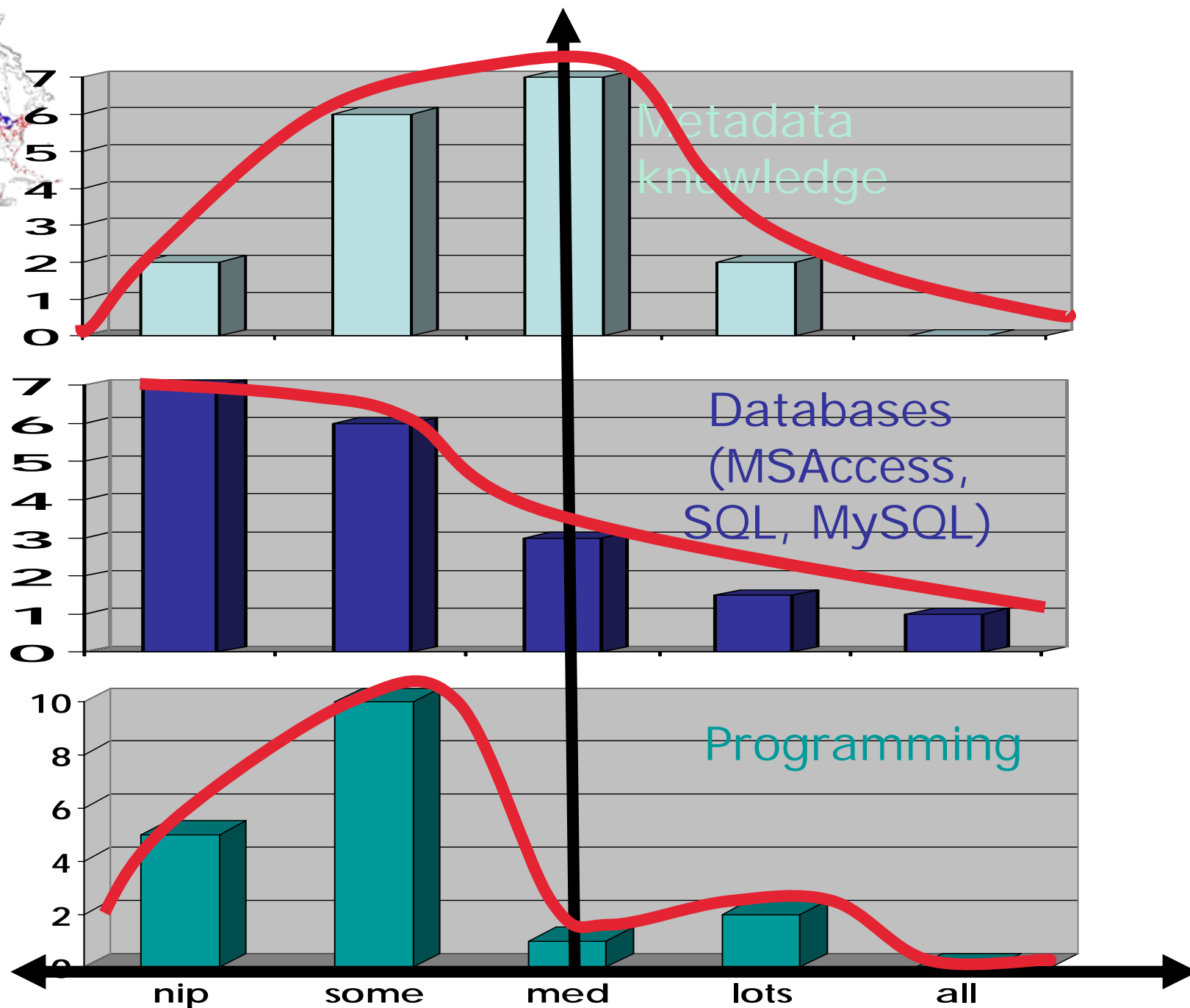
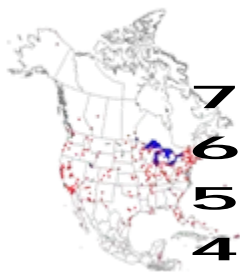




Legacy Metadata

Iñigo San Gil
Metadata assistant
LTER Network Office
Albuquerque, Nuevo México
Oct 31 2005







What is legacy metadata?

