



esri

# Approaches to Eliminating Data Redundancy

*Canadian Hydrographic Conference - 2010*

*Quebec, Canada*

*Beata Van Esch (presenting)*

*Timothy Kearns, Tom De Puyt, Rafael Ponce (co-authors)*

# Trends for modern production systems

Existing Data is compiled into a central database

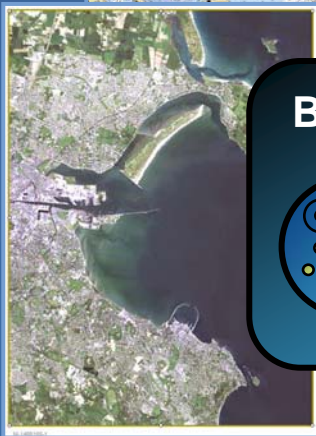
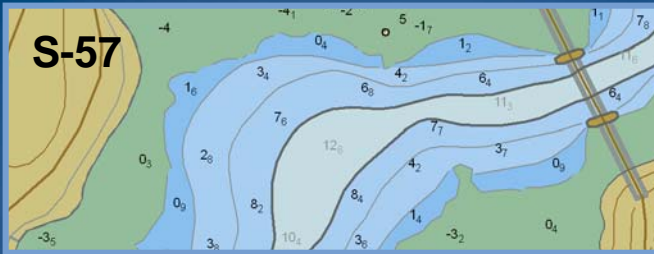
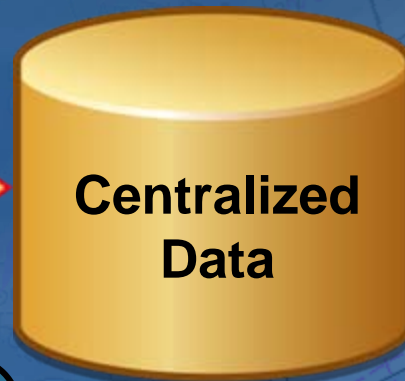
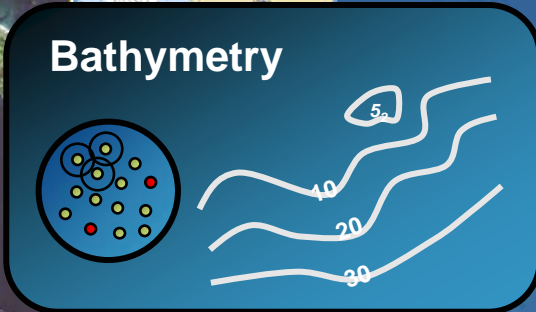
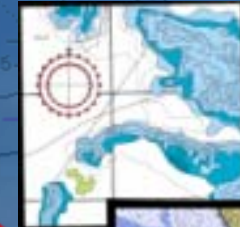


