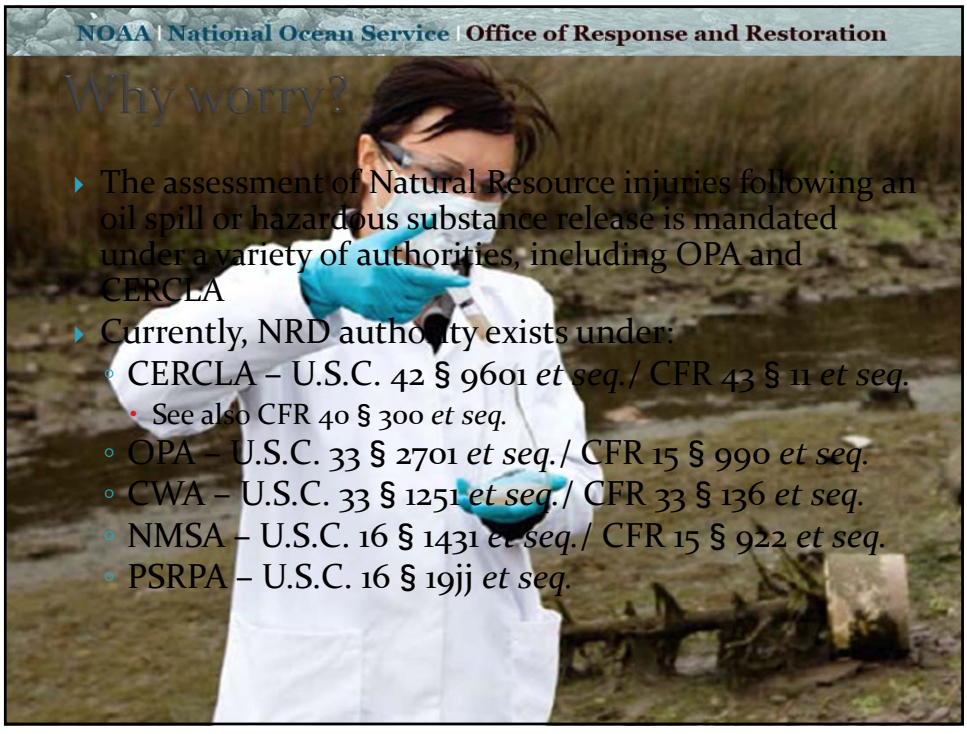
NOAA | National Ocean Service
Office of Response and Restoration

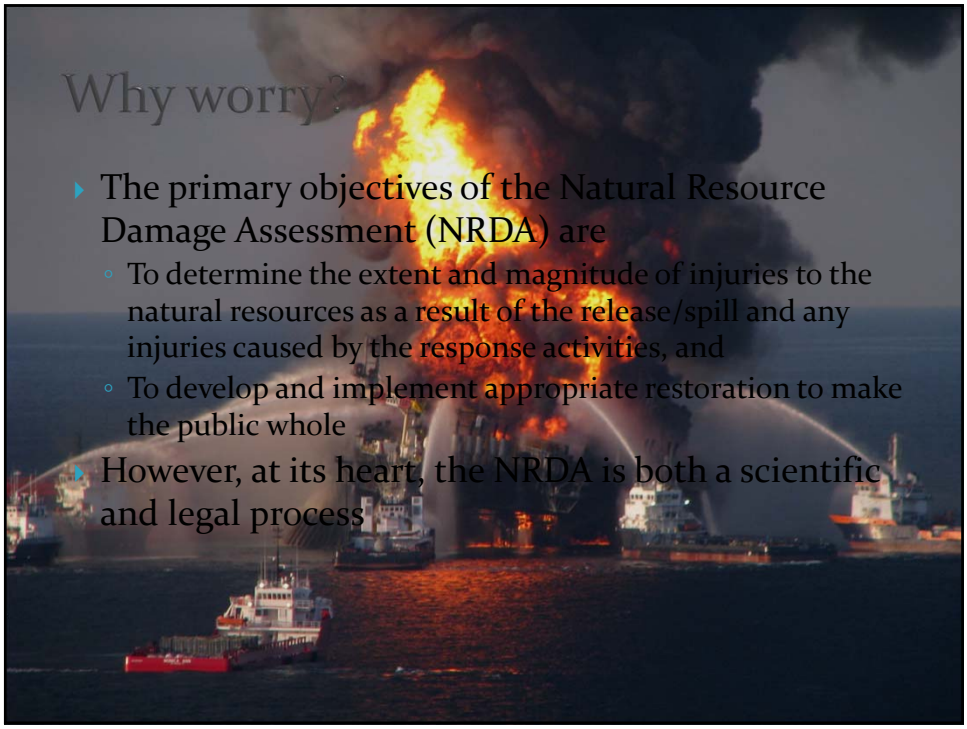
Data and Information Management Needs for Natural Resource Damage Assessments

