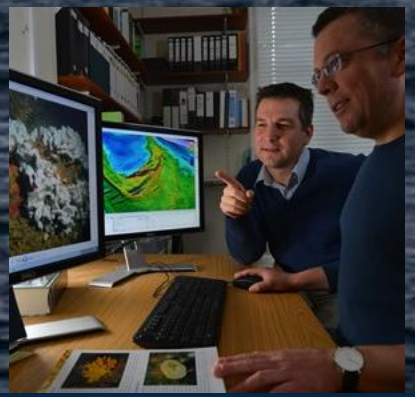


Centre for Environment
Fisheries & Aquaculture
Science

Cefas

Our vision is to make a real difference for society as recognised leaders in marine and aquatic science.

"Cefas is a world leader in marine science and technology, providing innovative solutions for the aquatic environment, biodiversity and food security"

Knowledge is Power

Not everything that counts can be counted

and

not everything that can be counted

counts

Dr David Morris

29th September 2015 - ECOWinds Conference

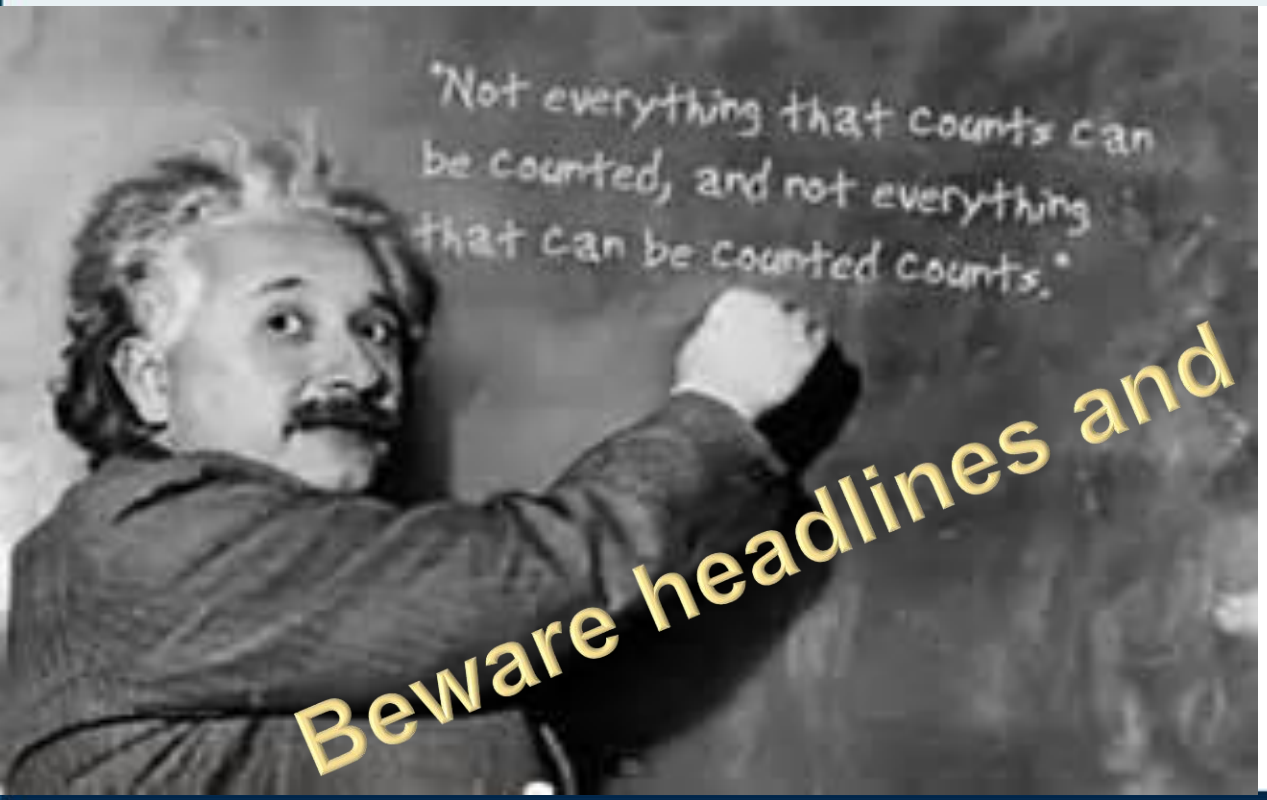
Lowestoft

Centre for Environment
Fisheries & Aquaculture
Science

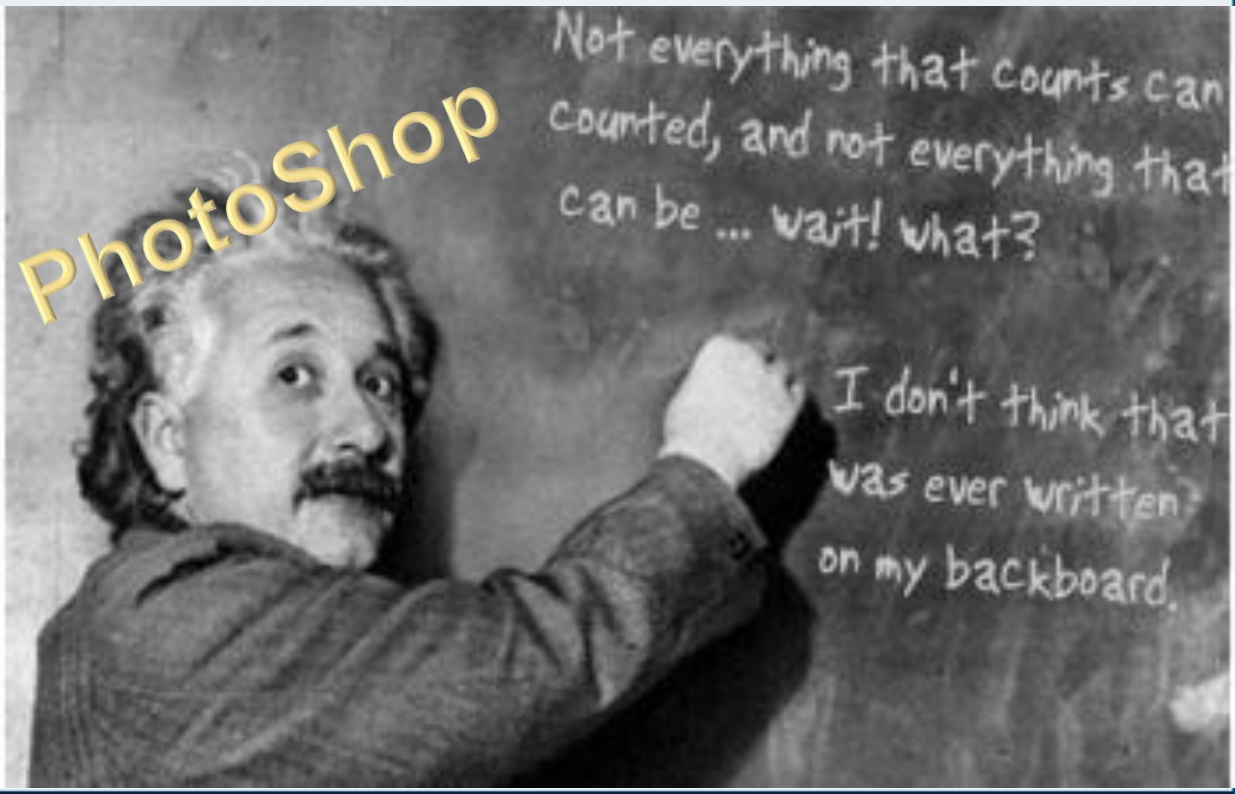
Cefas

Report | McKinsey Global Institute

Big data: The next frontier for innovation, competition, and productivity



Beware headlines and



BIG DATA- Cefas doesn't have any!

