Communities caring for catchments

Staghorn coral after recovery in May 2002



Staghorn coral during bleaching in March 2002



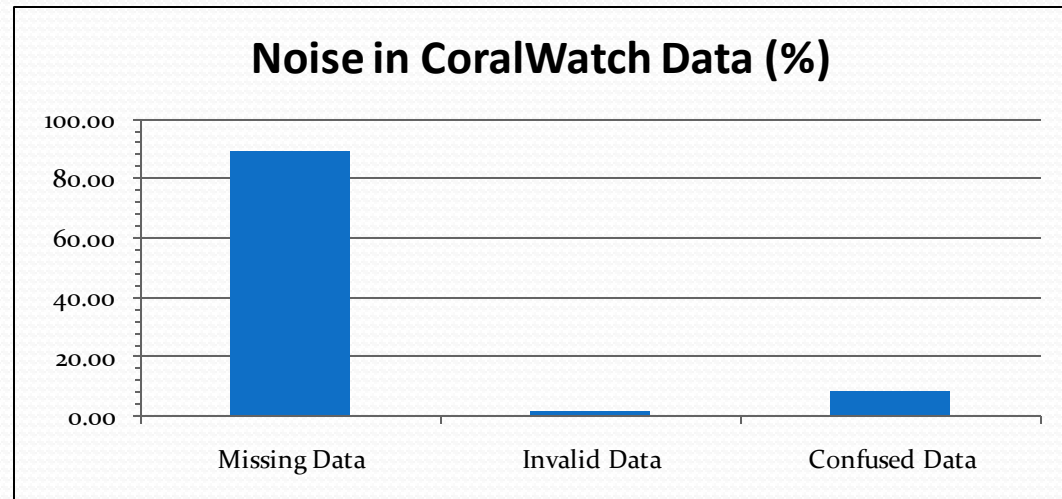
Brain coral after recovery in May 2002



Brain coral after recovery in May 2002

# Challenges

- Poor data quality
  - Absence of “scientific method”
  - Insufficient training
  - Lack of tools to identify outliers
    - automatically compare overlapping or complementary data sets
    - Non-standard and poorly designed tools and formats
- Potential anonymity – lack of authentication of users
- No measure of data reliability/certainty
- Lack of trust on data by scientists
- Limited filtering and visualisation services
- Lack of appropriate feedback
- Lack of volunteers – attracting and retaining



# Aims

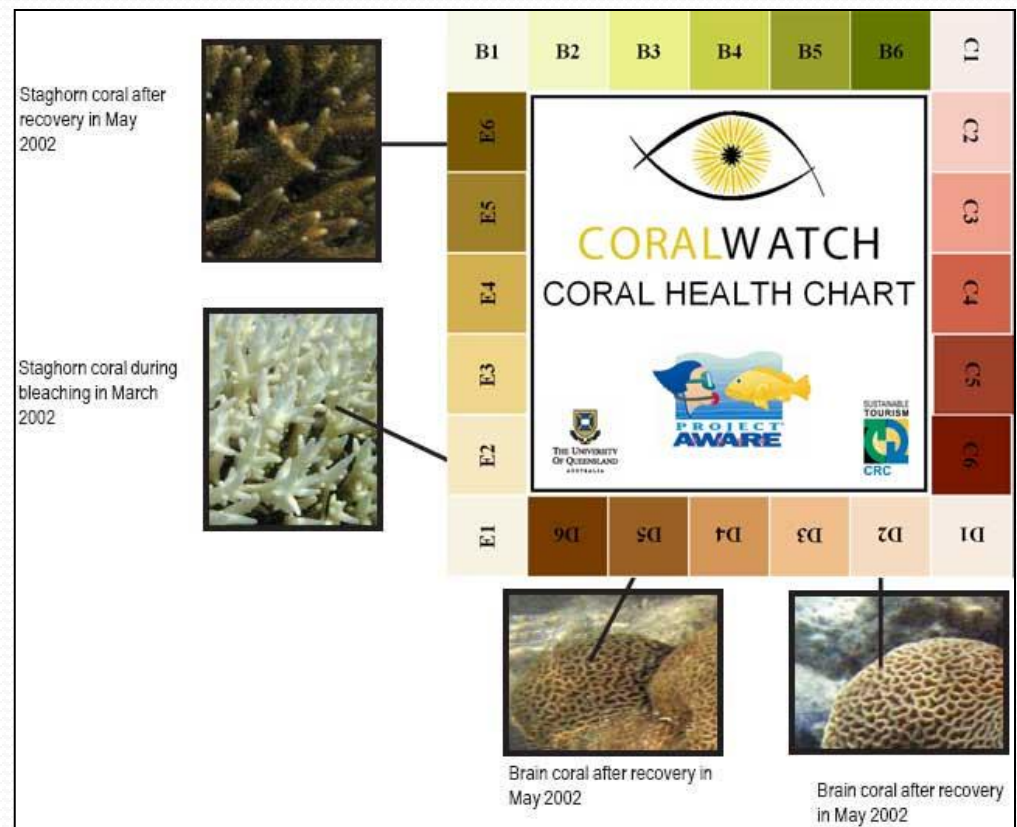
- **Quality:** improve the quality and reliability of the data/metadata without adversely impacting on the complexity or usability of the data capture tools.
  - Controlled vocabularies/schemas
  - Automate data capture e.g. GPS location/date/contributor
  - Automatic validation (XML Schemas) on input
  - Identify gaps in data – encourage volunteers specifically in these areas
  - Consistency across datasets from different sources
  - Identify and remove malicious data
- **Trust:** address the low level of trust associated with citizen science data as perceived by the scientific community; ways to measure trust, display explicitly and take into account in decision-support
  - Rank users - reliability/trust
  - Rank reliability of datasets
  - Filter searches based on data reliability
- **Understand the optimum interaction/balance between quality improvement and trust metric services**

