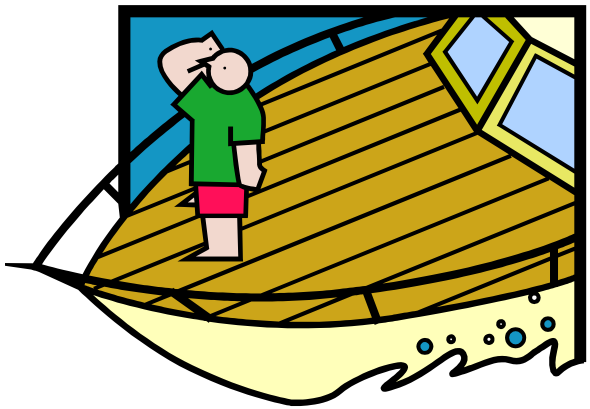
Can the maritime crowd be trusted for sourcing Bathymetric Data?

Brendan Mason – Envitia
Tim Clarke - Dstl

Can we trust the maritime crowd?

Yes



Maybe

No

Data Types for this Study?

- Authoritative – provided by a recognised or official production source.
- Institutional – provided by academic bodies
- Crowdsourced – provided by non-professional or volunteer bodies.

Crowdsourced datasets

The following crowdsourced datasets were reviewed. (These were constrained by relevance to the maritime domain.)

OpenSeaMap	www.OpenSeaMap.org
TeamSurv	www.teamsurv.eu
ActiveCaptain	www.activecaptain.com
Navionics	www.navionics.com
Google Ocean	www.earth.google.com
Icewatch	www.naturewatch.ca/english/icewatch/
Olex	www.olex.no/teknikk_e.html

Issues with Crowdsourced Data

Automated measurement

- visually observed measures are not accurate or consistent

Automated location detection

- clicking on a map may not be accurate or consistent

Structured attribution input

- free text input is inconsistent

Coverage

- coverage in the maritime areas are poor

User training

- need confidence in user input particularly for non-automated input

Licencing

- need clarity of licence criteria

Data Availability

- data accessible in bulk

(Source: ©Tim Sheerman-Chase)

Selection of Crowdsourced Data

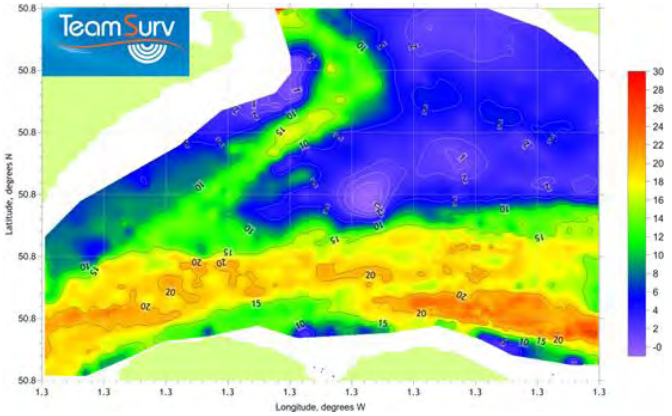
	Availability	Features of Interest	Product Specification	Automatic location	Automatic Measurement	Formalised attribution	Trained Users
OpenSeaMap	N	Y	Partial	N	N	Partial	Partial
TeamSurv	Y	Y	Partial	Y	Y	N	Partial
ActiveCaptain	N	Y	N	N	N	Y	N
Navionics	N	Y	N	N	Partial	Y	N
Google Earth\Ocean	N	Y	N	N	N	N	N
Icewatch	Y	N	N	N	N	Partial	Partial

And the Winner Is?

TeamSurv

- TeamSurv is the public face of the EU sponsored CoSuDEC project, which investigated the use of crowdsourcing to log position and depth data from seagoing vessels, and then to process that data to enhance the quality of nautical charts, and data sets for GIS systems. It is primarily focused on littoral areas where modern multi-beam survey equipment has difficulty getting close inshore.