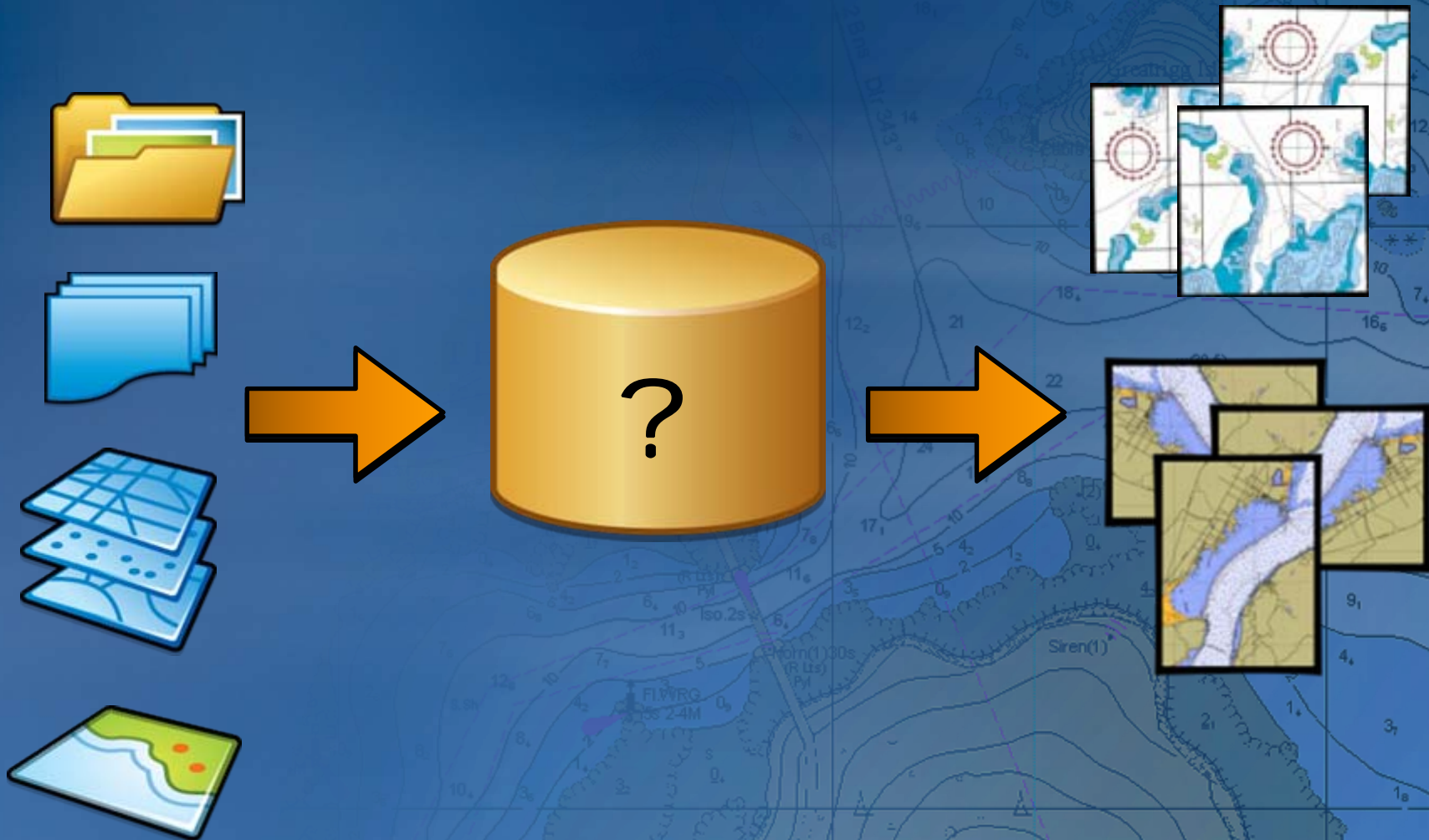
Image based sources



Changes are pushed to products



# Data Migration

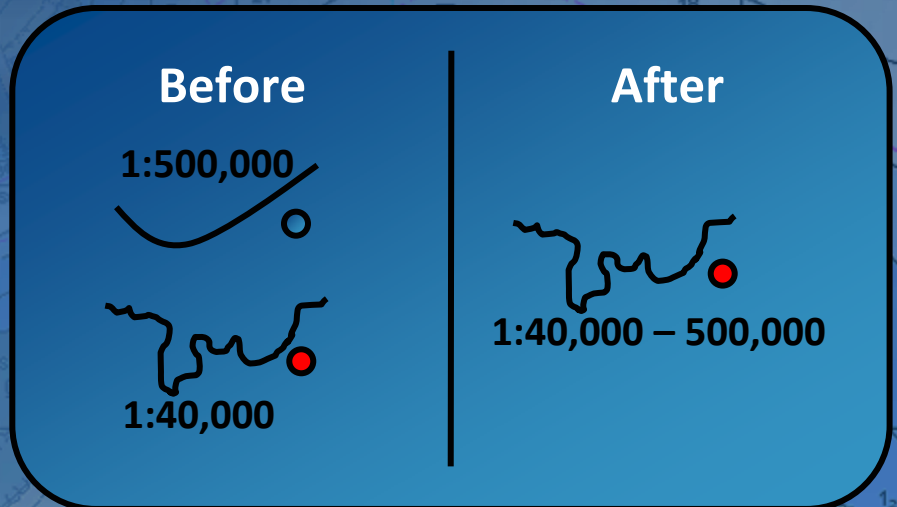


# Major considerations

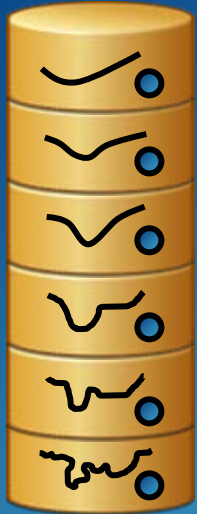
- Database platform
  - Oracle, SQL Server, PostGres, DB2, Informix, etc.
- Database architecture
  - Federated, partitioned, single instance
- Data loading
  - COTS tools, quality control, validation
- Data model
  - S-57, Navigational, facilities, comprehensive
- Data management
  - Edge-matching, compiling source, conflating features

