Marine Data Formats

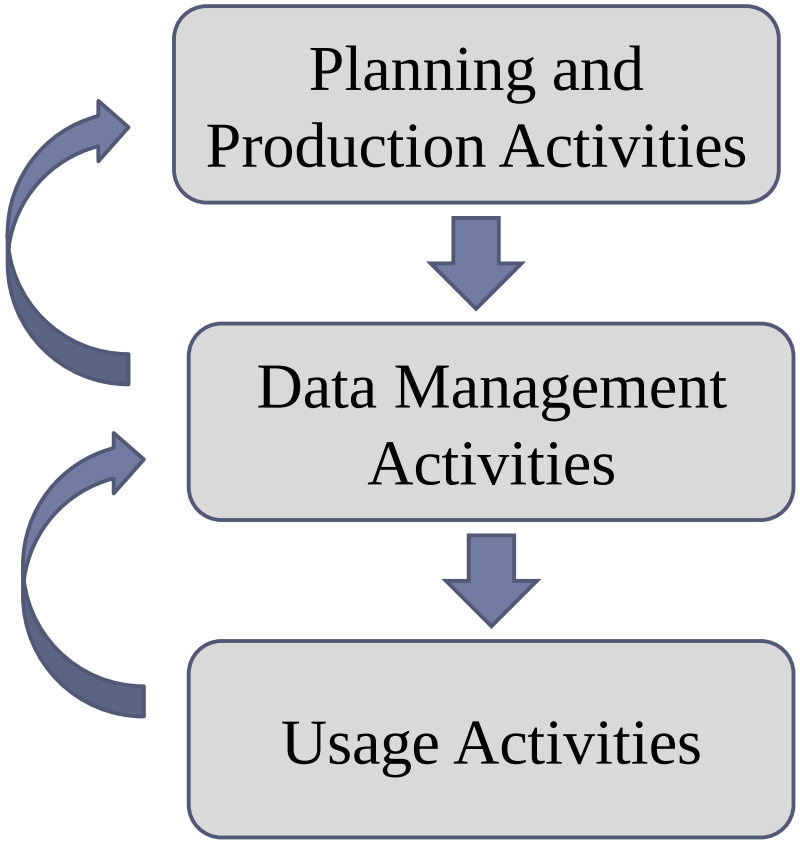
R. Venkat Shesu

Scientist

Indian National Centre for Ocean Information Services (INCOIS)