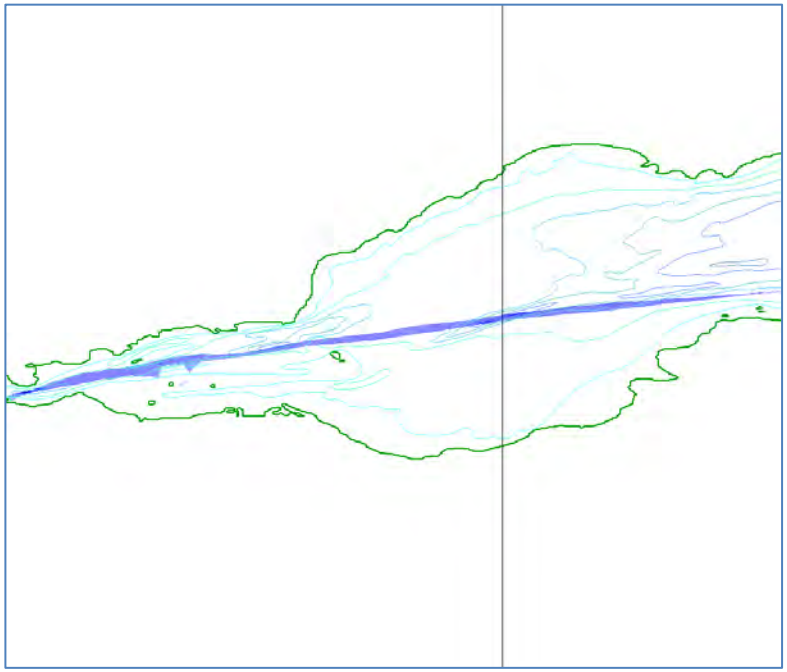
- The TeamSurv community collects information from small vessels by making use of hardware or software loggers attached to on-board NEMA0183 instrument feeds from leisure or commercial vessels (see www.Teamsurv.eu).



(Source: ©TeamSurv)

TeamSurv Sample Dataset

- Crowdsourced data for the Firth of Forth was obtained from TeamSurv providing coverage of the area 3.41434,55.9948 to -2.76844,56.0866 (WGS84). This provides in excess of 5300 grid cells in the area of interest.



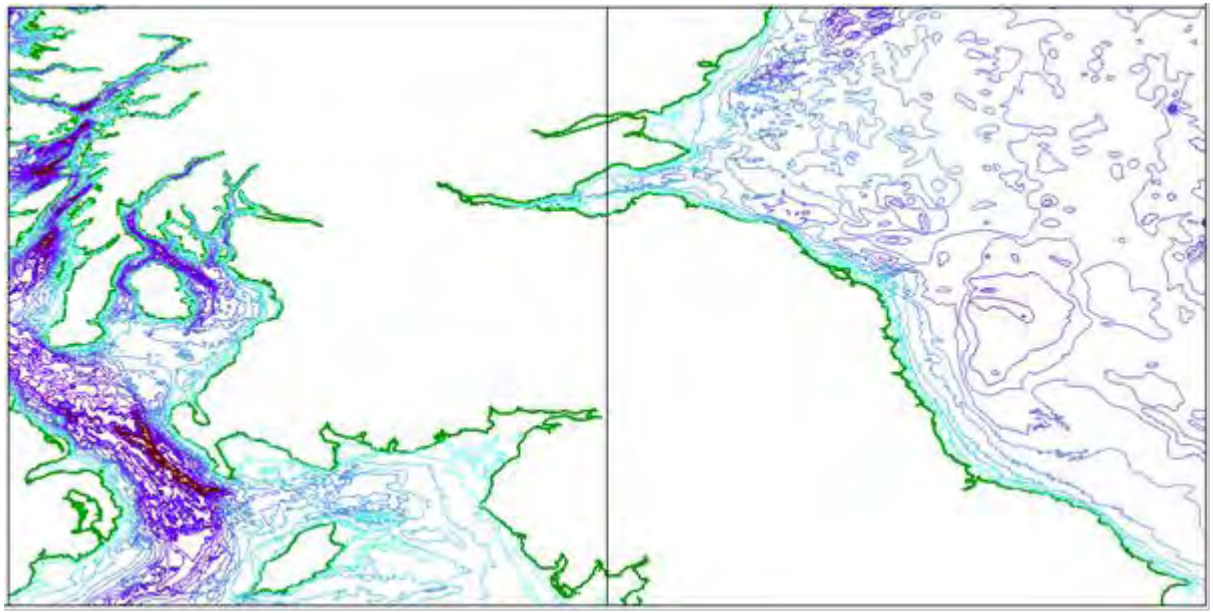
Data provided by TeamSurv
Map generated by Envitia MapLink Studio
TIN generated from grid cell centroids

What is our benchmark?