# Eliminating data redundancy through conflation

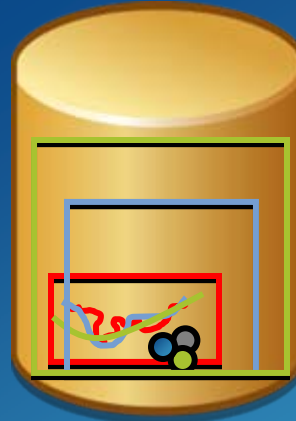
- Conflation is the process of **removing redundant features** that co-exist in the same database.
  - These features may or may not share the same:
    - Geography
    - Scale
    - Attributes
- The result is a more efficient database.
- Also referred to as:
  - **1 Touch Editing**
  - **One Feature One Time**
  - **Deconfliction**



# Approaches to Centralized Data



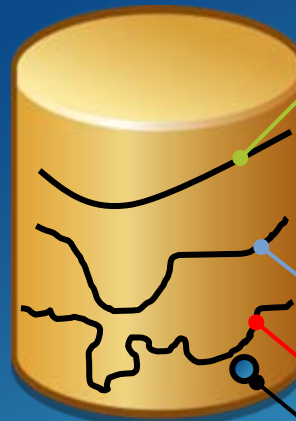
Physically separated



Product Warehouse



Best Scale



1:500,000

Scaled Features

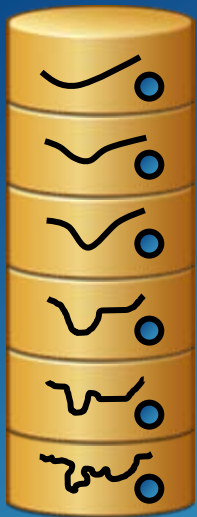
1:250,000

1:40,000

1:40,000 – 500,000

# Centralized Data

## *Physical databases*



Physically  
separated

## Practicalities

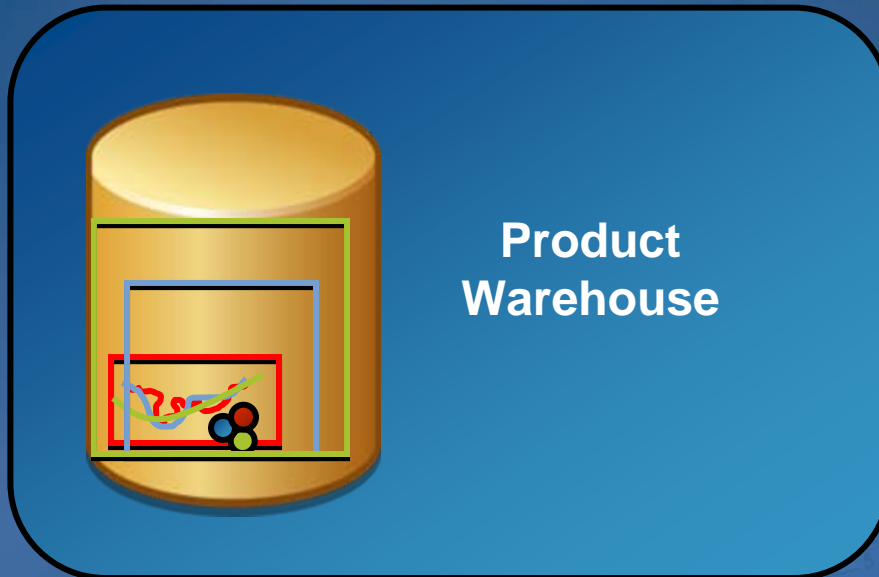
- Scale/usage bands are breakpoints
- No overlap of features within a usage band
- Easy display of features
- Potential to ‘link’ features across usage bands