- BIG DATA is usually defined by a number of '**V**'s
- I've seen requirements for how many ranging from 3 to 7
- Usually 3, 4 or 5
- Often in different orders
- Usually with claims that all are needed

VOLUME, VELOCITY, VARIETY

- **VOLUME** - scale of data (size or rate of production)
- **VELOCITY** - analysis of streaming data - speed of generation (see VOLUME above - they really are all mixed up – both the data and the people talking about it)
- **VARIETY** - different types of data - structured and unstructured (20:80 ratio often quoted here)

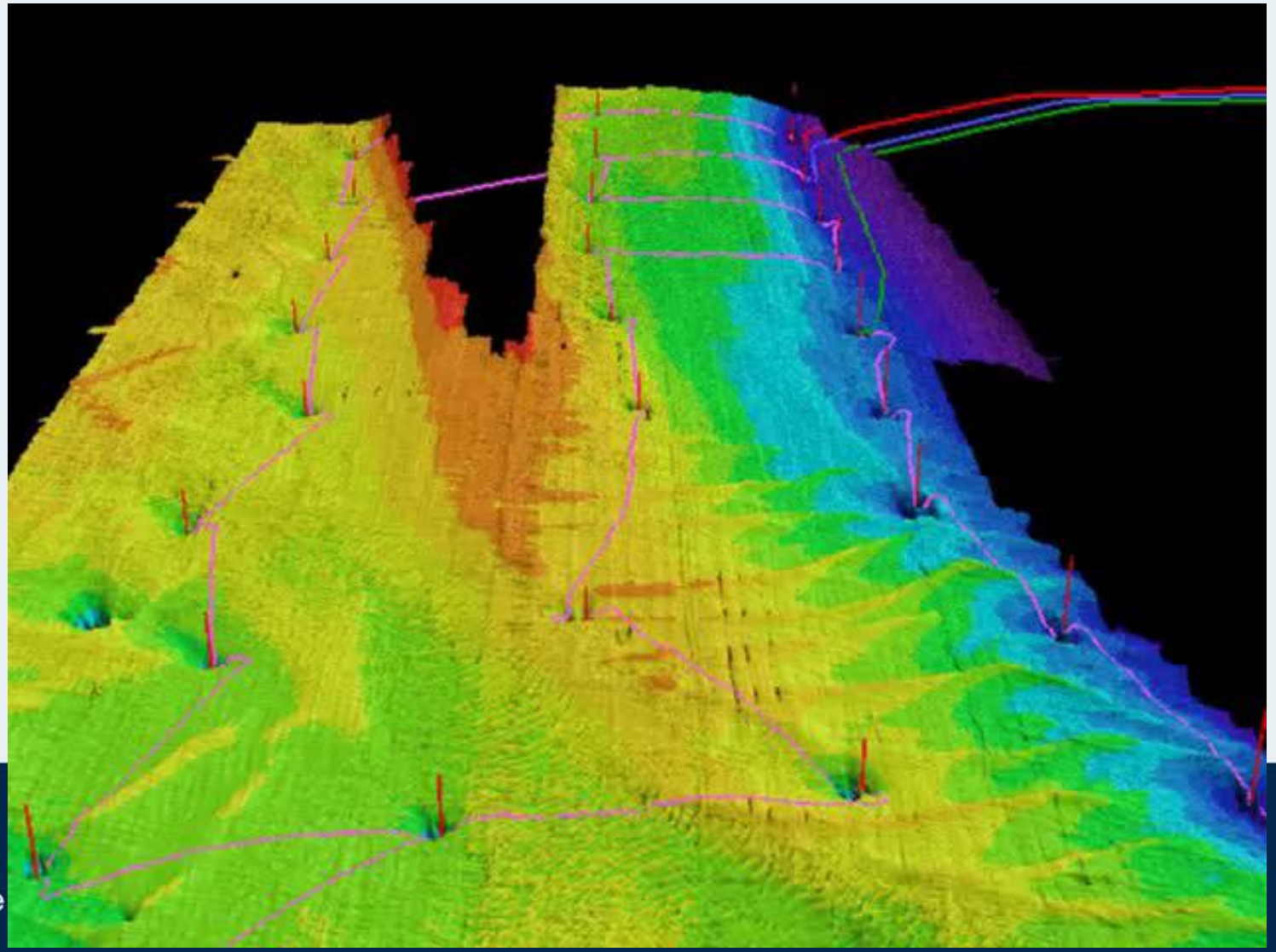
VERACITY, VALUE, VARIABILITY

- **VERACITY** - uncertainty of data (in terms of its value and usefulness for YOUR problem)
- **VALUE** - no point unless we can get this - most of the Twitter stuff has no intrinsic value for me (not really) but it may have value to someone else who can analyse it with, say, location data.
- **VARIABILITY** - rapidly changing “meaning” - to whoever and others