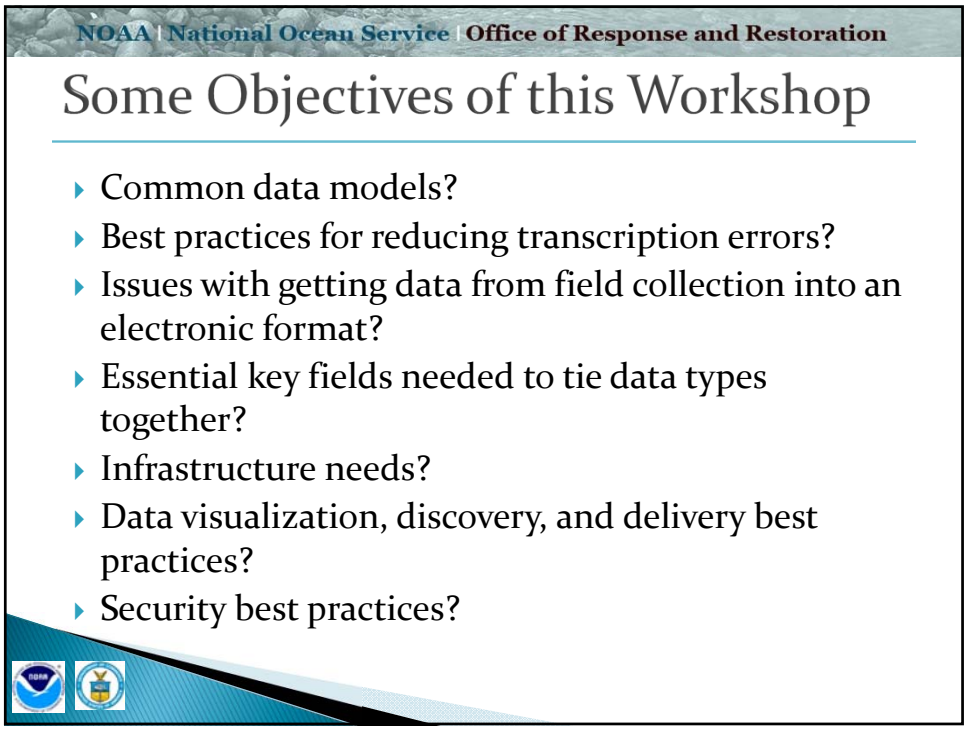
(why this is important to the users and abusers!)