## Realities:

- Lack of feature ‘awareness’
- Cannot conflate across usage bands
- Products are dependent on usage or scale bands
- Feature edits for each usage band

# Centralized Data

## *Product Warehouse*



### Practicalities

- Simple queries
- Product ‘warehouse’
- 1:1 relationship between feature and product

### Realities:

- Cannot conflate, features are dependent on products
- Feature edits for each product



# Centralized Data

## *Best Scale*



Best Scale

## Practicalities

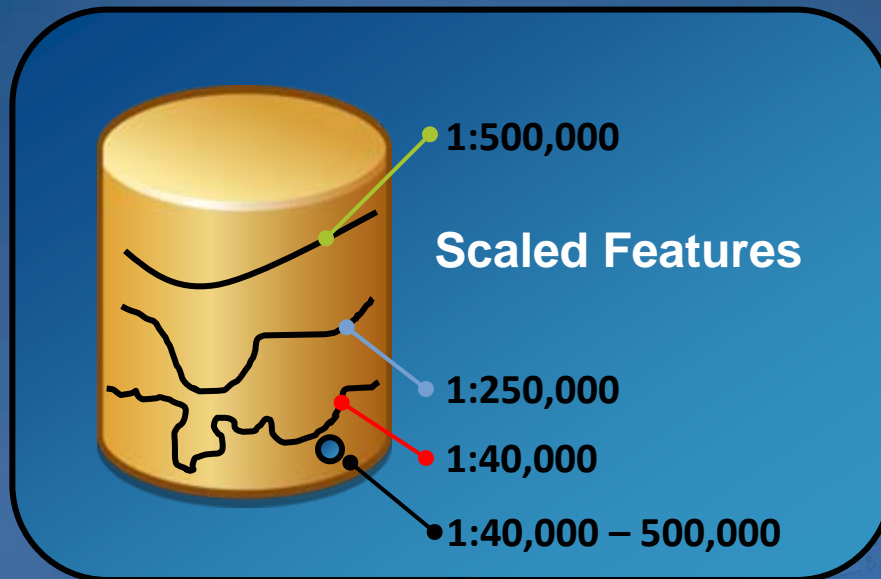
- Decreased feature load on database
- True one feature, one time for all features (100% conflation)
- Only largest (best) scale features are retained

## Realities:

- All feature generalization lost upon data load
- On the fly generalization of features required
- Edgematching of all features will be required

# Centralized Data

## *Scaled Features*



### Practicalities:

- Same feature, multiple representations in database
- Features can participate in products across scales
- Dependency on scale, not product
- Potential exists to 'link' features across scale bands
- True one feature, one time for selected features

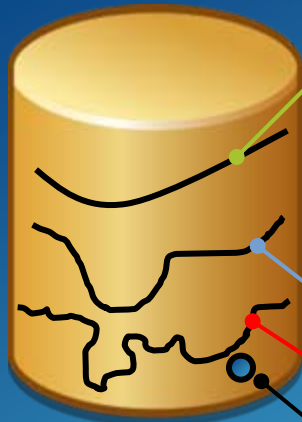
### Realities:

- Feature edits for each scale band

# Recommended Approach



Best Scale



1:500,000

Scaled Features

1:250,000

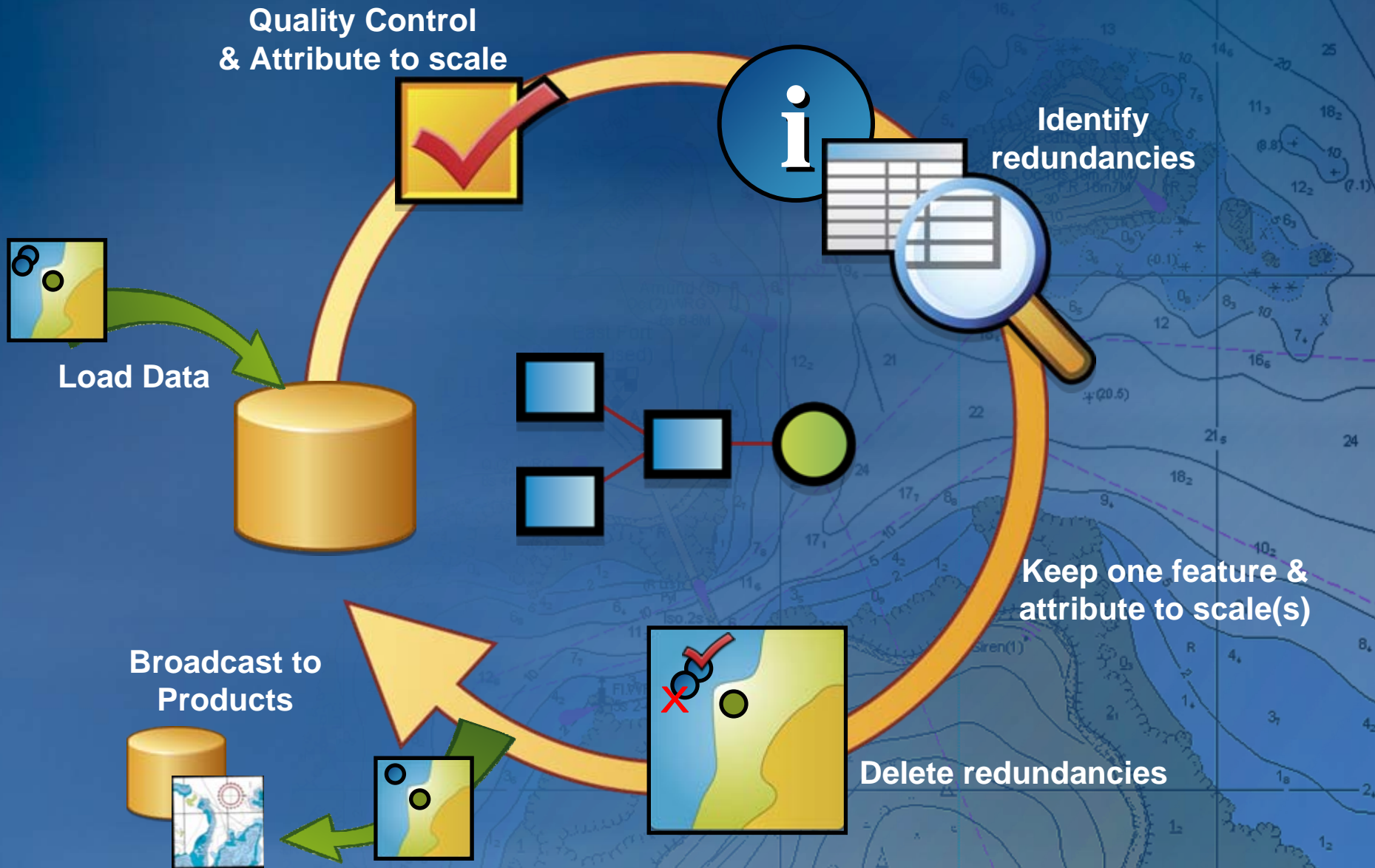
1:40,000

1:40,000 – 500,000

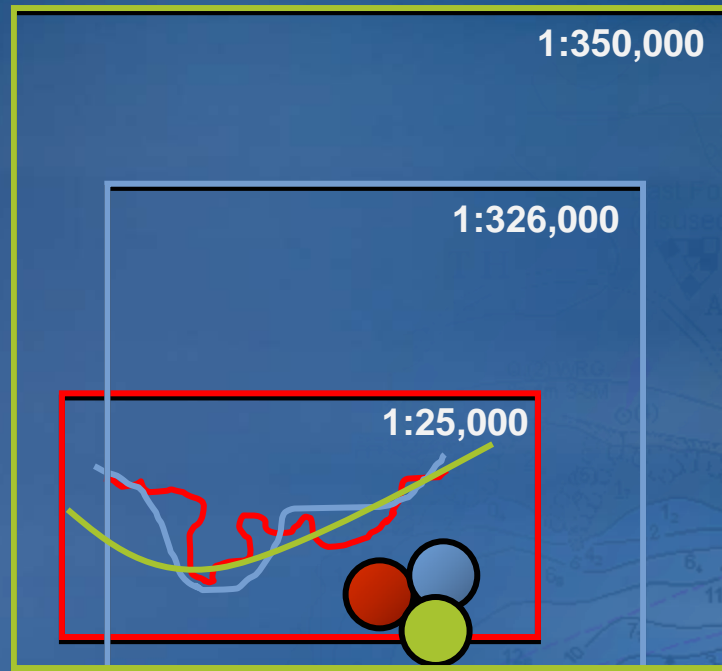
## Hybrid:

- Scale bands pre-defined
- Product neutral/seamless DB
- Features attributed by scale(s)
- Spatial / Attribute queries
- Conflation of all features within a scale band
- Conflation of selected points across entire database
- Features can exist on many products across scales

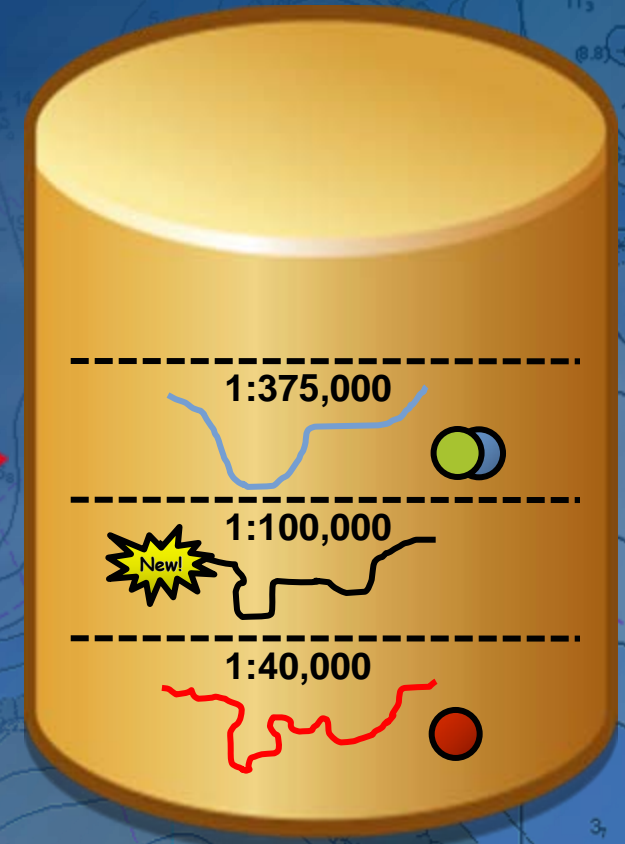
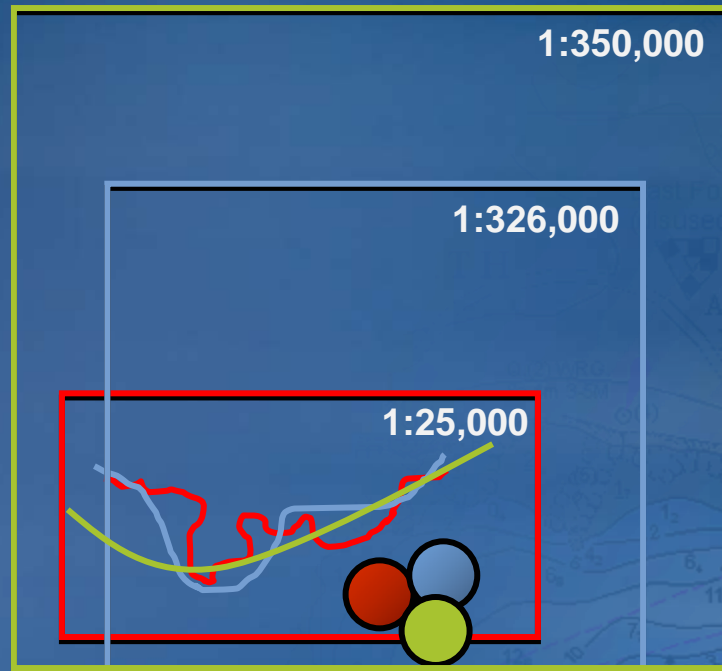
# Methodology for Conflation