Hyderabad, India

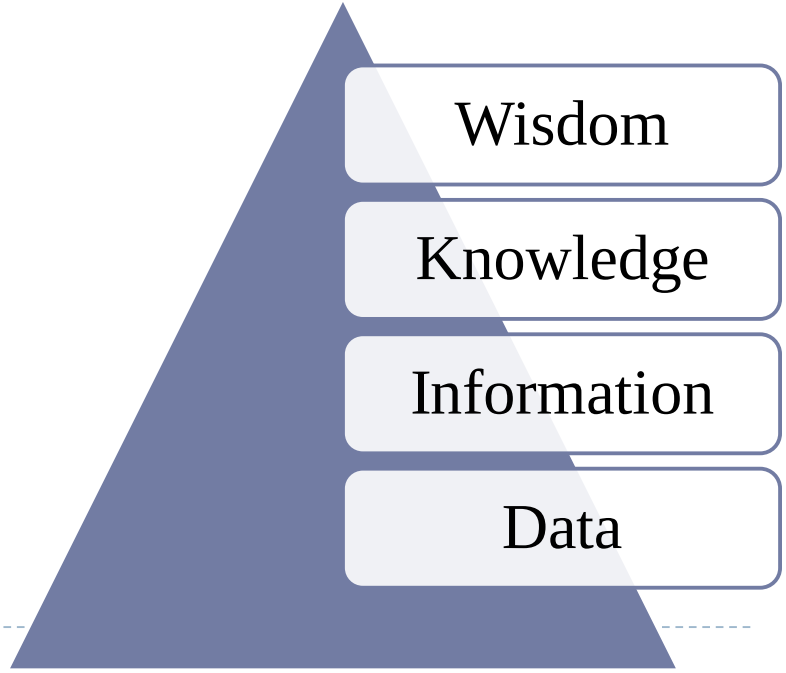
In data life cycle



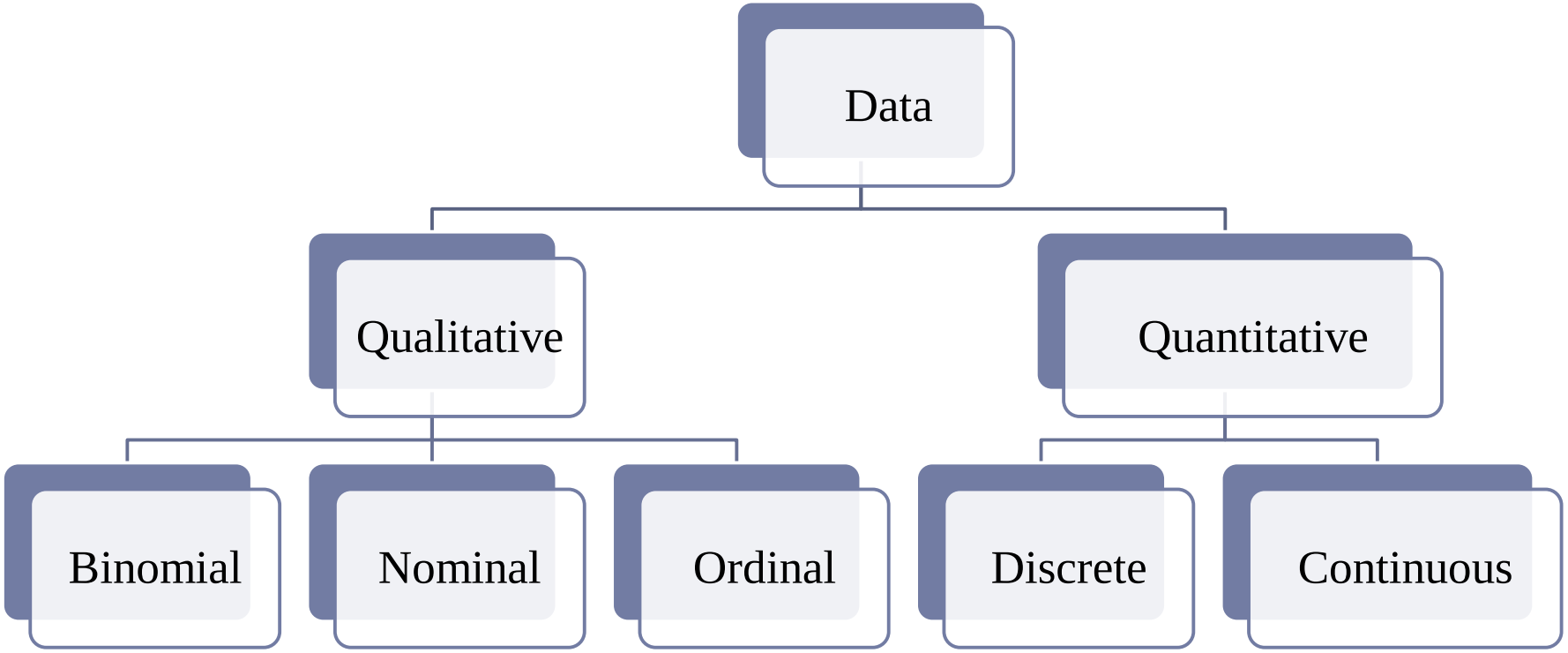
NOAA Environmental Data Management Framework

Data

- Data a collection of numbers or characters
- Lowest level of abstraction
- Qualitative / Quantitative
- It may be (https://data.library.virginia.edu/data-management/plan/format-types/)
  - Observational
  - Experimental
  - Simulated
  - Derived
  - Reference
- Primary / Secondary



Flavours of Data



Data formats

- Structure to organise information and store
- Sharing information
- Research datasets, documents, images , audio/video
- Open / Proprietary
- Choosing for appropriate data format
- Format migration

Choice of Data Format

Dimensions

- Eulerian / Lagrangian
- Univariate / multivariate
- Discrete time / Time series
- Point / Spatial
- Single / Multiple – levels

Closeness

- Accuracy and Precision

Support

- Free / Commercial
- Proprietary / non proprietary
- Communication / archival mechanism
- Software

Volume of Data

- Data growth

OceanTeacher's Data Formats Groups

- Archival
- Auxiliary
- Document
- Hard Copy
- Markup Language
- Message
- Compression
- Metadata
- Raster and Grid
- Relational Databases
- Self-Describing
- Spreadsheet
- Vector