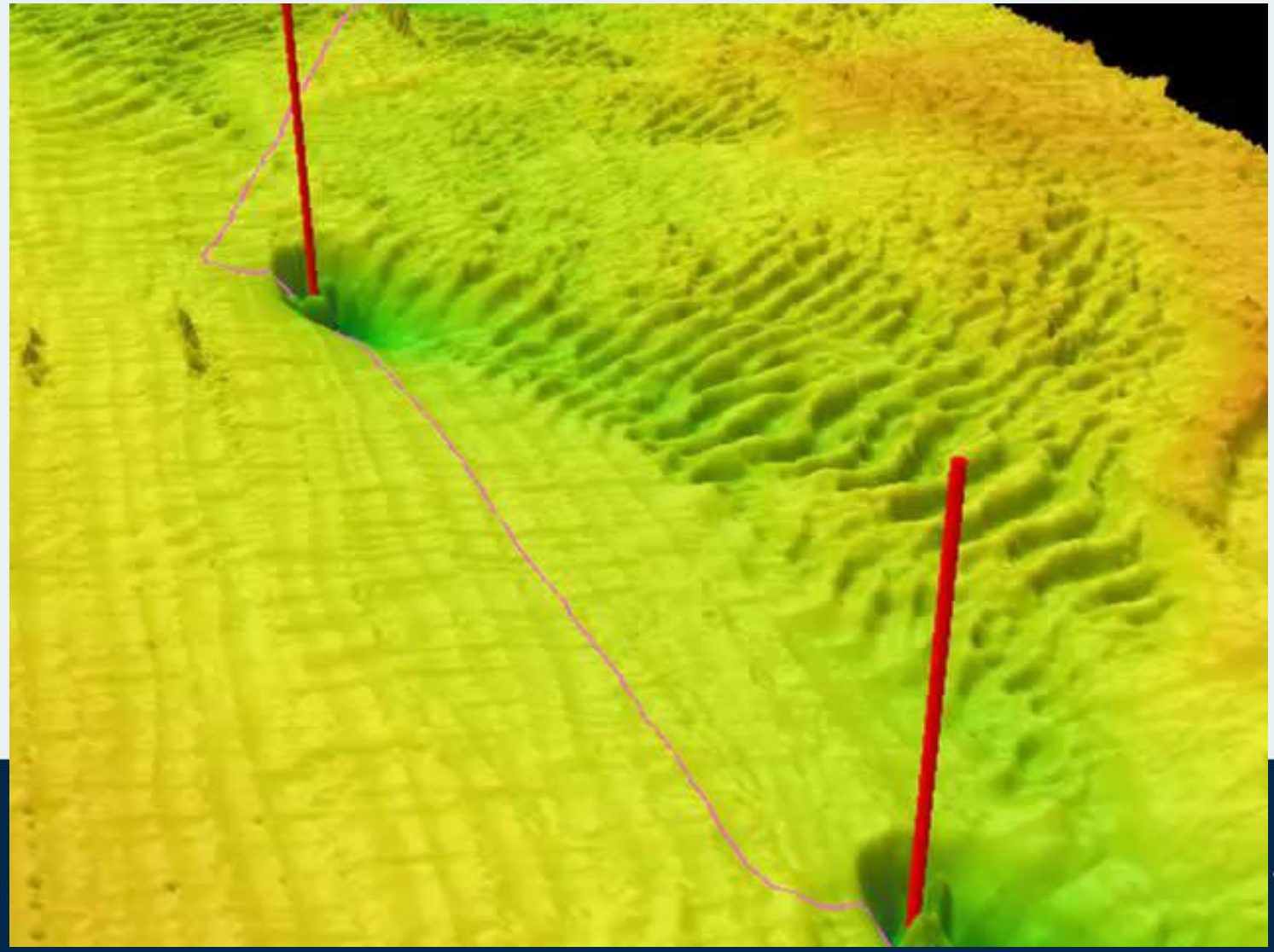
VISUALISATION

- **VISUALISATION** - so mere mortals can understand it and use it

Now VISUALISATION we CAN do!
