Appendix 3. Data Confidence Assessment Methods

A review of existing approaches to making confidence assessments of datasets and evidence was conducted. While the GIS literature contains a lot of information on types of error in GIS datasets, surprisingly, only a limited number of examples of confidence assessments were found.

The two that were closest to what we were trying to achieve were those developed by The Crown Estate to support their geographic Marine Resource System (MaRS) and that produced by The Marine Management Organisation (MMO) for reviewing evidence. We are grateful to these two organisations for providing details of their methodologies.

From considering these two we elected to use a simplified version of the MMO's approach, with two versions, one for geographic datasets and one for other evidence. This took the form of Excel spreadsheets where we could assess a dataset / piece of evidence for a number of factors in terms of high, moderate or low confidence. The factors were:

Geographic Datasets
- Is the data of an appropriate scale / resolution?
- Are you aware of when the data was collected and is it suitably up to date?
- Is the data complete for its intended use, suitably uniform?
- Is the data from an authoritative source?
- Any indication of errors?

Other Evidence
- "Source: High - Peer reviewed / scientific journal / Government body, Medium - Professional Report e.g. consultants, Low - Stakeholder"
- Are you aware of when the data was collected and is it suitably up to date?
- Is the data complete for its intended use, suitably uniform and of an appropriate level of detail?
- Does the data use appropriate technology or best practice?
- Any indication of errors?

Comments on these rating were entered where appropriate and an Overall Assessment was calculated based on the above.

The paragraphs below give more detail on assessing each of the factors.

Is the data of an appropriate scale / resolution?

High
- Point datasets with grid references of 100m or better
- Line or polygon datasets digitised from maps with scales better than 1:50,000
- Gridded datasets with resolutions better than 100 metres

Moderate

Appendix 3. Data Confidence Assessment Methods

- Locations that in reality are linear sections or areas that are represented by a single grid reference.
- Line or polygon datasets digitised from maps with scales 1:50,000 – 1:250,000
- Crude shapes used to indicate wide areas of a few kilometres.
- Gridded datasets with resolutions better than 200 metres

Low
- Line or polygon datasets digitised from maps with scales poorer than 1:250,000
- Gridded datasets with resolutions poorer than 200 metres

Are you aware of when the data was collected and is it suitably up to date?

High
- Data is suitably up to date

Moderate
- There may be minor changes to the data since it was collected

Low
- There may be major changes to the data since it was collected

Is the data complete for its intended use, suitably uniform?

High
- The data is fully complete and present for the whole area

Moderate
- The data is partially complete and present for the majority of the area e.g. data from surveys / sampling or collated from multiple but not comprehensive sources

Low
- The data is known to incomplete

Is the data from an authoritative source?

High
- Created from official and/or peer-reviewed sources

Moderate
- Created from unofficial "published" sources – reports, internet etc

Low

Appendix 3. Data Confidence Assessment Methods

- Created by unofficial unpublished sources – fieldwork, personal accounts etc

Any indication of errors?

High
- No indication of errors

Moderate
- Some errors evident – missing / incorrect / additional areas etc

Low
- Significant number of errors – obviously missing or incorrect data, reporting presence of a phenomenon within a whole reporting unit rather than mapping the phenomenon itself.