

Human Data Sources

Integrating Multiple Data Streams

Unique Properties of NOAA's Commercial Fishing Data

- Humans are the primary observing platform
- A single fishing trip is observed by multiple systems
 - Vessel logbooks
 - Dealer reports
 - NOAA observers on fishing vessels
 - Satellite Vessel Monitoring Systems
- Why does this matter?

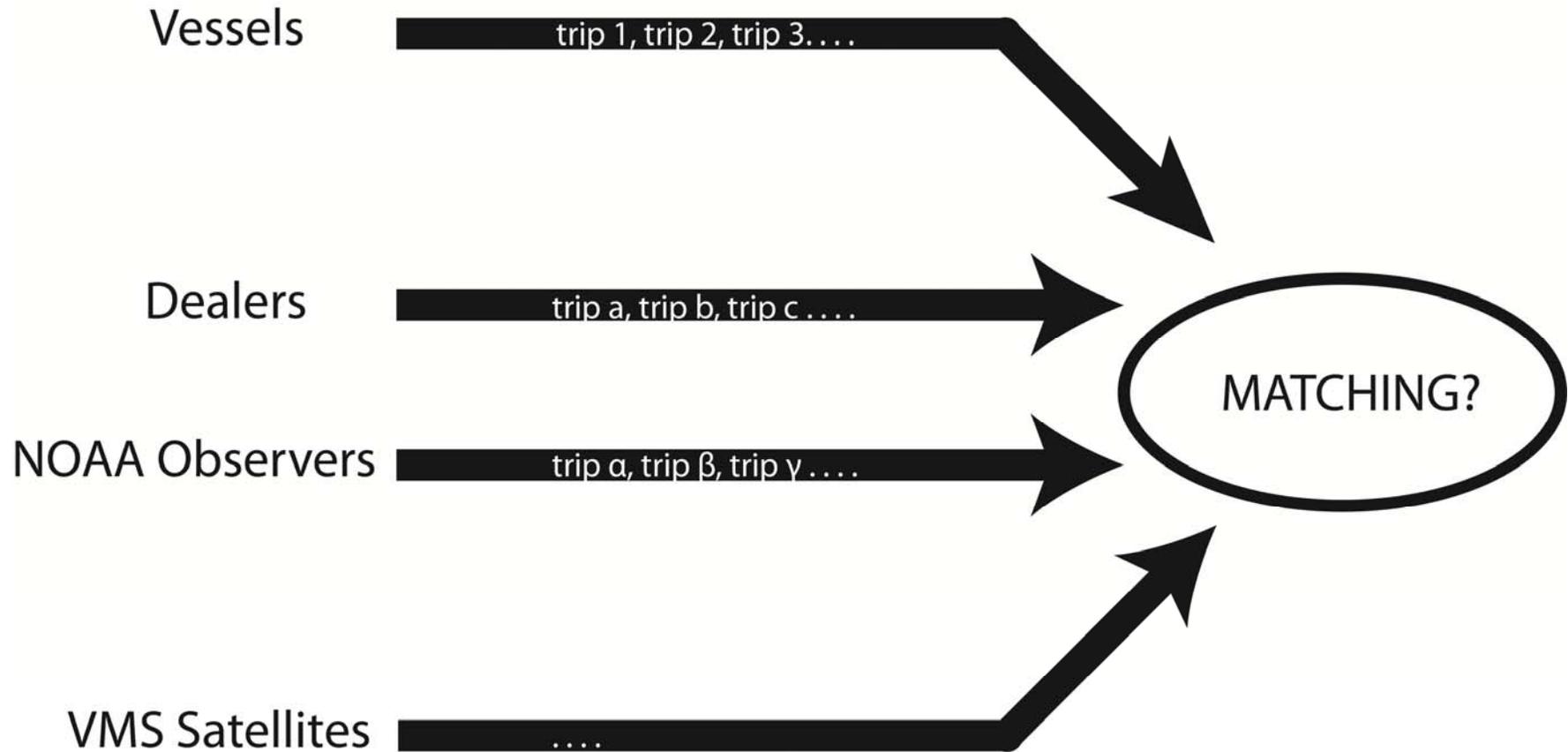
Human Data Sources Can be Unreliable

- Data entry errors
 - Typos
 - Handwriting
 - Confusion
 - Don't know (e.g. Vessels regarding Dealers)
- Other faults
 - Intentional
 - Negligent