Chances are you are handling data from some experiment or simulation. That's legacy data.

Documenting the data **now** will save you a lot of trouble down the road.

You will need the answers to these questions:

What was the data about?

What format did it have?

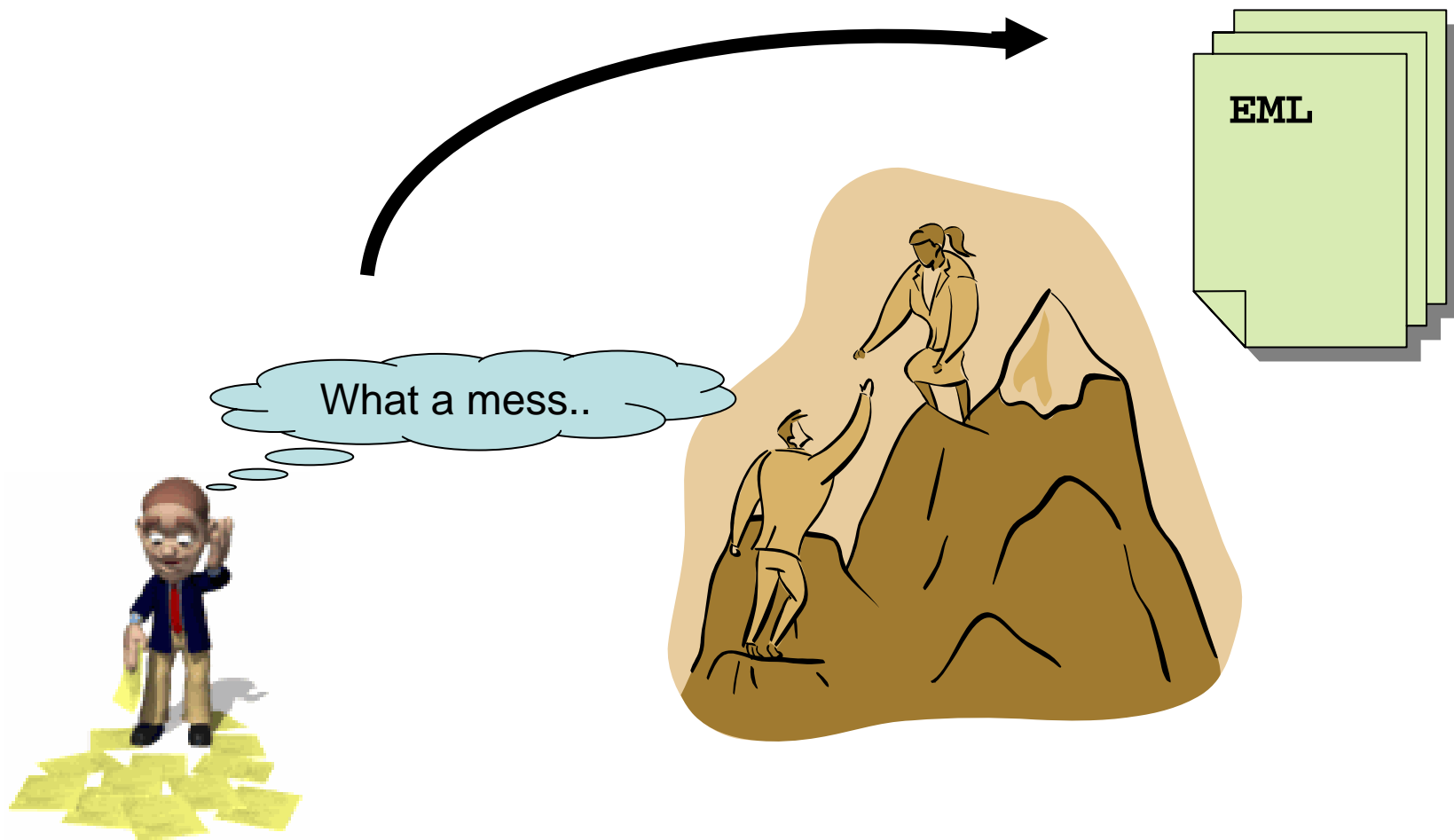
Where is it stored?