# Case Study CORALWATCH

- Citizen science project aims to “*improve the extent of information on coral bleaching events and coral bleaching trends*”
- Non-profit organisation based at UQ
- 1300 volunteers around the world (80 Countries)
- Publications (Books, CDs, Presentations etc)
- Website: <http://coralwatch.org>
- New website published 2 weeks ago

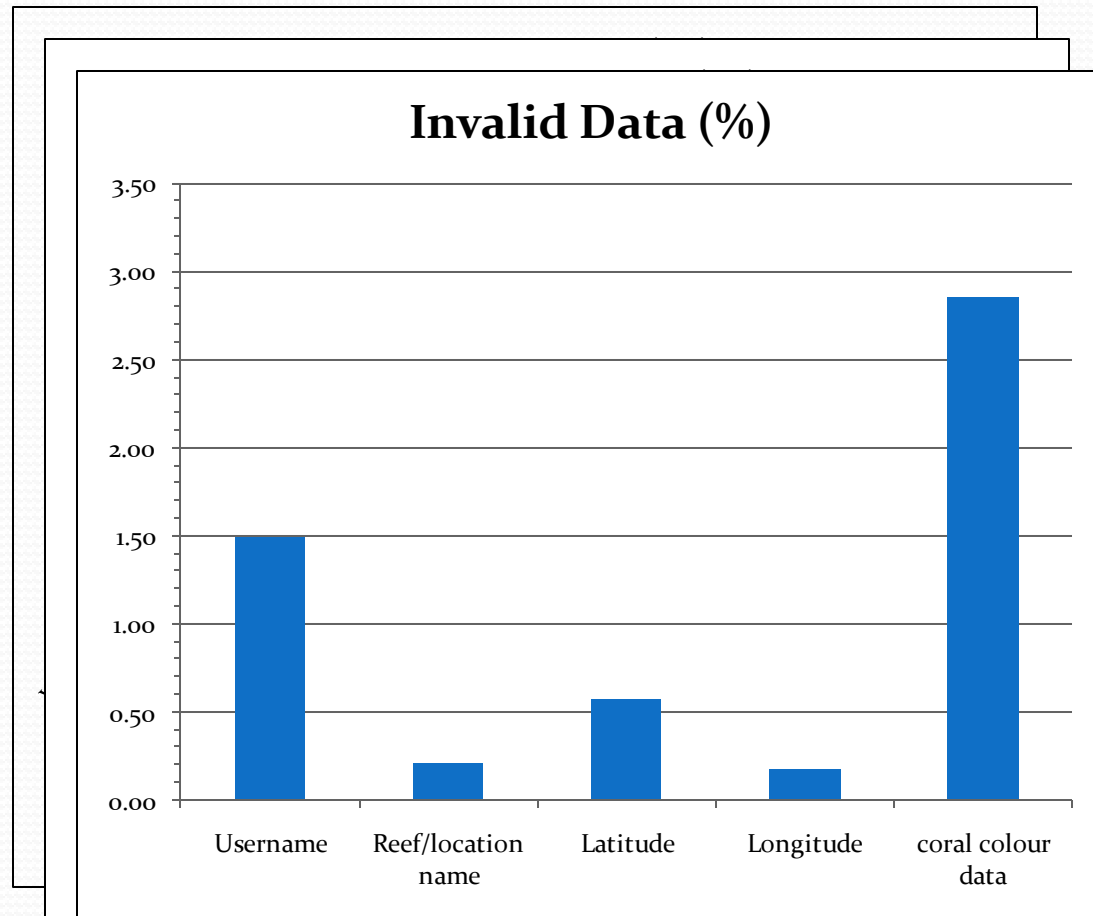
# CoralWatch Tools and Techniques

- Coral Health Chart
- Datasheet
- Reef education package
- Excel spreadsheet
- Online data entry form



# Issues with CoralWatch Data

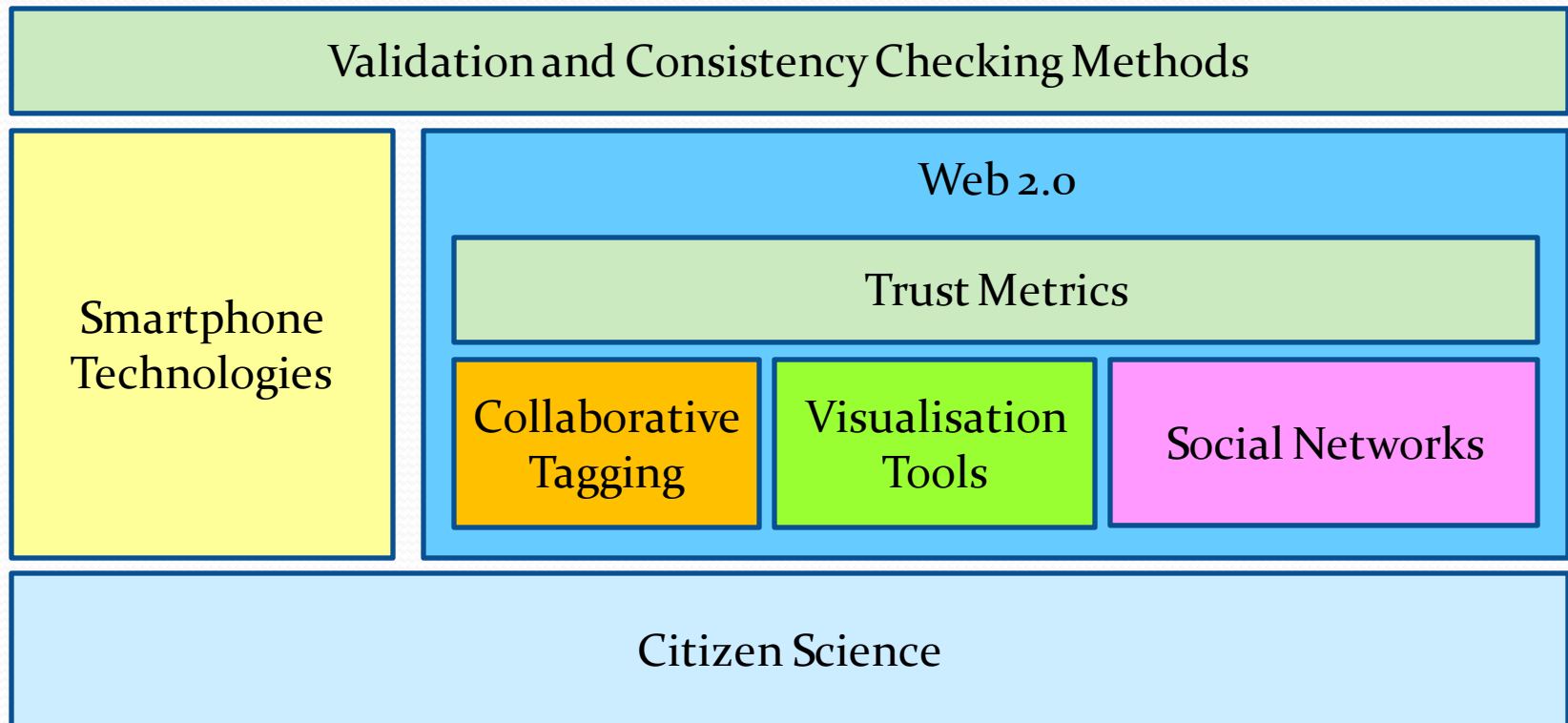
- July 2003 to Sep 2009
- 18569 Records
- No Authentication
- No Validation
- No data model
- 64% of GPS records missing