Error Types

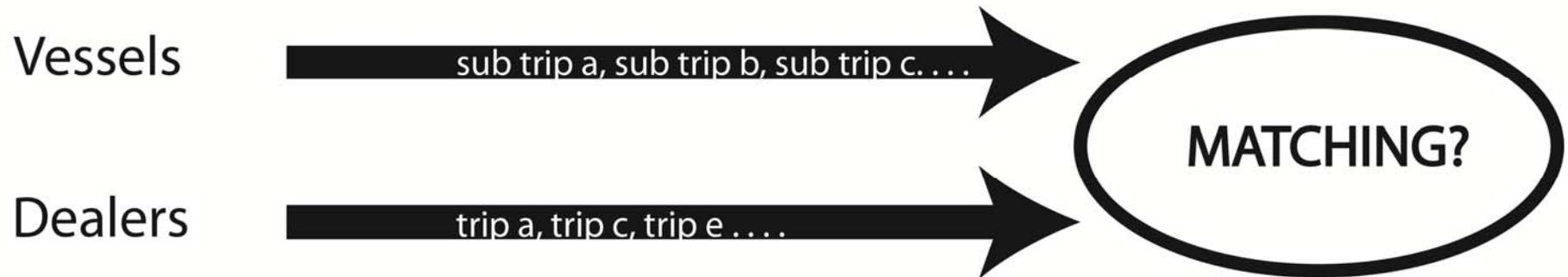
- Suppose there are 2 systems
- Invalid record in System 1 ->
- Mismatching between trips in Systems 1 & 2
- Hence, error types
 - Invalid records
 - Mismatches