Who helped in the data collection? How?





Goal: standardize metadata

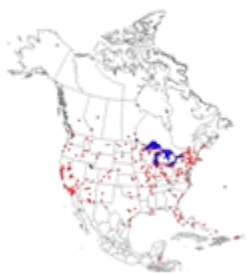




Outline -- Legacy Metadata

- How much, how rich and how is it stored?
- One site, one metadata management model (26 LTER sites, 26 Kingdoms.)
- Databases: MS Access, SQLs, Paradox...
- Flat text files: Structured Folders (or not)
- Case: transforming legacy metadata





How much metadata?

How wide and deep is your metadata
do we have it in a database?
do we have flat text files?

XML editor (Oxygen, XMLSpy, others)



Perl (Practical Extraction and Report Language)

Other tools (will talk about the possibilities later)





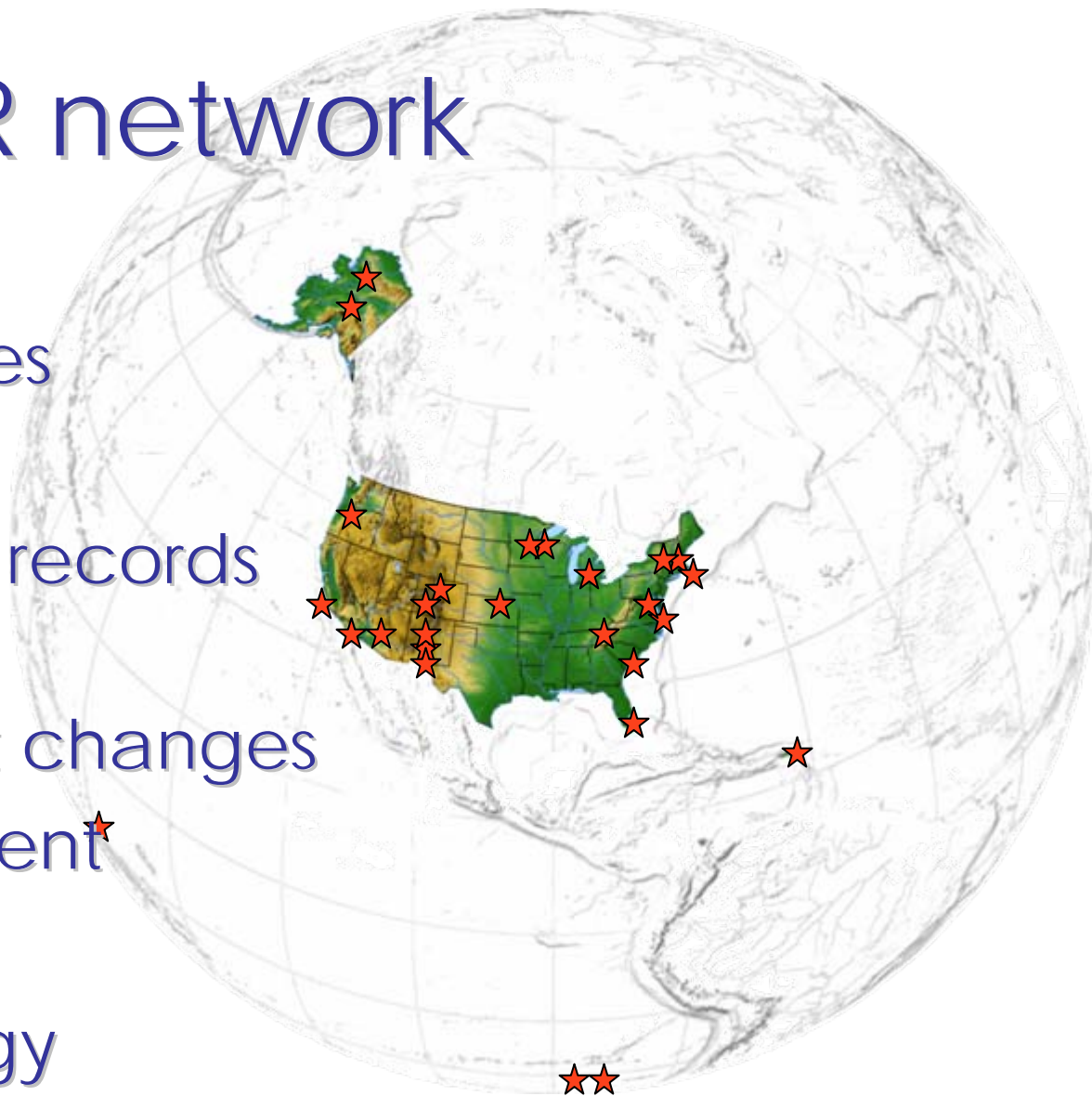
The LTER network

26 autonomous sites

20 yr+ length data records

Relatively frequent changes
in data management

Evolving technology





Data management models

Databases:

MySQL, MSAccess, Paradox, Metacat, etc



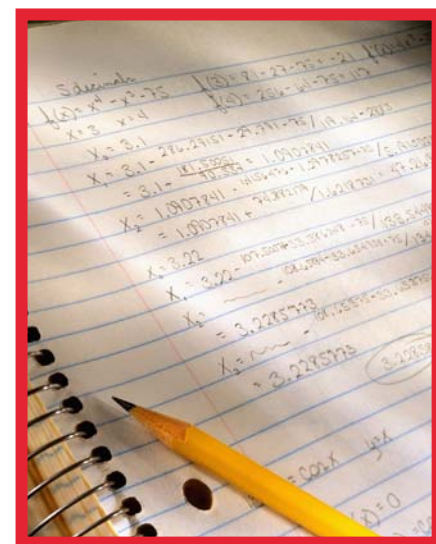
Spreadsheets:

Excel, etc



Text files:

Flat ASCII, Word documents.