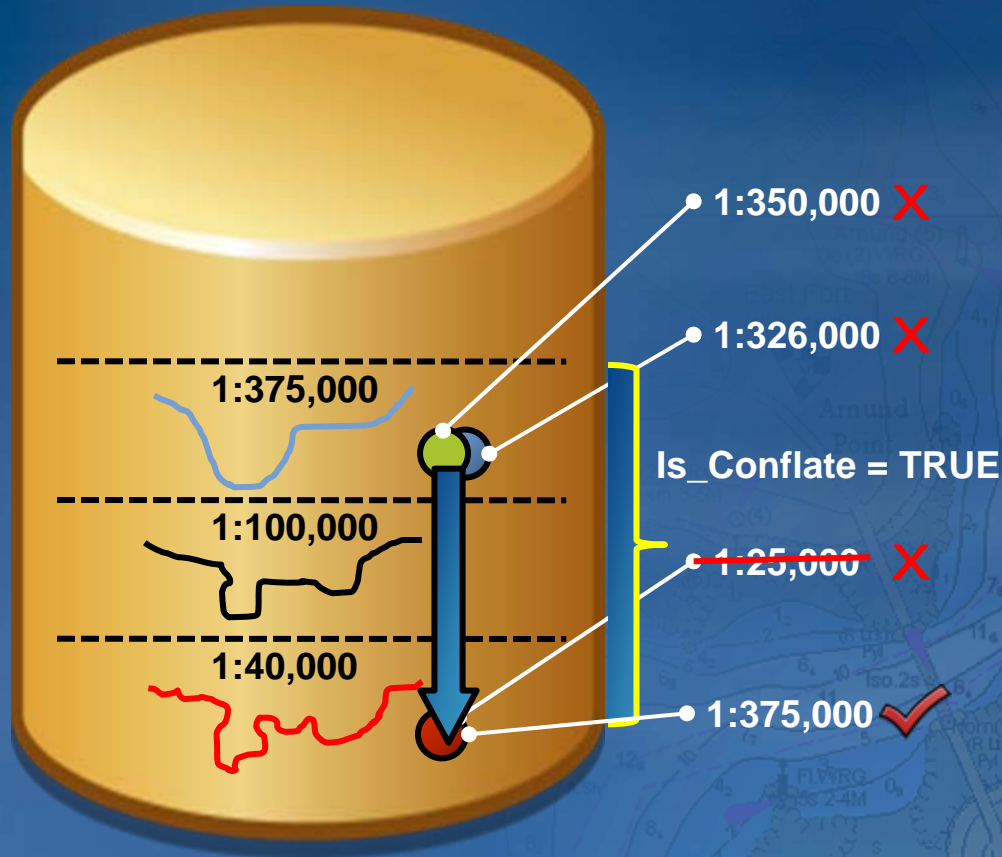
# Conflation within a scale band



# Conflation across scale bands

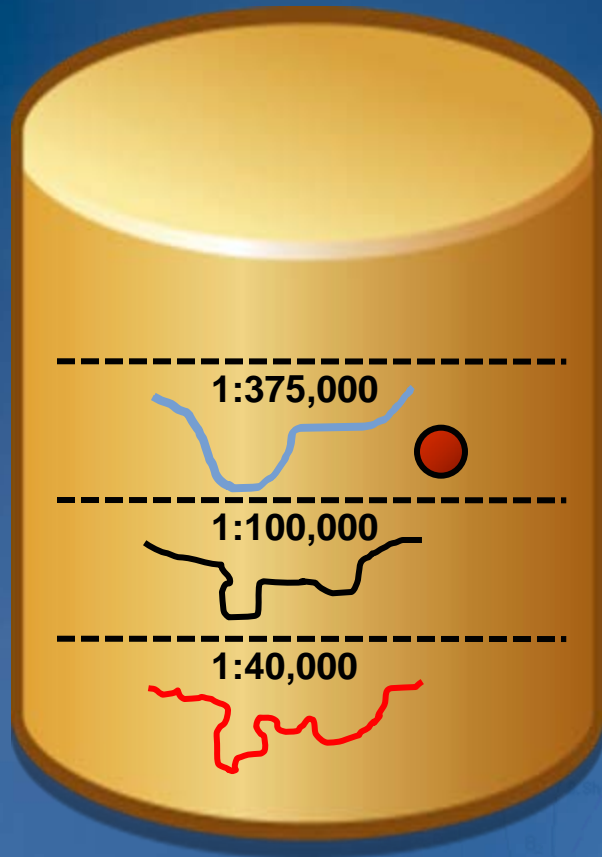


# Conflation across scale bands

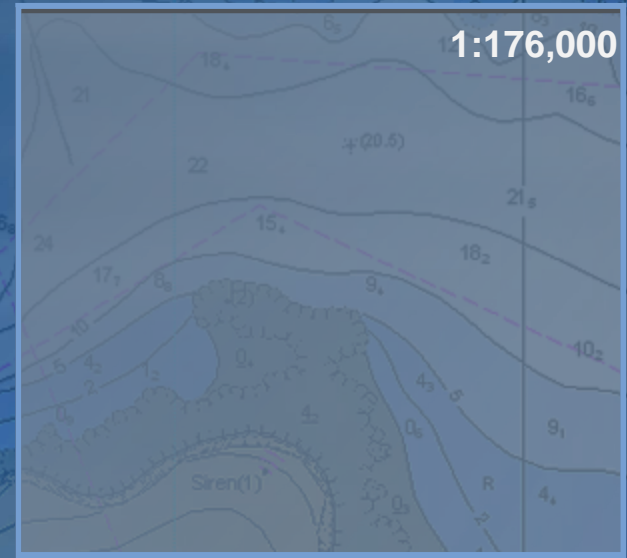


- Determination of most accurate position
- Participating scale bands
- Geometry/Attribute transfer
- Delete redundant features
- Set conflation attribute

# Creating/updating products *small scale products*

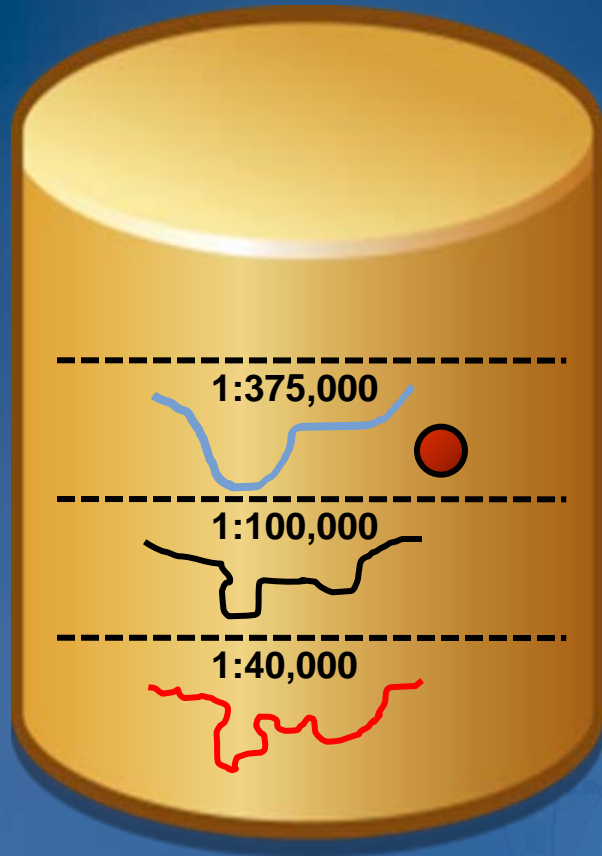


Feature  
Extraction  
Query by  
Scale and  
Conflation attribute

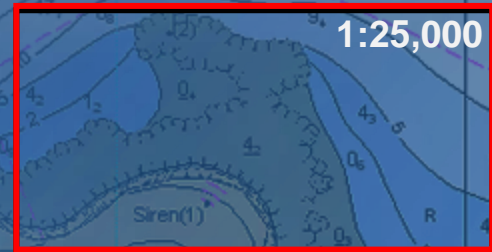




# Creating/updating products *large scale products*



Feature  
Extraction  
Query by  
Scale and  
Conflation attribute



# Summary

- Trends among hydrographic offices are:
  - Spatial data infrastructures, interoperability, centralized data
- Many significant considerations
  - Platform, architecture, data model, conflation
- Conflation is the process of **removing redundant features** that co-exist in the same database. (aka: 1F1T, One Touch Editing)
- There are at least 4 approaches
  - Best Scale, Physically separate, Scaled Features, Product warehouse
  - Practicalities and Realities to each approach
- Hybrid of Best Scale + Scaled Features is best
  - It is very difficult to truly conflate all features across all scale bands.



esri

# ESRI - Maritime *GIS for a Blue Planet*

*Beata Van Esch*

[bvanesch@esri.com](mailto:bvanesch@esri.com)

<http://www.esri.com/maritime>