Multiple Data Streams *different naming schemes*



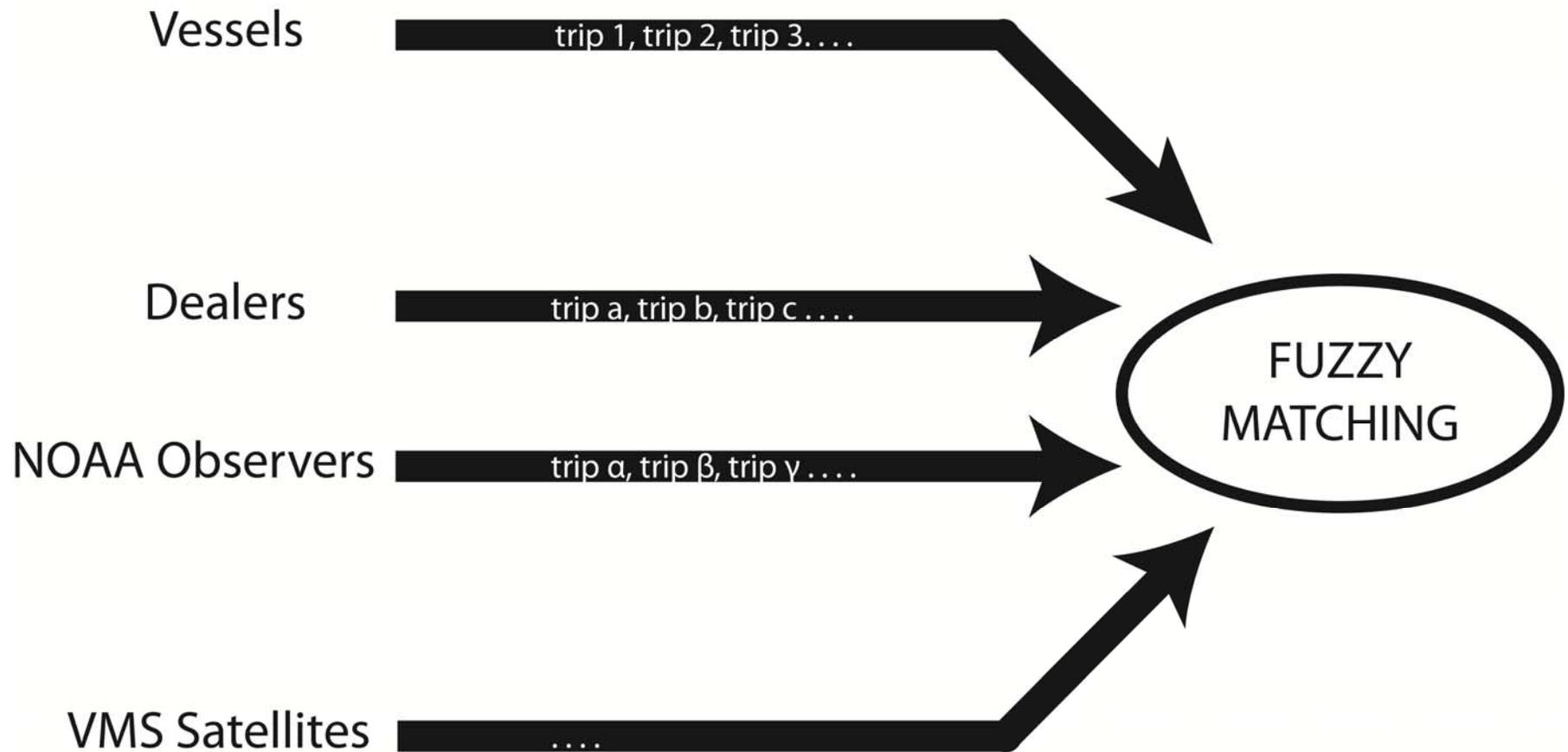
Multiple Data Streams

mismatched levels



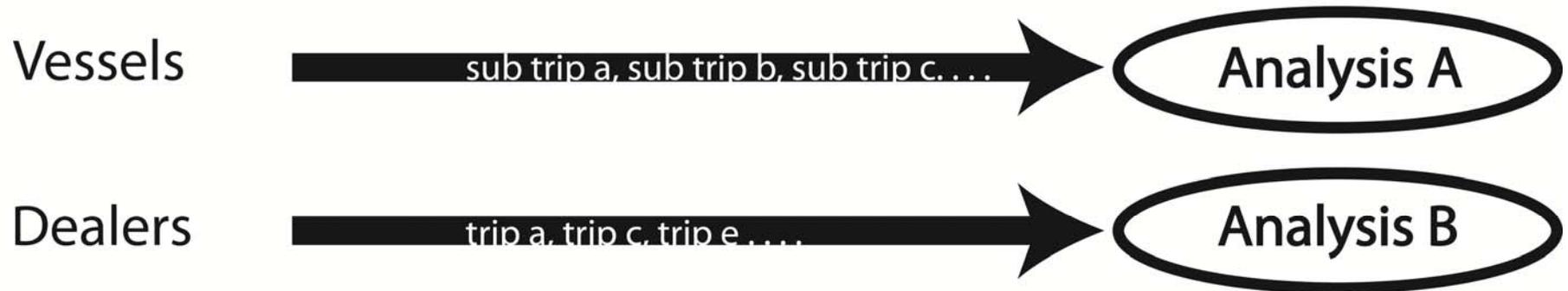
Multiple Data Streams

possible solution 1?



Multiple Data Streams

possible solution 2?