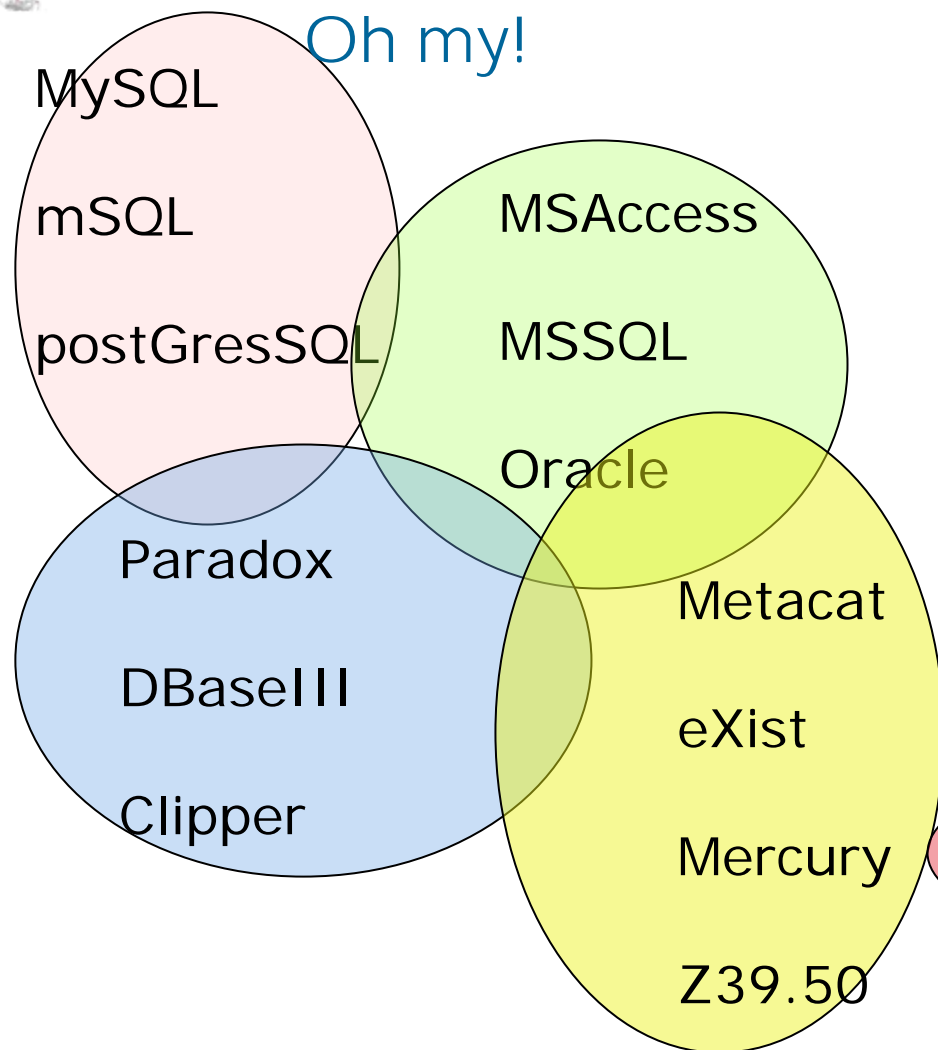
Combos of the above





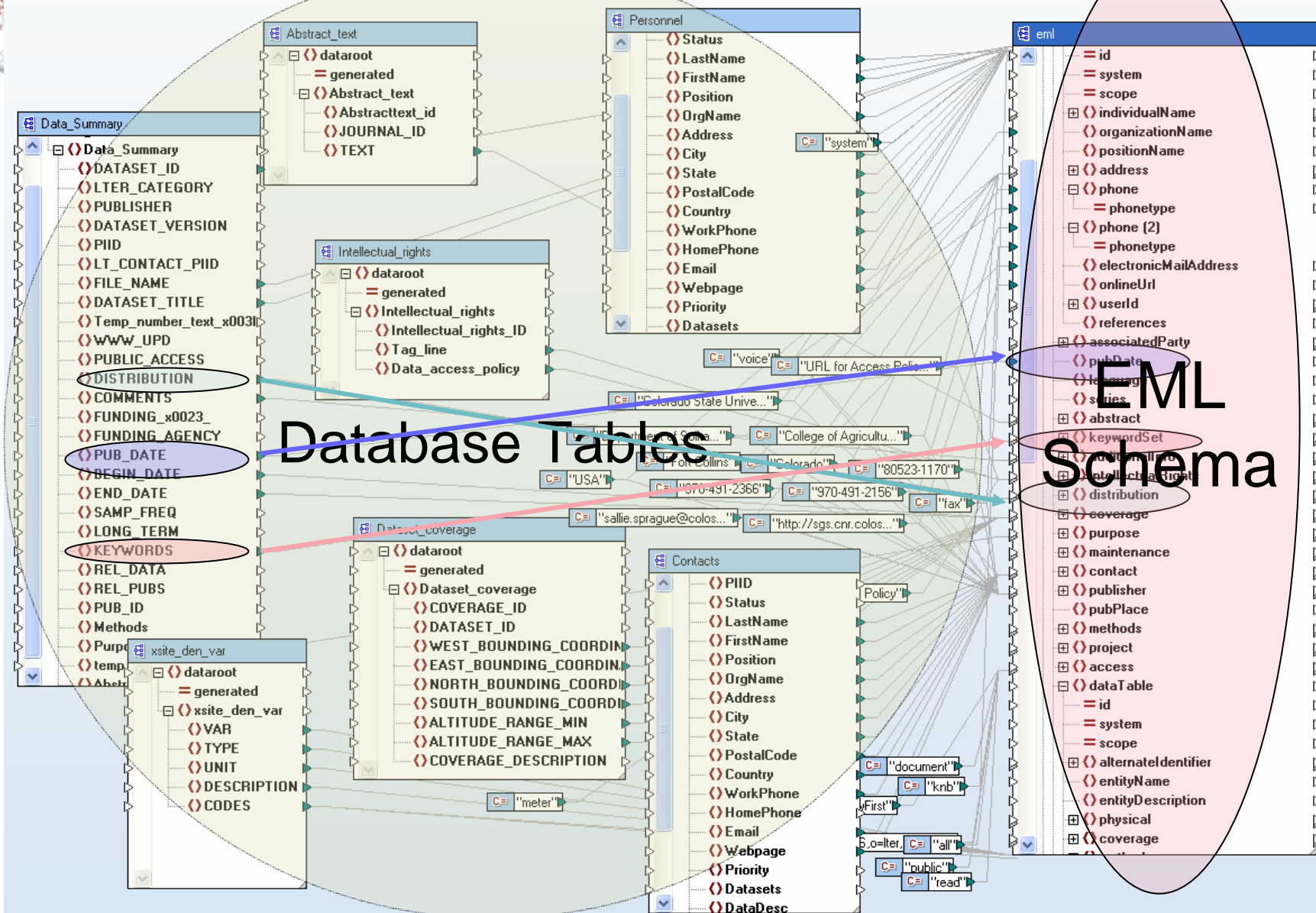
Database examples

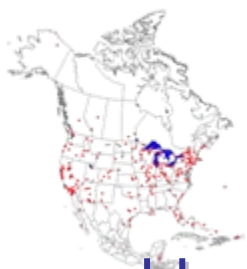
Relational, Hierarchical, Object, Hybrid,
Open source, Proprietary,
Oh my!



File Systems:
-Unix
-Windows







Flat text to EML

- Useful when text sources

ARCTIC LTER DATABASE

(1) FILE NAME: 2004e5chlr.txt

(2) YEAR: 2004

(3) PI: Dr. Anne Giblin

(4) OTHERS: Chris Crockett, Lori Winters, Sam Kelse

(5) BRIEF DESCRIPTION OF DATA FILE: Corrected Chlorophyll a for lake E5 during the summer of 2004

(6) KEYWORDS: Chlorophyll a

(7) SITE TYPE: AQUATICS-LAKES

(8) RESEARCH LOCATION: Toolik Field Station, North Slope, Alaska.

(9) EXPERIMENTAL DESIGN AND METHODS:
Water samples are collected at a minimum of the epi, meta, and hypolimnion, samples at other depths are collected as needed dependent on lake

```
<?xml version="1.0" encoding="UTF-8"?>
<eml:eml xmlns:eml="eml://ecoinformatics.org/eml-2
C:\eml-2.0.1\eml.xsd" packageId="" system="">
```

```
<dataset>
```

```
<title> </title>
```

```
<creator>
<individualName>
```

```
<title> </title>
```

```
<creator>
<individualName>
```

```
<abstract>
```

```
<section>
```

```
<para>Corrected Chlorophyll a
for lake E5 during the summer
of 2004</para>
```

```
</section>
```

```
</abstract>
```

```
<surName/>
</individualName>
</contact>
```