Archival formats

- Similar to spreadsheets, but more than one row format
- E.g. One format for Cruise information, other for station and another for actual data

```
2.P3M5                    1   2 90 51316 01244 N 8341 E      153
   02920    52920   102920   152920   202920   252920   302920   352920   402920
  452920   502880   552830   602810   652800   702780   752770   802730   852690
  902600   952500  1002380  1052290  1102240  1152140  1202080  1252030  1301990
 1351940  1401900  1451850  1501800  1551750  1601710  1651660  1701640  1751600
 1801570  1851530  1901500  1951480  2001460  2051440  2101410  2151400  2201380
 2251340  2301320  2351290  2401280  2451270  2501250  2551240  2601230  2651220
 2701200  2751190  2801180  2851180  2901170  2951160  3001160  3051150  3101150
 3151140  3201140  3251130  3301120  3351120  3401110  3451110  3501110  3551100
 3601100  3651100  3701090  3751080  3801080  3851080  3901070  3951060  4001060
 4051050  4101040  4151040  4201030  4251030  4301030  4351020  4401020  4451020
 4501020  4551010  4601010  4651010  4701010  4751000  4801000  485 990  490 990
 495 980  500 980  505 980  510 980  515 970  520 970  525 970  530 960  535 960
 540 950  545 940  550 940  555 930  560 930  565 930  570 920  575 920  580 910
 585 910  590 910  595 900  600 900  605 900  610 890  615 890  620 890  625 880
 630 880  635 870  640 880  645 880  650 880  655 870  660 870  665 870  670 860
 675 860  680 850  685 850  690 840  695 840  700 840  705 830  710 830  715 820
 720 810  725 810  730 800  735 800  740 790  745 790  750 780  755 780  760 770
```