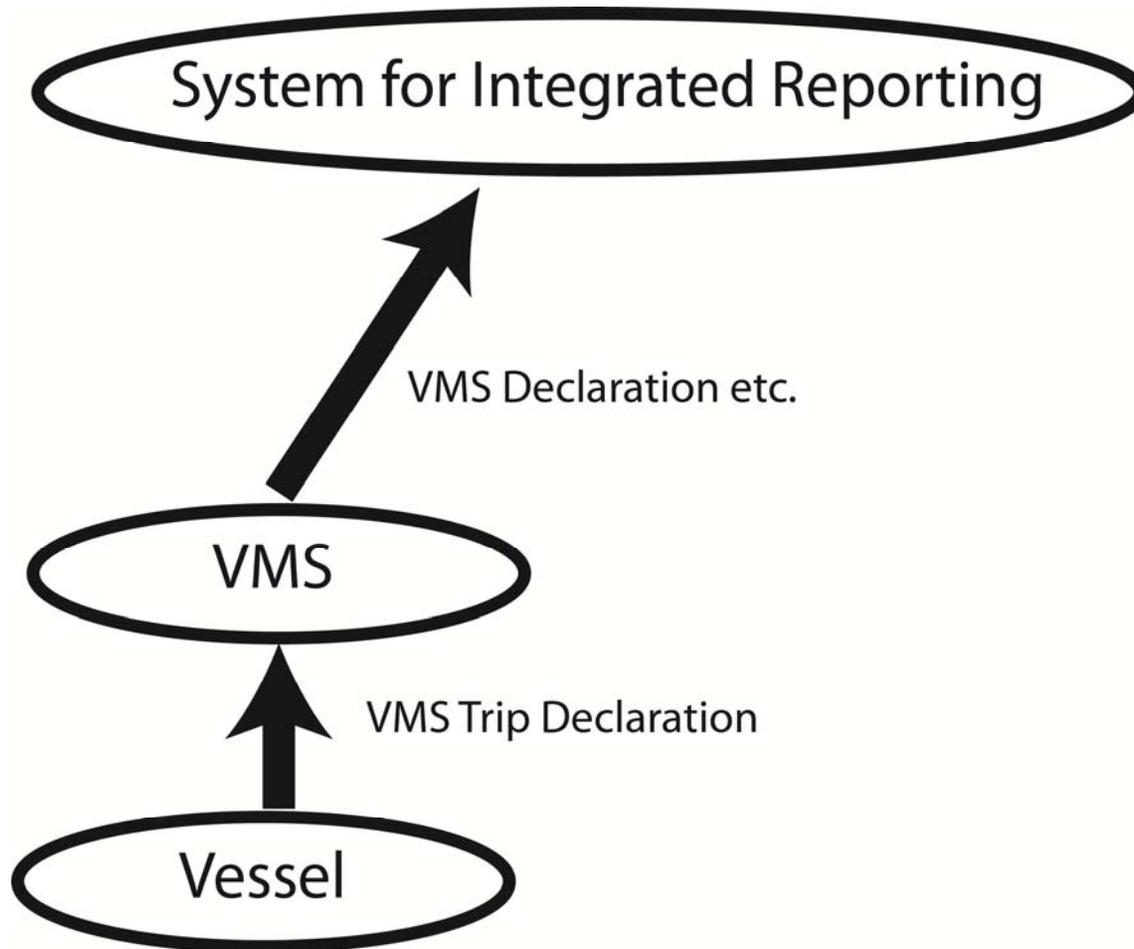


Remaining Issues

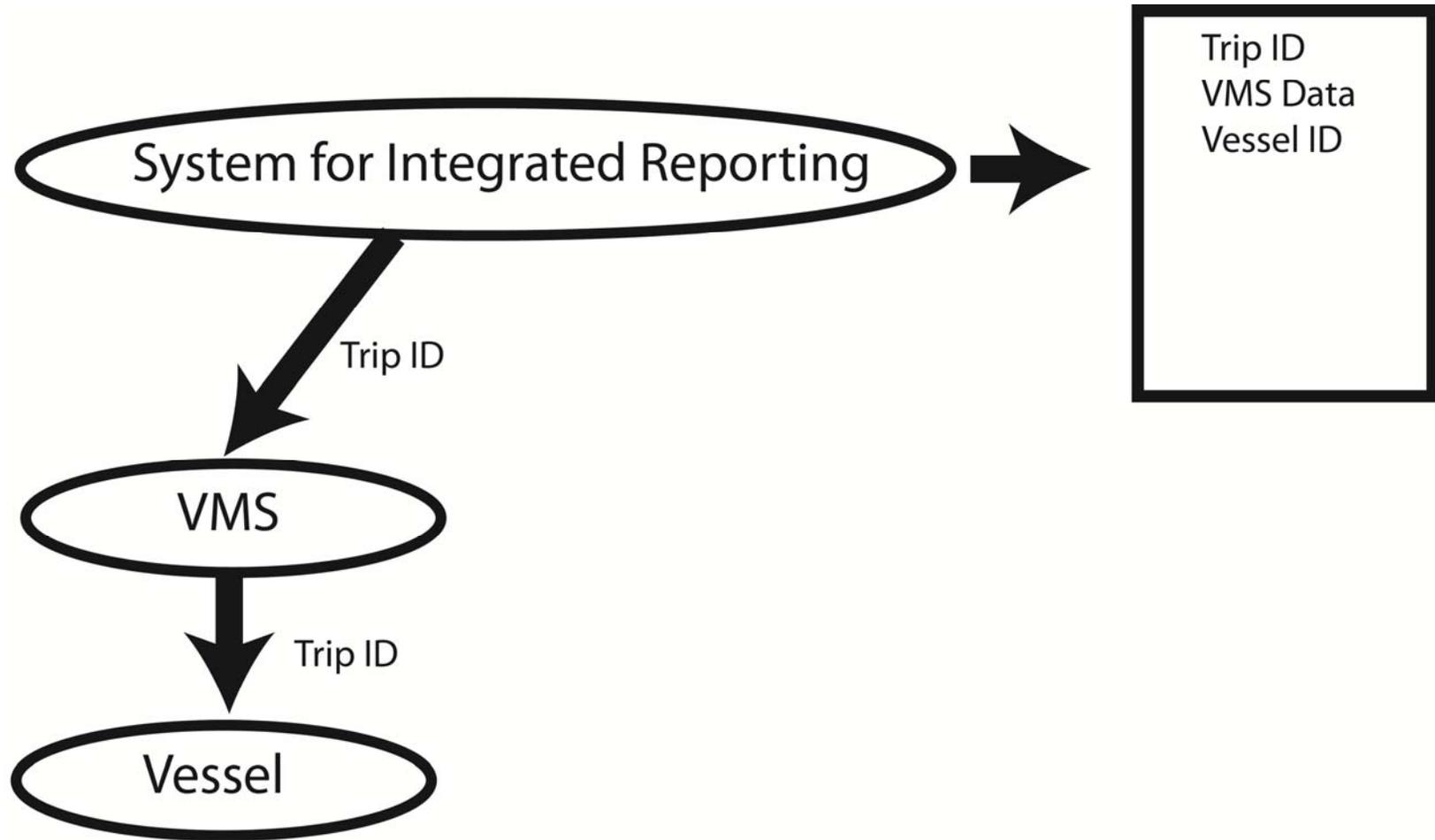
- Fuzzy matching introduces errors
- No matching sacrifices analytical power
- Human errors not addressed