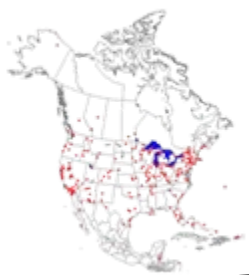
```
</dataset>
```

```
</eml:eml>
```

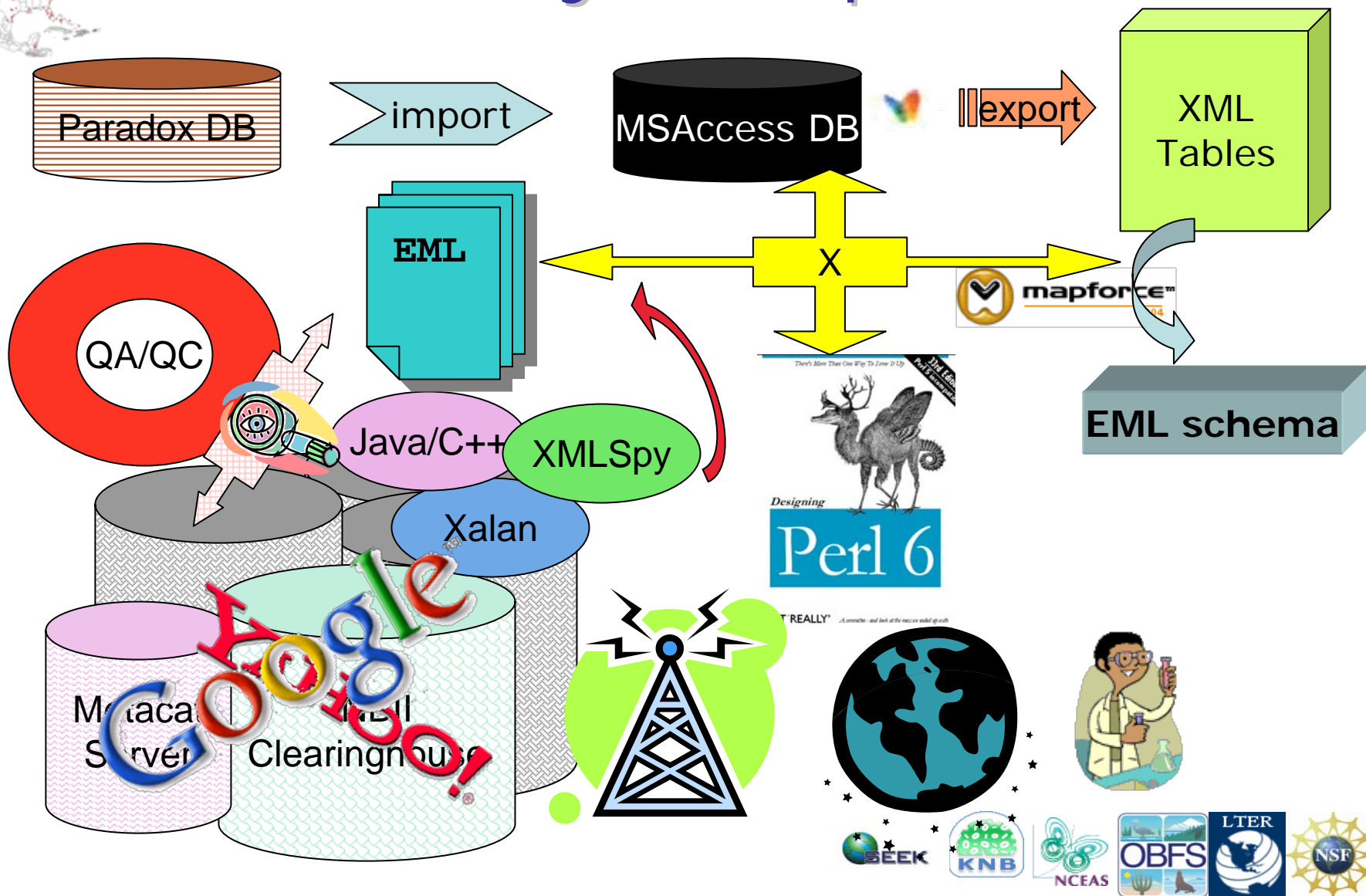


Case study – Luquillo





Case study – Luquillo





Links, book references

- www.altova.com (XML Spy editor)
- www.oxygenxml.com (XML editor)
- www.activestate.com (Free Perl runtime compiler)