Light Colour (E6) Dark Colour (E1)

# Methodology

- Develop a technological framework for enhancing the quality and reliability of citizen science data



# Metadata and Data Validation

- Aim: improving the quality of submitted data
- Validation and handling of errors at the submission process
- User friendly interface with strict validation rules
- Metadata standards e.g. Dublin Core, RDF/XML Schemas
- Controlled vocabularies, Value ranges/formats
- Ontologies/trend analysis to cross check with other data
  - e.g. Compare citizen science data with sensor or satellite data.
- Authentication and authorisation

NAME	EMAIL	COUNTRY	DATE	TIME	REEFNAME	WEATHER	TYPE	LIGHTEST	DARKEST	TEMPERATURE	LATITUDE	LONGITUDE
NULL	NULL	Australia	12/08/2004	00:00	Heron Island	Full Sunshine	Plate	E1	E4	0	NULL	NULL

# Trust Metrics

- *“Trust in a person is a commitment to an action based on a belief that the future actions of that person will lead to a good outcome.” (Golbeck, 2009).*
- Used in online community sites
  - e.g. Blogs, Facebook, eBay, Amazon.com
- Challenges/Questions—
  - Subjective: Web-based social trust must be focused and simplified.
  - Not Binary: value within range e.g. Ratings
  - Entering trust values for all people/datasets in a network is time-consuming - dealing with people you don't know
  - Can you infer data is reliable if person is trusted?
  - Best algorithms for measuring trust of person/data from multiple metrics?
  - How to measure changing trust values over time?

# Trust Metrics cont.


- Recommender System
  - Aim: Finding reliable and trusted data
  - e.g. movie ratings, amazon.com
- Generate a predictive trust value between users
- Calculate trust transitivity

$$t_{is} = \frac{\sum_{j \in \text{adj}(i)} t_{ij} t_{js}}{\sum_{j \in \text{adj}(i)} t_{ij}}$$

Home » Dashboard » Users » Bakarr

Details Surveys

<b>Display Name:</b>	Bakarr
<b>Email:</b>	<a href="mailto:bakar@coralwatch.org">bakar@coralwatch.org</a>
<b>Member since (d/m/y):</b>	07/02/2010
<b>Occupation:</b>	Biology Student
<b>Address:</b>	3/351 Coronation Dr Toowong QLD 4044
<b>Country:</b>	Australia
<b>Surveys:</b>	3 surveys
<b>Photos:</b>	No photos yet
<b>Videos:</b>	No videos yet
<b>Community Trust:</b>	★★★★☆ (2.5)
<b>Your Trust:</b>	★★★★☆ (4)



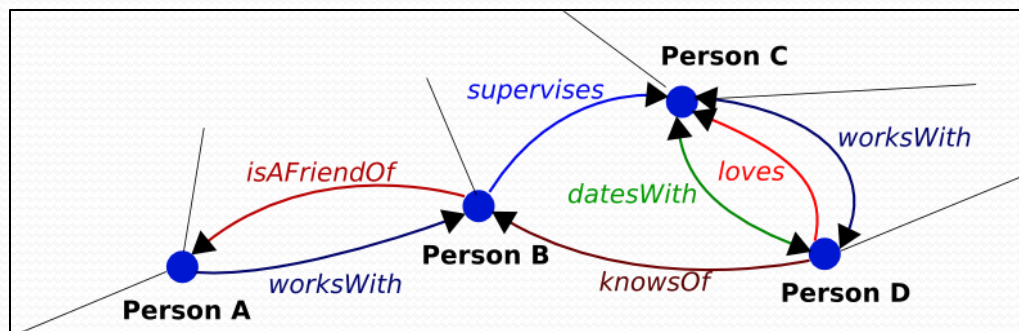
# Trust Metrics cont.