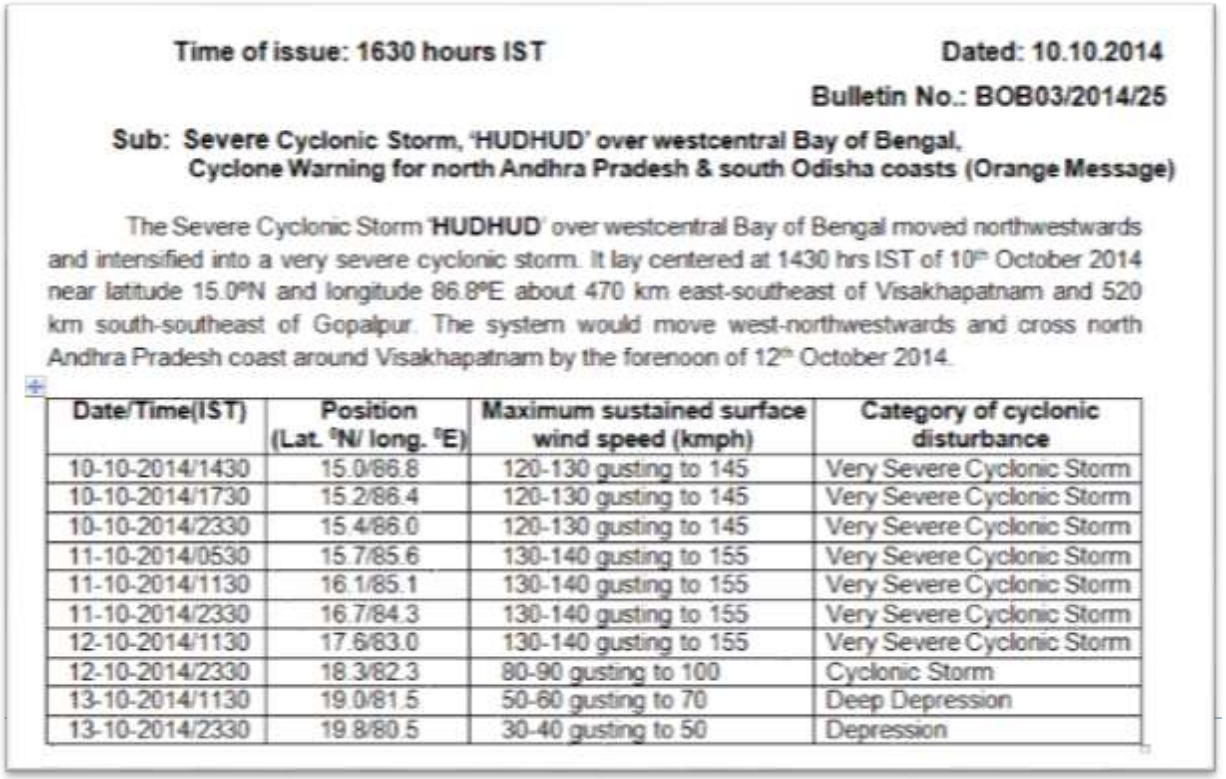
Auxiliary formats

- Containing no data
- To assists programs to find and/or use data files
- Header files

```
ObsTime : 2017-02-28 23:44:55
X : 80.25
Y : 05.25
deltaX : 0.25
deltaY : 0.25
lengthX : 100
lengthY : 67
```

Document formats

- Some data may only be available in document format
- Proprietary or elaborately formatted ASCII text
- Most of these dataset exported to other easily readable formats



Time of issue: 1630 hours IST **Dated: 10.10.2014**

Bulletin No.: BOB03/2014/25

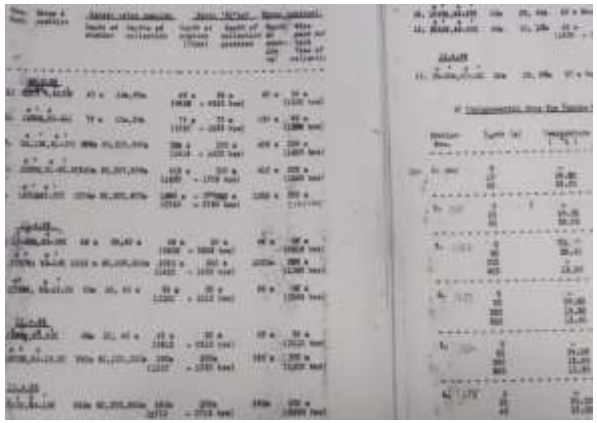
Sub: Severe Cyclonic Storm, 'HUDHUD' over westcentral Bay of Bengal, Cyclone Warning for north Andhra Pradesh & south Odisha coasts (Orange Message)

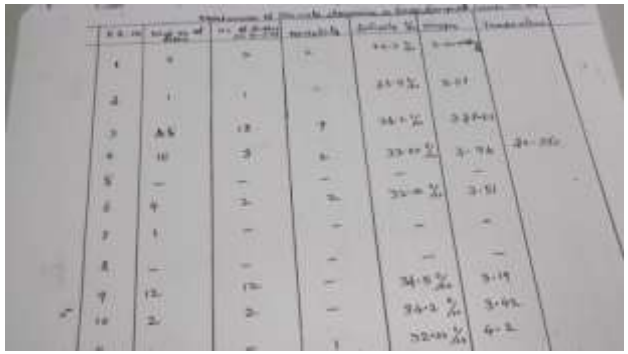
The Severe Cyclonic Storm **'HUDHUD'** over westcentral Bay of Bengal moved northwestwards and intensified into a very severe cyclonic storm. It lay centered at 1430 hrs IST of 10th October 2014 near latitude 15.0°N and longitude 86.8°E about 470 km east-southeast of Visakhapatnam and 520 km south-southeast of Gopalpur. The system would move west-northwestwards and cross north Andhra Pradesh coast around Visakhapatnam by the forenoon of 12th October 2014.

Date/Time(IST)	Position (Lat. °N/ long. °E)	Maximum sustained surface wind speed (kmph)	Category of cyclonic disturbance
10-10-2014/1430	15.0/86.8	120-130 gusting to 145	Very Severe Cyclonic Storm
10-10-2014/1730	15.2/86.4	120-130 gusting to 145	Very Severe Cyclonic Storm
10-10-2014/2330	15.4/86.0	120-130 gusting to 145	Very Severe Cyclonic Storm
11-10-2014/0530	15.7/85.6	130-140 gusting to 155	Very Severe Cyclonic Storm
11-10-2014/1130	16.1/85.1	130-140 gusting to 155	Very Severe Cyclonic Storm
11-10-2014/2330	16.7/84.3	130-140 gusting to 155	Very Severe Cyclonic Storm
12-10-2014/1130	17.6/83.0	130-140 gusting to 155	Very Severe Cyclonic Storm
12-10-2014/2330	18.3/82.3	80-90 gusting to 100	Cyclonic Storm
13-10-2014/1130	19.0/81.5	50-60 gusting to 70	Deep Depression
13-10-2014/2330	19.8/80.5	30-40 gusting to 50	Depression