- Choice of authoritative data sources informed by UKHO and DefStan 00-102 documents
  - ARCS raster charts (navigation)
  - ENC electronic navigation charts (navigation)
  - AML additional military layers (situational awareness)
    - Contour Line Bathymetry (CLB)
    - Large Bottom Objects (LBO)
    - Small Bottom Objects (SBO)
    - Environment Seabed and Beach (ESB)
    - Maritime Foundation and Facilities (MFF)
    - Routes Areas and Limits (RAL)
    - Network Model Bathymetry (NMB) - not actively produced
    - Integrated Water Column (IWC)
    - Atmospheric and Meteorological Climatology (AMC)

Reference Dataset

- AML Contour Line Bathymetry
  - Authoritative (UKHO)
  - Depth Area, Depth Contour*, Sea Area, Sounding



AML Data provided by UKHO
Map generated by Envitia MapLink Studio

Quality Assessment Techniques

- OGC (Open Geospatial Consortium)
  - Generally points to the ISO standards
  - UncertML (a conceptual model to reflect uncertainties in data)
  - Reference to other specifications e.g. INSPIRE
- DefStan 00-102
  - Asserts the use of ISO standards
- ISO Standards
  - ISO 19113 Geographic Information – Quality principles
  - ISO 19114 Geographic Information – Quality evaluation procedures
  - ISO 19138 Geographic information – Data quality measures

Quality Assessment Techniques

- There is no absolute right answer!
- Pick your question!