Environmental Disasters Data Management Workshop
Sept 16-17, 2014
National Conservation Training Center

Robert Haddad, PhD
Amy Merten, PhD
NOAA Office of Response & Restoration and
Damage Assessment, Remediation, & Restoration Program




Why worry?
The primary objectives of the Natural Resource Damage Assessment (NRDA) are
To determine the extent and magnitude of injuries to the natural resources as a result of the release/spill and any injuries caused by the response activities, and
To develop and implement appropriate restoration to make the public whole
However, at its heart, the NRDA is both a scientific and legal process


NOAA | National Ocean Service | Office of Response and Restoration
Some Objectives of this Workshop
Common data models?
Best practices for reducing transcription errors?
Issues with getting data from field collection into an electronic format?
Essential key fields needed to tie data types together?
Infrastructure needs?
Data visualization, discovery, and delivery best practices?
Security best practices?

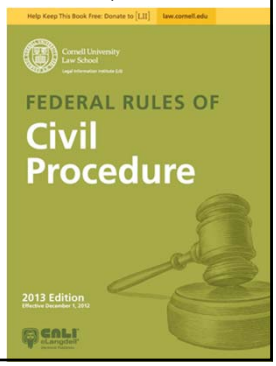
My Drivers… Science

- Science performed in any system, but especially within the complexity of natural ecosystems requires strong hypothesis testing
  - In turn, our ability to test hypotheses and reduce or minimize natural variability inherent in these systems requires a considerable amount of high quality data
  - Access to these data, to the underlying QA/QC information and other metadata is critical for any scientific investigation
  - The ability to integrate all of the different data types and then visualize the results of our analyses are also critical to our success in identifying and quantifying injury

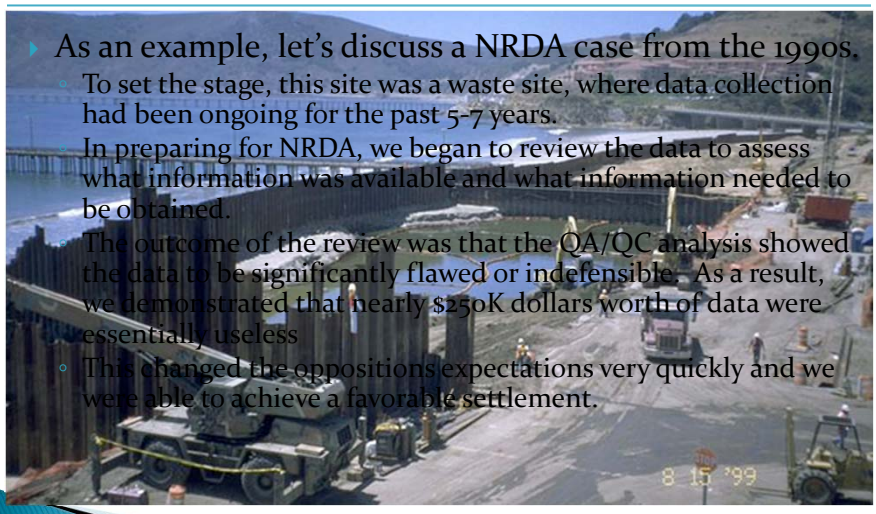
My Drivers…Legal Framework

- Every NRDA is conducted within a legal framework. It is always our desire to settle, but we have to be prepared for litigation
- Within this legal framework, all of the data we collect and use to develop our injury analyses are subject to the highest level of scrutiny

Thus, we must be able to explain and defend the appropriateness of how these data were collected, analyzed, and interpreted in the adversarial arena of the courts



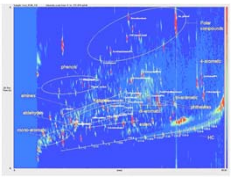
Really??



- As an example, let's discuss a NRDA case from the 1990s.
  - To set the stage, this site was a waste site, where data collection had been ongoing for the past 5-7 years.
  - In preparing for NRDA, we began to review the data to assess what information was available and what information needed to be obtained.
  - The outcome of the review was that the QA/QC analysis showed the data to be significantly flawed or indefensible. As a result, we demonstrated that nearly $250K dollars worth of data were essentially useless
  - This changed the oppositions expectations very quickly and we were able to achieve a favorable settlement.

What does this have to do with DM?