Hard copy format

- Data on paper, not digitized
- Publications, journals, books, gray literature
- Data rescue initiatives
  - IODE's Global Oceanographic Data Archaeology and Rescue (GODAR)





Markup language formats

- Instructions, how to display text
- Computer typesetting
- Word-processing systems
- Hypertext Markup Language (HTML)
- Extensible Markup Language (XML)
- Keyhole Markup Language (KML)

```
<observations>
        <cruise_id>   ..... </cruise_id>
        <station_id> ..... </station_id>
        <temperature> ..... </temperature>
        <salinity> ..... </salinity>
</observations>
```

Message formats

- WMO's Global Telecommunication System (GTS)
- Formatted codes
- Buoy Format – FM-18

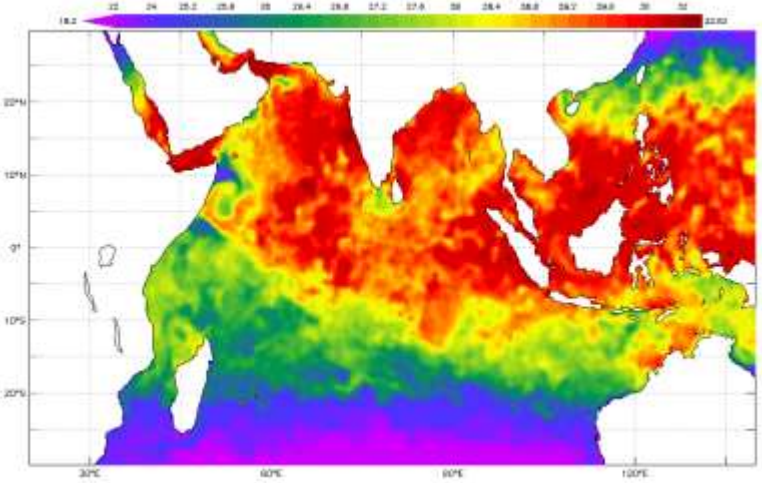
```
SSVX01 DEMS 151240
ZZYY 23459 15027 12001 113988 086941 111// 00405 10/// 40158 222// 00276 1////=
 ZZYY 23451 15027 12001 114970 069017 111// 0//05 10275 40145 222// 00278 10502=
 ZZYY 23452 15027 12001 111765 068623 111// 00206 10276 40136 222// 00283 1////=
ZZYY 23453 15027 12001 108267 073343 111// 00203 10293 40128 222// 00292 10703=
```

Metadata formats

- Data about the data
- Describing the datasets
- May or may not be interoperable
- Standards
  - ISO 19115 (Geographic information – metadata)
  - Directory interchange format (DIF)
  - Cruise summary report (CSR)
  - Federal Geographic Data Committee (FGDC)
  - Darwin Core

Raster and Grid formats

- Gridded data
- Rectangular grid of pixels
- Remote sensing data
- GeoTiff
- ArcGIS Grid
- Image files with varying formats



Relational databases

- Tabular structures
- Universally binary, invisible to users
- Proprietary formats
- SQL quering, ODBC and export utilities
- Important databases – Oracle, DB2, PostgreSQL, MySQL, MS-Access .....
- Ocean Data View (*.var, *.odv)

Self describing formats

- Extensive internal metadata
- GRIB, BUFR – Meteorological
- HDF – Satellite data
- netCDF – Oceanographic observations
- Best for gridded and raster data
- Data discovery and practical usage