If it's a clear day look North

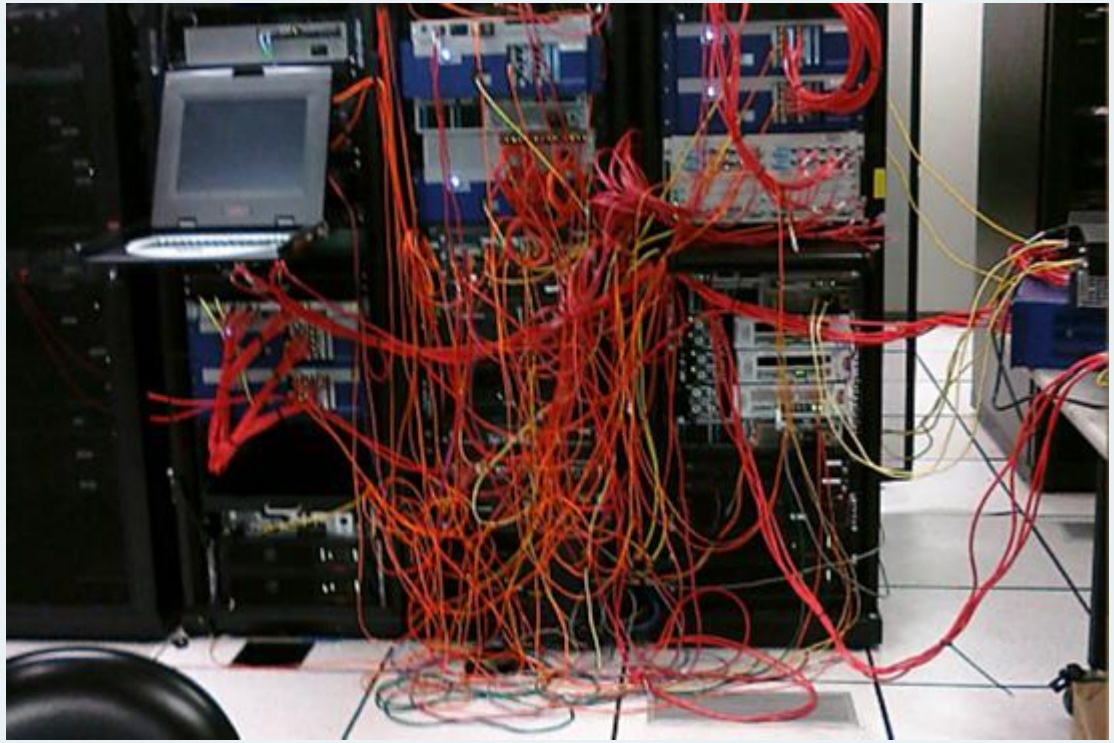


and in a bit more detail (and of direct interest)



So what else data related DOES Cefas have?