Standards-based Evaluation

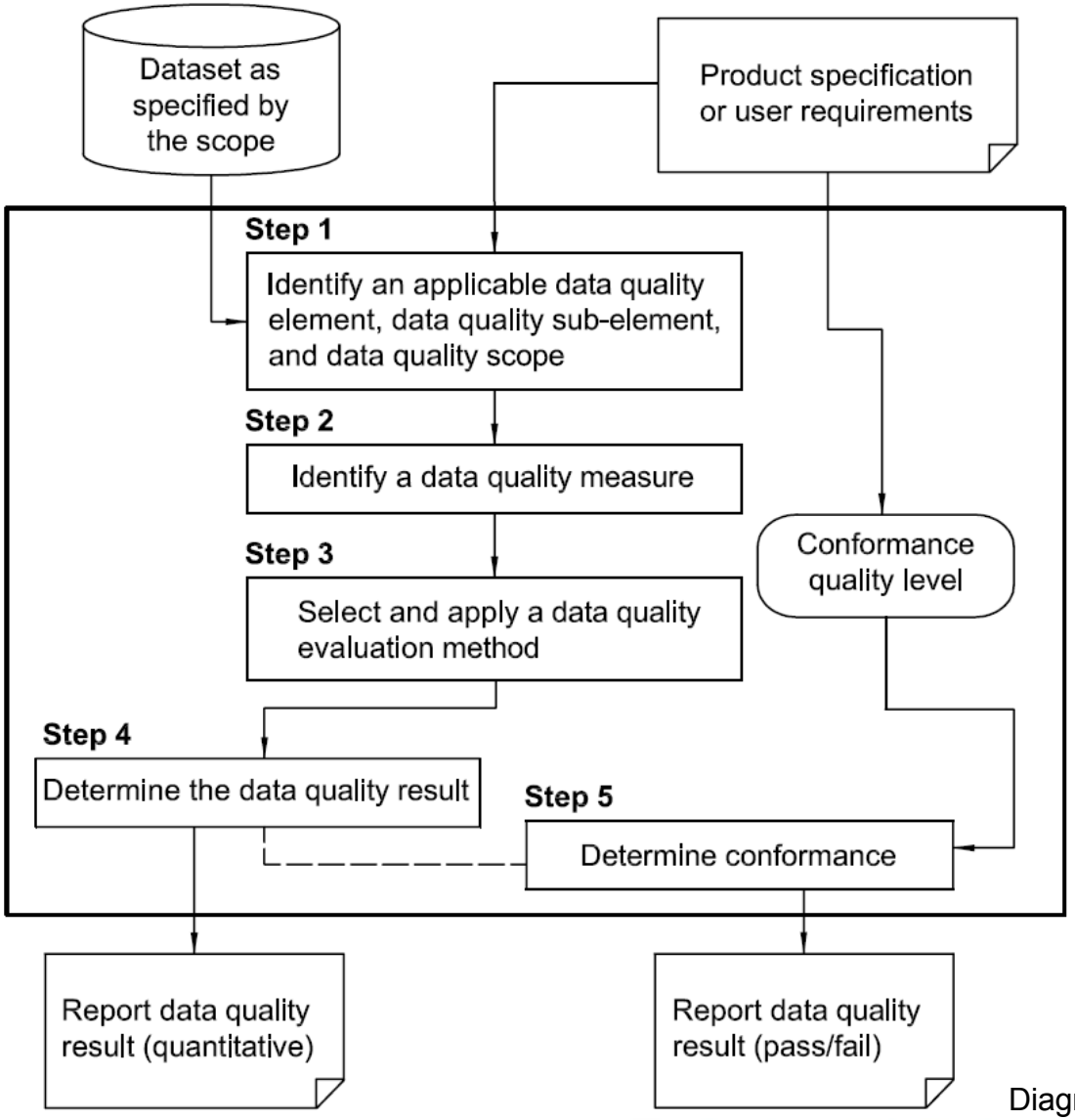


Diagram © ISO 