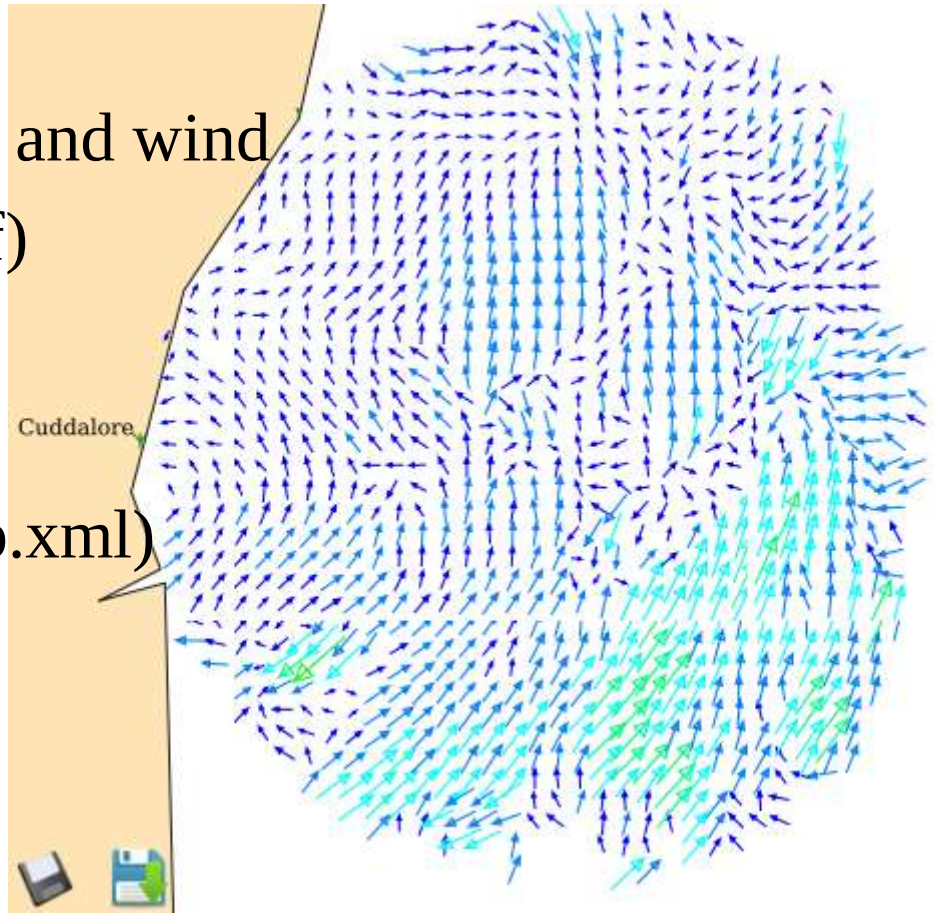
Spreadsheet format

- Data tables
- Rows and columns
- Header as ancillary data
- TSV, CSV

```
Cruise Station Type mon/day/yr hh:mm Longitude Latitude Bot. Depth Depth Temperature
WOD05_06000082 506503 B 06/03/1967 10:00:00.000 13.98 -26.65 400 5 17.43
WOD05_06000082 506503 B 06/03/1967 10:00:00.000 13.98 -26.65 400 -99999 -99999
WOD05_06000082 506591 B 06/04/1967 08:00:00.000 14.07 -26.68 400 5 16.8
WOD05_06000082 506591 B 06/04/1967 08:00:00.000 14.07 -26.68 400 -99999 -99999
WOD05_06000082 506701 B 06/05/1967 08:00:00.000 14.1 -26.77 390 5 15.92
WOD05_06000082 506701 B 06/05/1967 08:00:00.000 14.1 -26.77 390 -99999 -99999
WOD05_06000082 506829 B 06/06/1967 11:00:00.000 13.97 -26.5 385 5 16.38
```

Vector Formats

- Lines, points, curves
- Shapes / polygon
- Vector data such as current and wind
- Shape files (.shp, .shx, .dbf)
- Project information (.prj)
- Spatial index of feature
- Metadata information (.shp.xml)



Compression format

- To reduce the volume
- Efficient use of storage
- Packaging of data
  - Zip
  - Gzip
  - Bzip
  - KMZ
  - Tar
  - E00

List of common data formats

ASC	**DOC, DOCX**	**HDF**	**NCML**	**SHP**	**TFW**
BMP	**DXF**	**JPG, JPEG**	**NCWMS**	**SHX**	**VRT**
BPW	**E00**	**JGW**	**ODV**	**SPRJ**	**World File**
BSB	**FLT**	**KAP**	**PGW**	**SPRM**	**XLS, XLSX**
BUFR	**GIF**	**KML**	**PNG**	**S-57**	**XML**
CDL	**GFW**	**KMZ**	**PPT, PPTX, PPS, PPSX**	**TAB**	**XYZ**
CNV	**GINI**	**MGD77**	**PRJ**	**TAR**	**ZIP**
CSV	**GML**	**MGRD**	**RAR**	**TSV**	
DAT	**GRIB, GRIB2**	**NC, NetCDF**	**SDAT**	**TXT**	
DBF	**GZ**	**NCD**	**SGRD**	**TIF, TIFF**	

?

Thank you