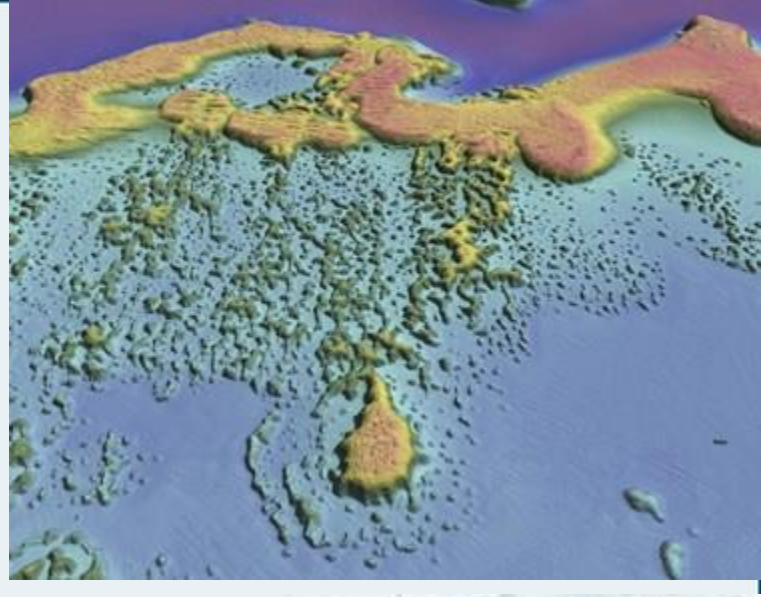
Lots of Legacy Data



Lots of Instruments and connections

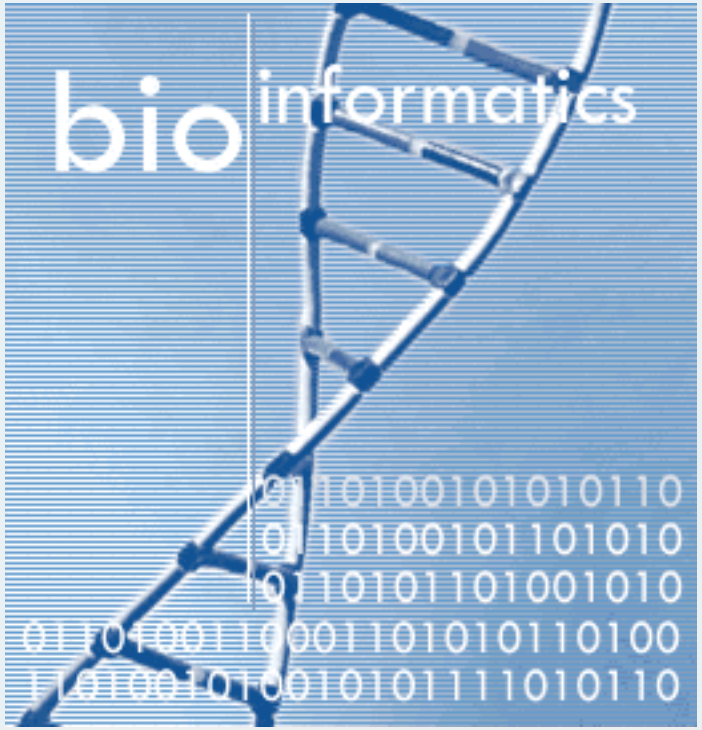
Lots of buoys

- **Sensors**
- **Samples**
- **Analyses**
- **Satellite data**
- **Remote Piloted Vehicle images**
- **Industry Data**



The usual kit, some unusual kit, models (50 y hind cast for N Sea), more legacy data (cold).
And customers everywhere (and an Office in Kuwait)