19114 standards document

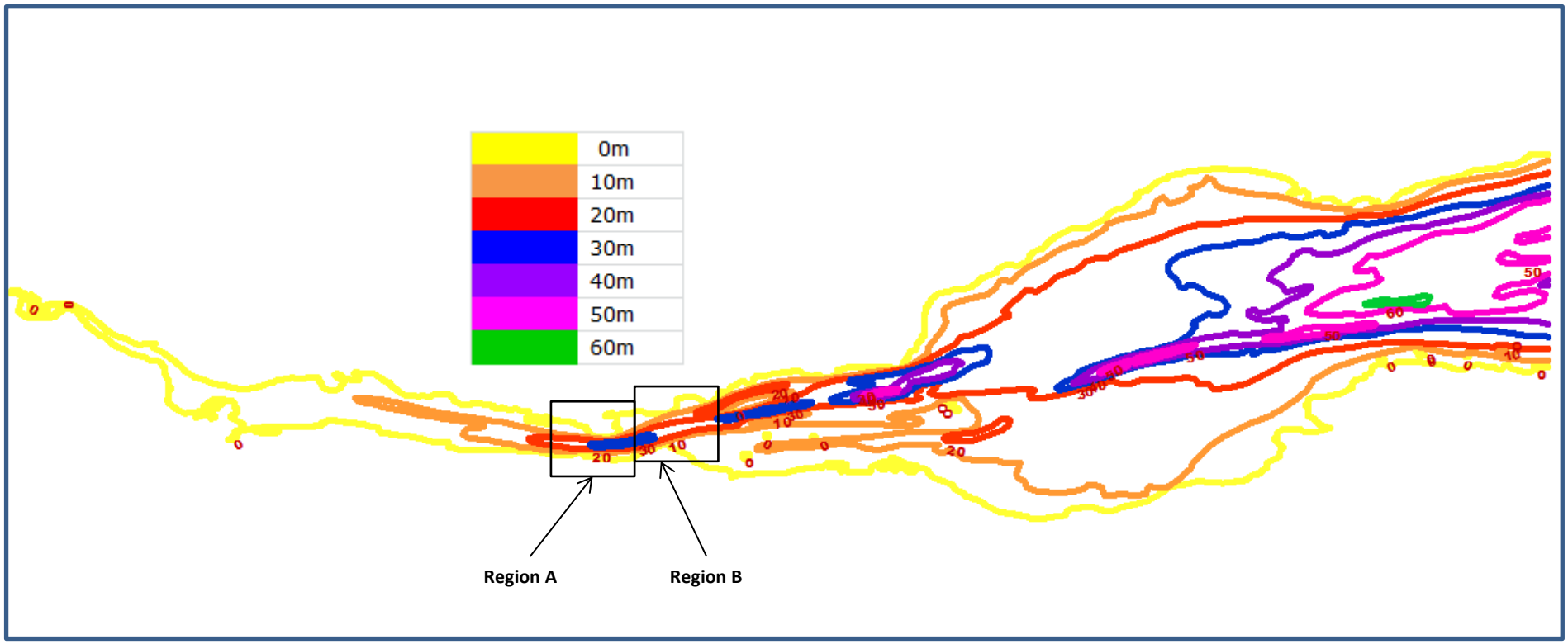
What question?