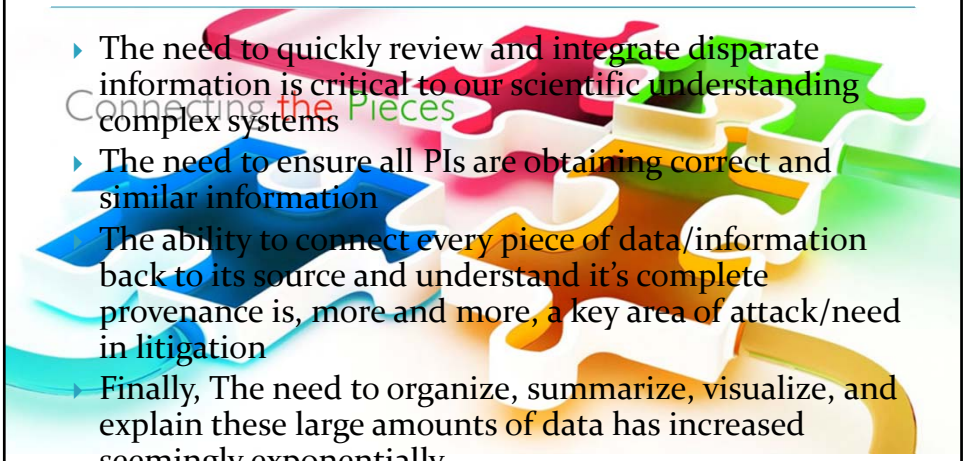
- The amount of information collected today far exceeds that collected historically – in almost every case
  - More laboratory analytical samples
  - More *in situ* instrument-derived results
  - More telemetry information
  - More digital photos
  - More modeling results



- And for assessment purposes, we need to capture data and information not only derived from the assessment
  - We need to be able to access and integrate historical baseline data/information
  - We need to be able to access and integrate much of the data/information developed during the response by the response agencies (think SCAT, dispersant usage, oil trajectory information [models, photos, observations, & remote sensing], etc.)
  - All of these data become pertinent to the NRDA

NOAA | National Ocean Service | Office of Response and Restoration

What does this have to do with DM?

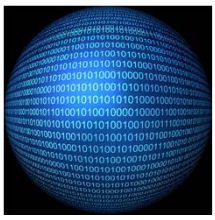


- The need to quickly review and integrate disparate information is critical to our scientific understanding complex systems
- The need to ensure all PIs are obtaining correct and similar information
- The ability to connect every piece of data/information back to its source and understand it's complete provenance is, more and more, a key area of attack/need in litigation
- Finally, The need to organize, summarize, visualize, and explain these large amounts of data has increased seemingly exponentially

NOAA | National Ocean Service | Office of Response and Restoration

What do we mean when we say Data?

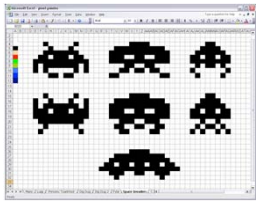
- There are others in the room who will talk about what types of data are important – and Amy touched on this point, earlier
- However, from the NRDA perspective, we see data as an encompassing term – A few broad examples of what we use:
  - Field and laboratory collected analytical data & methodologies (*e.g.*, analytical lab derived chemistry, in situ measurements of DO, etc.),
  - climactic/meteorological data,
  - Photos and data derived from remote sensing
  - field observations (*e.g.*, SCAT observations, species identification),
  - field determinations (*e.g.*, how many critters in a quadrat?),
  - telemetry output,
  - laboratory observations,
  - mathematical model inputs & outputs,
  - QA/QC data,
  - all associated meta data, etc.