The Real Solution

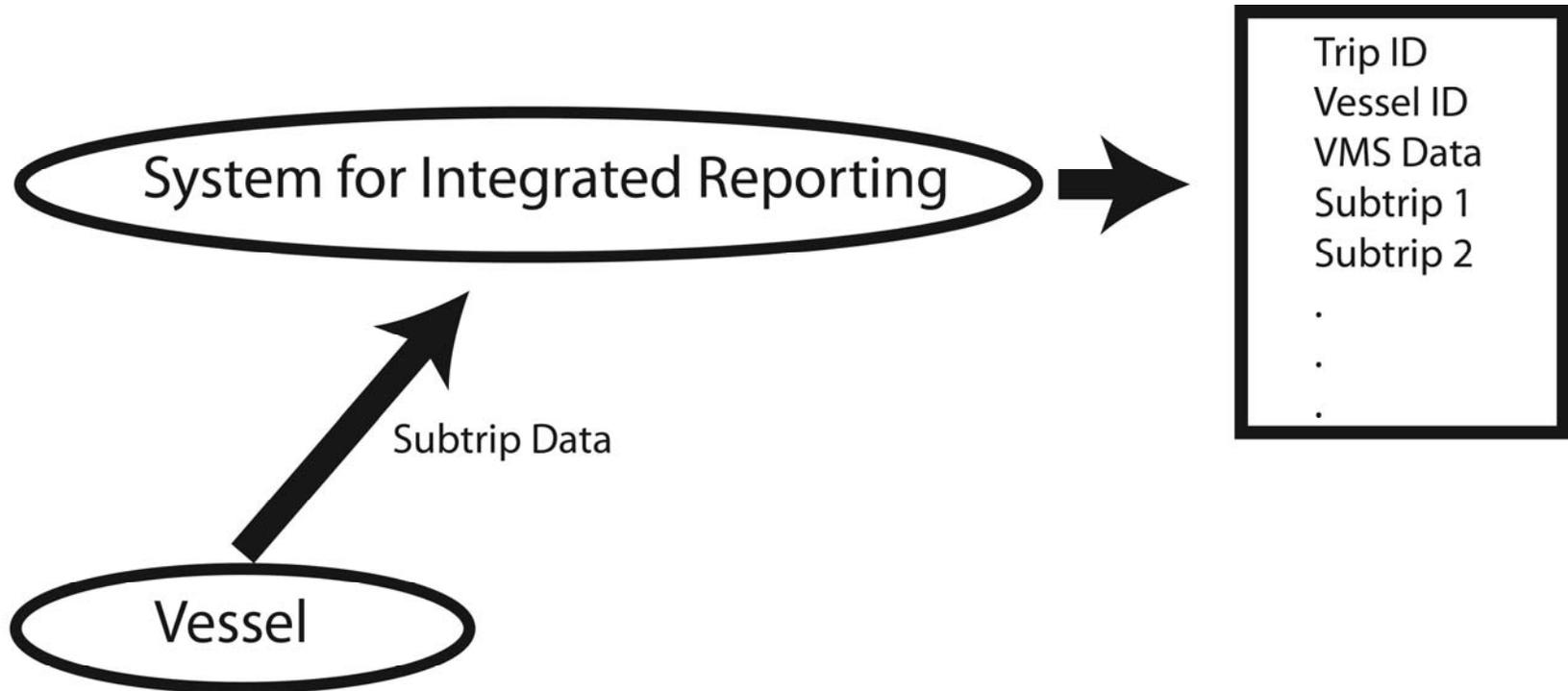
Step 1