- Dependent on source data
  - TeamSurv provides location (WGS84,UTM) and depth
- Data quality elements/sub elements that could be measured are;
  - **positional accuracy** (absolute accuracy against a reference set)
  - completeness (omission against a reference set)
  - logical consistency (internal checks that UTM and WG84 locations match accurately)

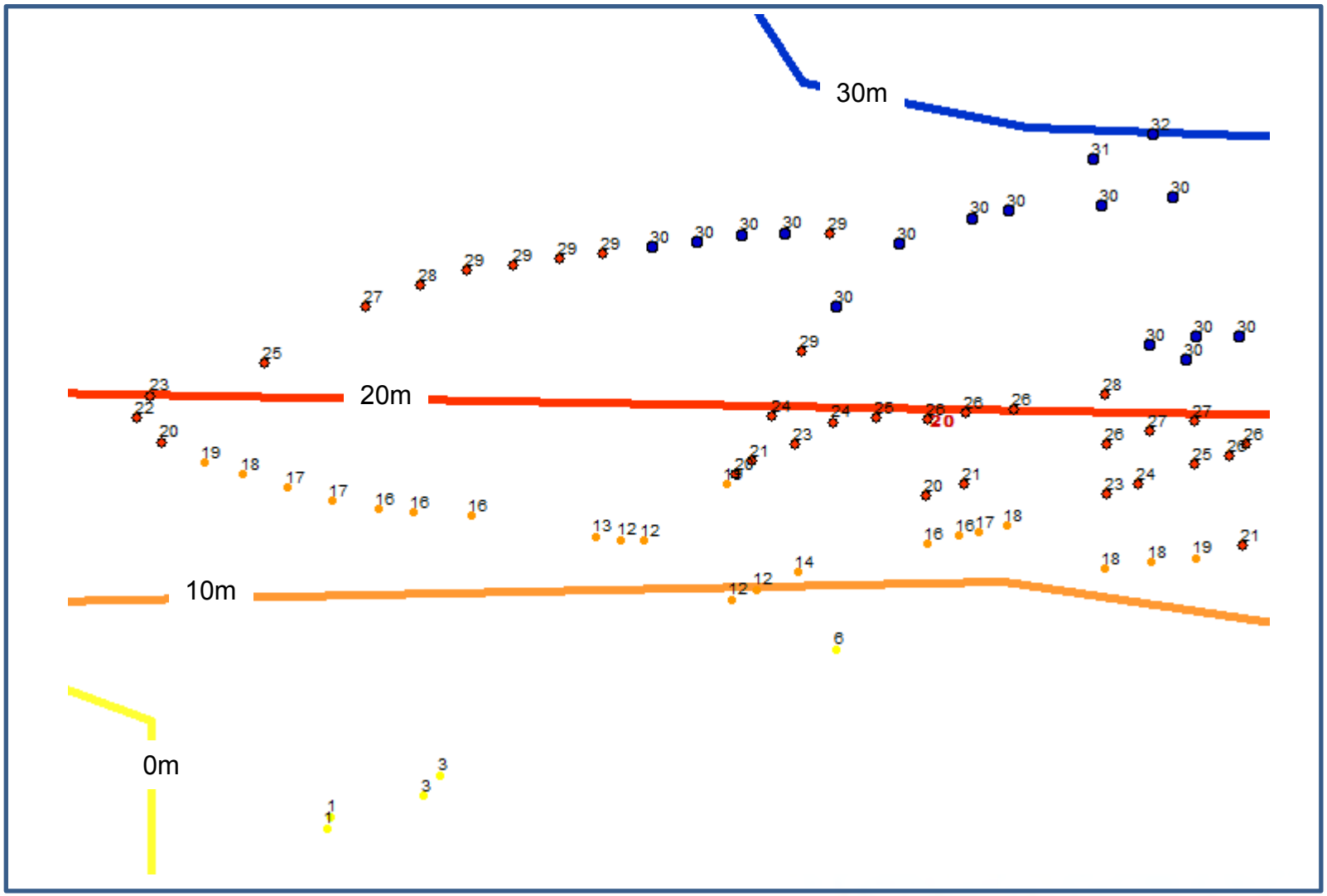
What's MY question

Data Quality Components (Shortname)	
DQ_Name	Depth accuracy
DQ_Scope	Depth values from Firth of Forth
DQ_Element	5-thematic accuracy
DQ_Subelement	3- quantitative attribute accuracy
DQ_Measure	
DQ_MeasureDesc	Percentage of consistent depth values
DQ_MeasureID	
DQ_EvalMethod	
DQ_EvalMethodType	2- Direct external (reference against AML CLB)
DQ_EvalMethodDesc	Consider 360 degree set of rays emanating from the centroid of the depth cell. Record the depth and distance for each intersection with the reference set contour line. Calculate the minimum and maximum depth of the surrounding contours (factored by the distance, closer intersections will have a higher factor.) This will provide the minimum reference depth and maximum reference depth values. Values which are greater than the maximum reference depth or less than the minimum reference depth are considered to be inconsistent. In special case where only one contour is encountered. Remove that contour and recast to determine whether it is a plateau or plain. In which case the depth value must be greater or less than the plateau or plain value respective.
DQ_QualityResult	
DQ_ValueType	4 – Percentage
DQ_Value	XX%
DQ_ValueUnit	Percent
DQ_Date	dd/mm/yyyy
DQ_ConformanceLevel	To be determined e.g. 95% of depth values to be consistent
Example dataset parameters	Omitted.
Example quality result meaning	e.g. Dataset pass. 97% of depth values are consistent with the reference dataset.