NOAA | National Ocean Service | Office of Response and Restoration

Evolution of ARD Data Management

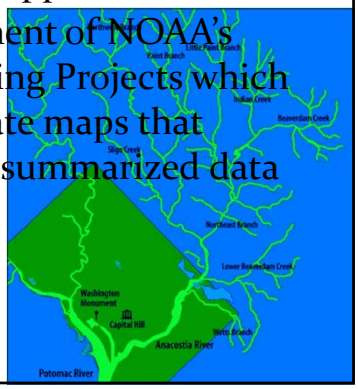
- In our Division (ARD), we have been evolving along multiple pathways
- Early efforts where focused on what was needed for a specific case and generally involved a number of excel spreadsheets shared between staff



NOAA | National Ocean Service | Office of Response and Restoration

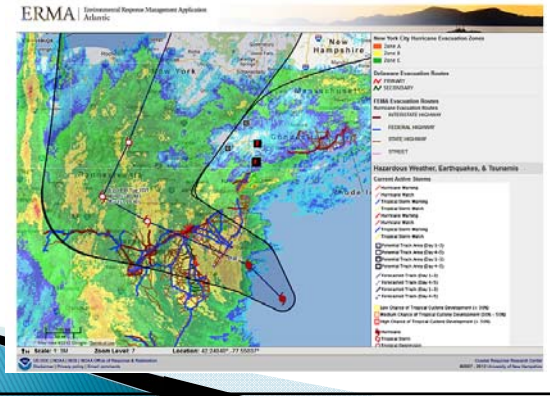
Evolution - continued

- Later evolution (*e.g.*, Query Manager) resulted from a need to better integrate analytical data from multiple sites to draw more universal conclusions and develop widely applicable models
- QM was a foundational component of NOAA's Watershed Database and Mapping Projects which provided a rapid method to create maps that displayed analyzed, sorted, and summarized data on a watershed-wide basis



Evolution - continued

- ERMA – the Environmental Response Management Application enhanced our needs for and ability to visualize many different types of data (including live feeds) and rapidly share these views with stakeholders



NOAA | National Ocean Service | Office of Response and Restoration

Evolution - continued

- DIVER – Data Integration, Visualization, Exploration, and Reporting – represents our current evolution of data management. This is a collection of tools and processes to standardize and make available to the principle investigators/scientists the vast range of data we have already discussed. This includes the ability for data mining across diverse data types with the ability to ask spatially explicit questions.

DIVER
Data Integration Visualization Exploration and Reporting

What this means to me

- As a scientist, I need to have a high degree of confidence in the data and other information upon which I will base my conclusions
- I also need to have access to the widest base of knowledge available. Many answers will not be simple, and instead will be identified as a probability in a weight of evidence analysis

NOAA | National Ocean Service | Office of Response and Restoration

What this means to me

- As an NRDA practitioner, I need to know that the data and information I use is scientifically valid and that the interpretations I draw from those data will be scientifically and legally defensible
- As head of NOAA's Damage Assessment group, I need to know that at a programmatic level, our Damage Assessment Claims are scientifically and legally defensible

Data Management Plans – why?

From C.Titus Brown, May 2010 (http://ivory.idyll.org/blog/data-management.html)

"*Now, as to my actual data management plan, here is how I plan to deal with research data in the future.*

I will store all data on at least one, and possibly up to 50, hard drives in my lab. The directory structure will be custom, not self-explanatory, and in no way documented or described. Students working with the data will be encouraged to make their own copies and modify them as they please, in order to ensure that no one can ever figure out what the actual real raw data is.

Backups will rarely, if ever, be done.

When required to make the data available by my program manager, my collaborators, and ultimately by law, I will grudgingly do so by placing the raw data on an FTP site, named with UUIDs like 4e283d36-61c4-11df-9a26-edddf420622d. I will under no circumstances make any attempt to provide analysis source code, documentation for formats, or any metadata with the raw data. When requested (and ONLY when requested), I will provide an Excel spreadsheet linking the names to data sets with published results. This spreadsheet will likely be wrong -- but since no one will be able to analyze the data, that won't matter."