Accumulative trust value of a user is based on:

- Expertise of the member – role, qualifications
- The member's frequency and duration of participation (number of surveys, images, videos, comments)
- Trust ranking from other members (1 – 5 stars)
- Social network analysis (FOAF)
- Quality of past data contributed

Accumulative trust value of survey is based on:


- Direct rating from other members
- Inferred rating from contributor's rating
- Consistency with related data (Reef Check, Satellite Data)



# Reporting and Visualisation

- Enable the synthesis and understanding of citizen science data
- Educate the volunteers about implications of their data  
“The big picture”
- Reporting services - using geospatial & statistical (R) tools
- Enable searching, querying and filtering
- Take into account trust/ranking of data

# Reporting and Visualisation



Home   Graphs   Education

**CoralWatch Surveys**

Find Location:

- 185 Coral Watchers
- 26 Coral Reefs
- 254 Surveys

Filters

Country: All Countries

Reef Name: All Reefs

Rating: All Stars

Options


- Dashboard
- Upload Data
- Request Kit

Legend

- 5 stars rating
- 4 stars rating
- 3 stars rating
- 2 stars rating
- 1 star rating

Updates on

twitter   facebook



Details   Surveys

Home » Dashboard » Users » Bakar

**Display Name:** Bakar

**Email:** [bakar@coralwatch.org](mailto:bakar@coralwatch.org)

**Member since (d/m/y):** 14/01/2010

**Occupation:**

**Address:**

**Country:**

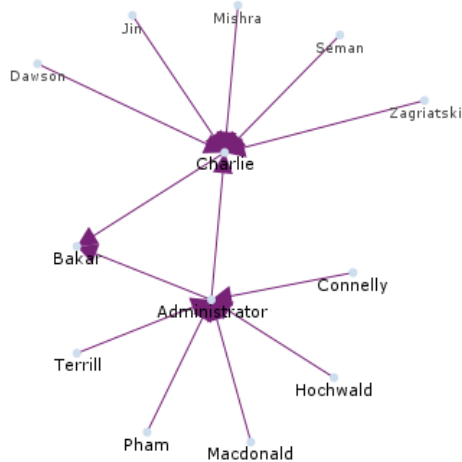

**Surveys:** 3 surveys

**Photos:** No photos yet

**Videos:** No videos yet

**Community Trust:** ★★★★★ (2.6) [Trust Cloud](#) [Trust Network](#)

**Your Trust:** ★★★★★ (1)



Network graph showing connections between users: Charlie (central node), Mishra, Seman, Zagriatski, Connelly, Hochwald, Macdonald, Pham, Terrill, Administrator, Bakar, Dawson, Jin, Glover, Bunting, Muir, Sheld, Karunajeewa, Sheld, Pedersen, Jhan, Dumya, Lim, Vasudavan, H, Rafiq, Zhang, Hangrave, Modliam, Wu, Carter, Kamaruddin, Ribaibovich, Elzad, Madden, Aaz, Jadav, D.

# Evaluation

- Assessment criteria
  - Improvements in data quality – optimize the weightings and algorithms for calculating the aggregate trust/quality metric
  - Performance and efficiency of the tools
  - Scalability and adaptability
  - Usability tests
  - User feedback
    - Volunteers
    - Project managers
    - Scientists
- Methods
  - Automatic monitoring/logging of usage
  - Error detection
  - precision before and after – compare with benchmark (ground truth) data
  - Conduct surveys and interviews with stakeholders/users

# Conclusion

- Citizen science movement is rapidly expanding across many disciplines – astronomy, environmental, marine
- Inherent weaknesses and challenges
- Critical need for automatic techniques to improve the quality and trust of citizen science data
- Data quality and social trust metrics can potentially be combined and applied to improve the reliability of citizen science data.
- Providing reporting and visualization tools enables stakeholders to better synthesize and understand citizen science data.

# Acknowledgements

- Supervisors
  - Prof. Jane Hunter
  - Assoc. Prof. Eva Abal
- eResearch Lab's members
- CoralWatch organizers and members
- SEQ Healthy Waterways Partnership
- ARC Linkage LP0882957

# Questions?

- Contact

- Abdulmonem Alabri: [alabri@itee.uq.edu.au](mailto:alabri@itee.uq.edu.au)
- Coralwatch: [info@coralwatch.org](mailto:info@coralwatch.org)

- Websites

- eResearch Lab: <http://itee.uq.edu.au/~eresearch>
- CoralWatch: <http://coralwatch.org>