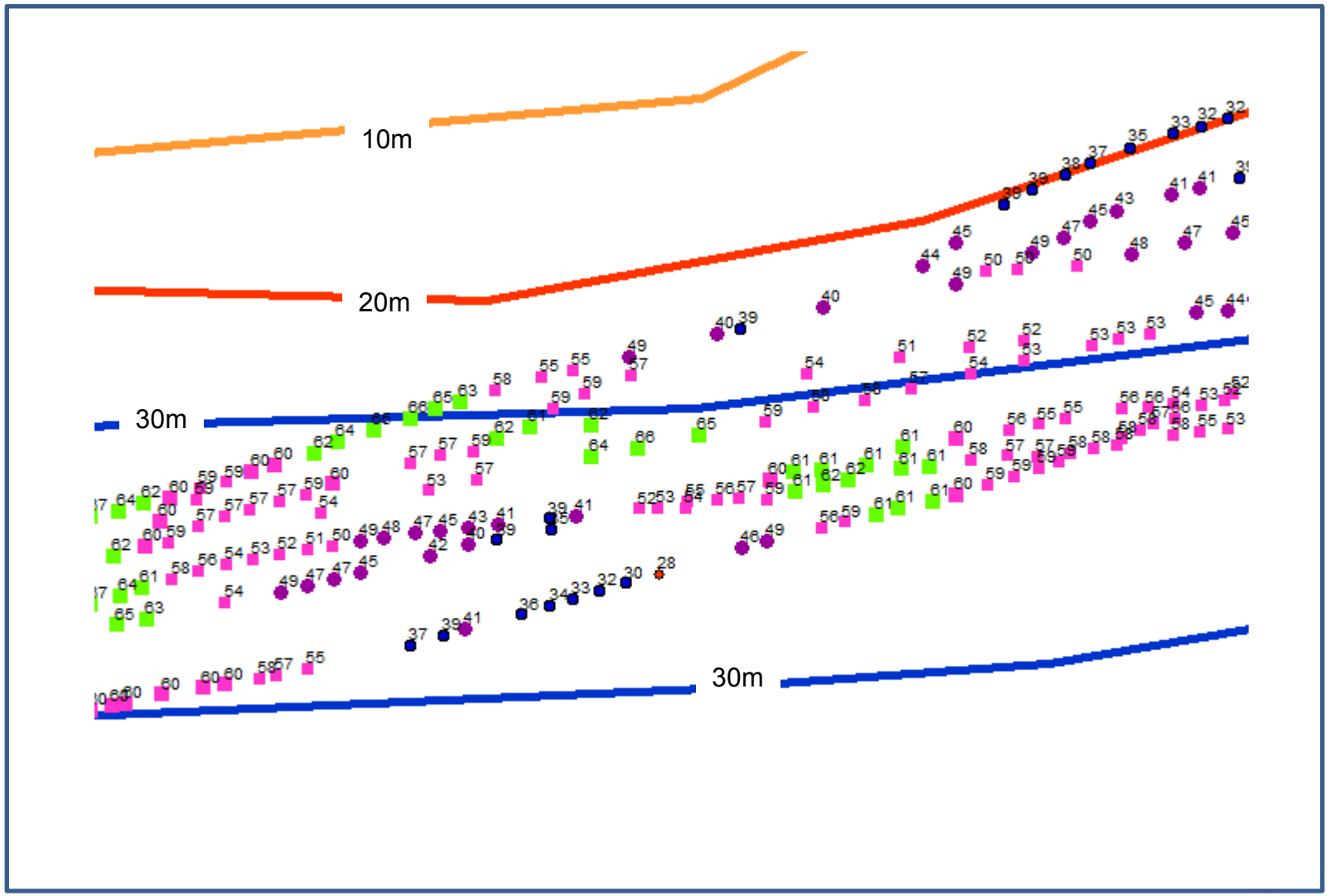
Visual Review - High Contrast AML



Visual Review - Region A



Visual Review - Region B



Validation

- Algorithmic Review
  - Convert data to consistent file formats
  - From cell centre find nearest neighbour contour lines
  - Record Lat, long, cell depth, minimum depth of contour line, maximum depth of contour line
  - Calculated statistical measure

And the answer is

Data Quality Components (Shortname)	
DQ_Name	Depth accuracy
DQ_Scope	Depth values from Firth of Forth
DQ_Element	5-thematic accuracy
DQ_Subelement	3- quantitative attribute accuracy
DQ_Measure	
DQ_MeasureDesc	Percentage of consistent depth values
DQ_MeasureID	
DQ_EvalMethod	
DQ_EvalMethodType	2- Direct external (reference against AML CLB)
DQ_EvalMethodDesc	Consider 360 degree set of rays emanating from the centroid of the depth cell. Record the depth and distance for each intersection with the reference set contour line. Calculate the minimum and maximum depth of the surrounding contours (factored by the distance, closer intersections will have a higher factor.) This will provide the minimum reference depth and maximum reference depth values. Values which are greater than the maximum reference depth or less than the minimum reference depth are considered to be inconsistent. In special case where only one contour is encountered. Remove that contour and recast to determine whether it is a plateau or plain. In which case the depth value must be greater or less than the plateau or plain value respectively.
DQ_QualityResult	
DQ_ValueType	4 – Percentage
DQ_Value	91%
DQ_ValueUnit	Percent
DQ_Date	16/08/2011
DQ_ConformanceLevel	>= 95% of depth values to be consistent with reference data set
Example dataset parameters	Omitted.
Example quality result meaning	Dataset Failed 91% of depth values are consistent with the reference dataset.

91% of depth values are consistent with the reference dataset.

Interim Conclusions

The study shows that

- We cannot trust crowdsourced data blindly
- Crowdsourced data is potentially useful given the right circumstances
- ISO Standards provide a suitable framework for establishing quality measures
- Different features and attributes require different quality measures and evaluation processes
- Algorithmic analysis is key to proving consistent assessment
- One size does not fit all