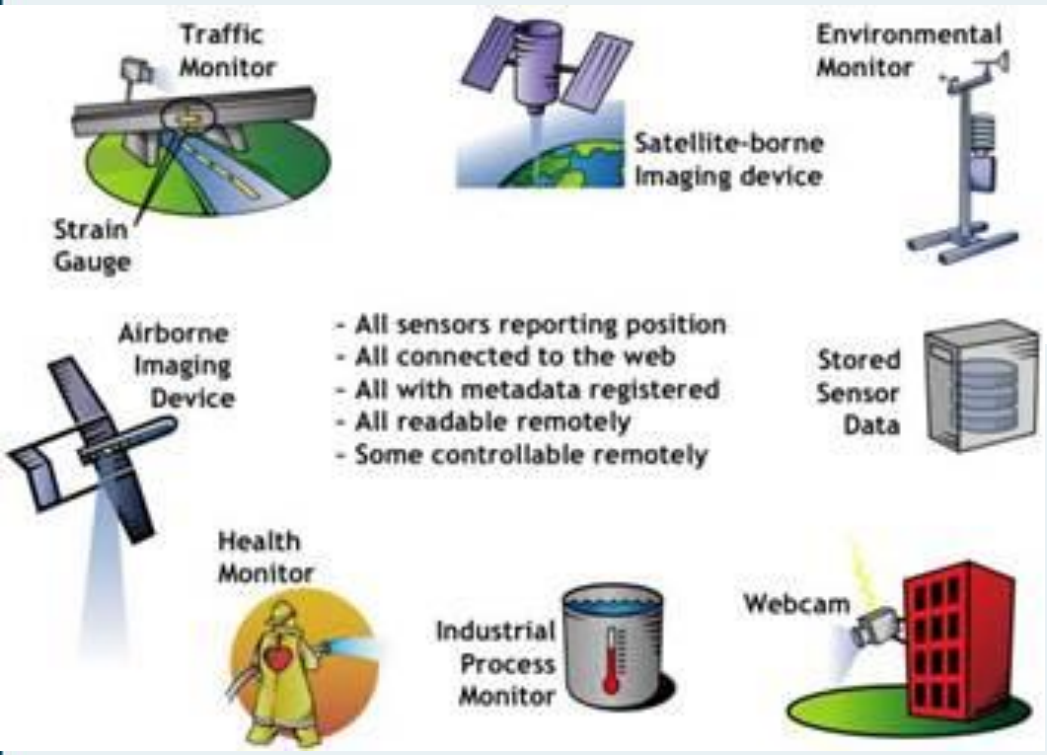
eDNA



Oh and ~50,000 otoliths a year.

And new stuff like Wave Gliders and e-DNA from water samples (Topmouth Gudgeon) – lots of 1s and 0s

And where are we going with it?



OGC®

Making location count.

www.opengeospatial.org

Where we will bump into Big Data and Big Data Analytics

From sensors to you AND the Hub

Cefas Data Hub

Cefas holds a wealth of data related to the marine environment and is working to make much of it directly available to the public. The Cefas Data Hub provides a central point from which information about our data can be discovered.

https://www.cefas.co.uk/publications-data/cefas-data-hub/

AGAIN - Cefas has tens of Tbytes of Big Data but not BIG DATA

- Nearly ALL is structured (not 20%) – collected by design BUT being deconstructed for re-use elsewhere
- Big by most standards: we think that "*if its hard to move around it's big*" and "*if its £10,000+ a software seat its big money*"
- We have significant **Volumes** that come in at reasonable **Velocities** from all sorts of sensors (from people to ships) so it has **Variety** and we work hard on **Veracity** by design and we get paid to collect it, so it has clear **Value,** and anything biological is highly **Variable**. And, as you have seen, it can sometimes need a lot of **Visualisation** effort.

7 out of 7! And STILL not BIG DATA!

Top 10 categories for Big Data sources and mining technologies

Getting over the gee-whiz factor of Big Data can be tough. Enumerating important Big Data sources and technologies can give us a good start in moving the discussion forward.

http://www.zdnet.com/article/top-10-categories-for-big-data-sources-and-mining-technologies/

Only 2/10 at a push – so Big Data it is!

- Social network profiles
- Social influencers
- **Activity-generated data**
  (but not all the location and mobile phone and home thermostat type stuff)
- Public
- Software as a Service (SaaS) and cloud applications
- Hadoop MapReduce application results
- Data warehouse appliances
- Columnar/NoSQL data sources
- Network and in-stream monitoring technologies
- **Legacy documents**

The value to you?

- Publicly funded data is/will be available “free”
- Under Open Government Licence

- Knowing what to do with the data, that’s Knowledge
- And “Knowledge is Power”
- Neither come “free”

REMEMBER (BIG DATA)

not everything that counts can be counted and

not everything that can be counted counts

CEFAS

knowing what counts, when, where and how to count it, turning counts into knowledge and knowing what then really counts, and what to do with that

Contact

david.morris@cefas.co.uk

http://www.cefas.co.uk

@CefasGovUK

www.linkedin.com/company/cefas