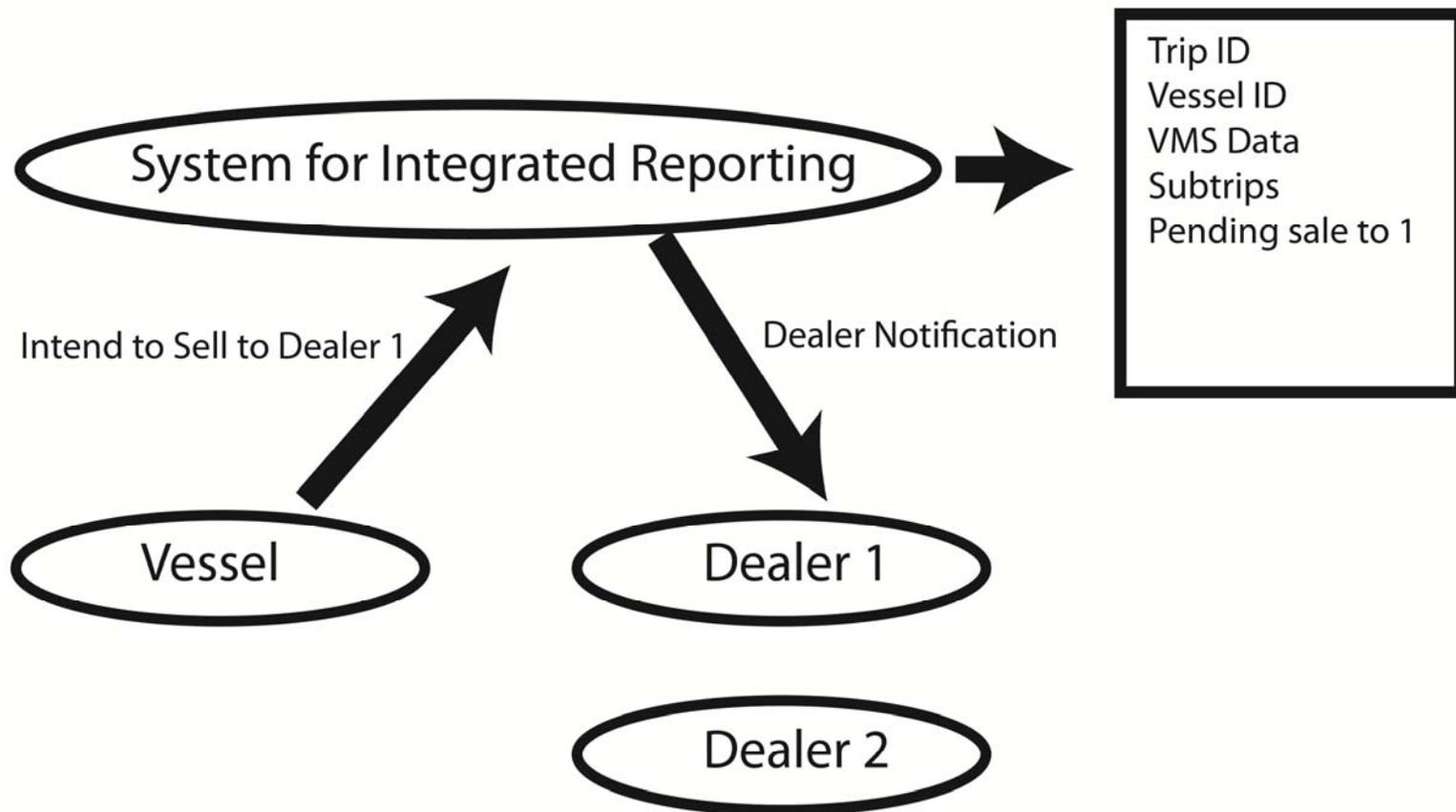
Step 2



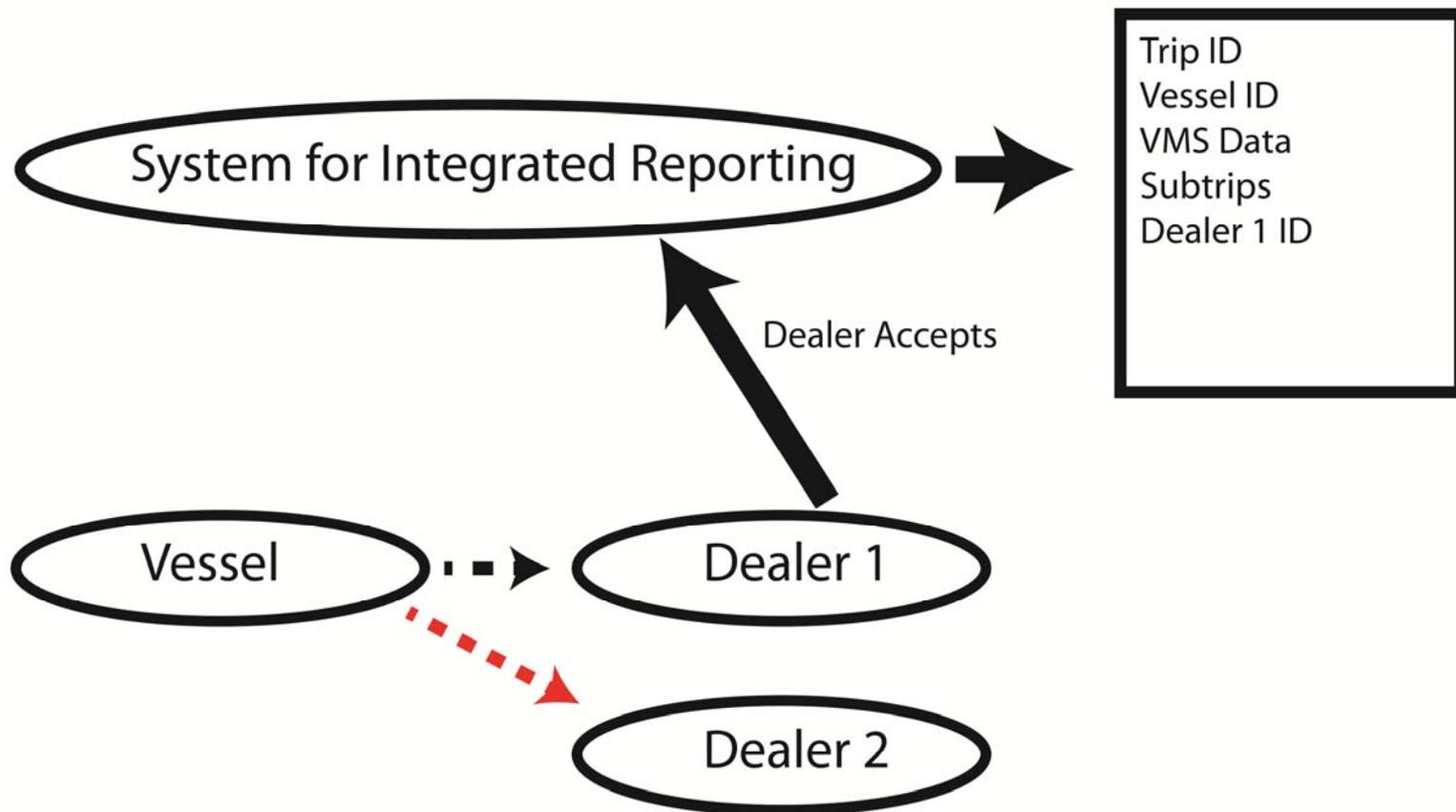
Step 3



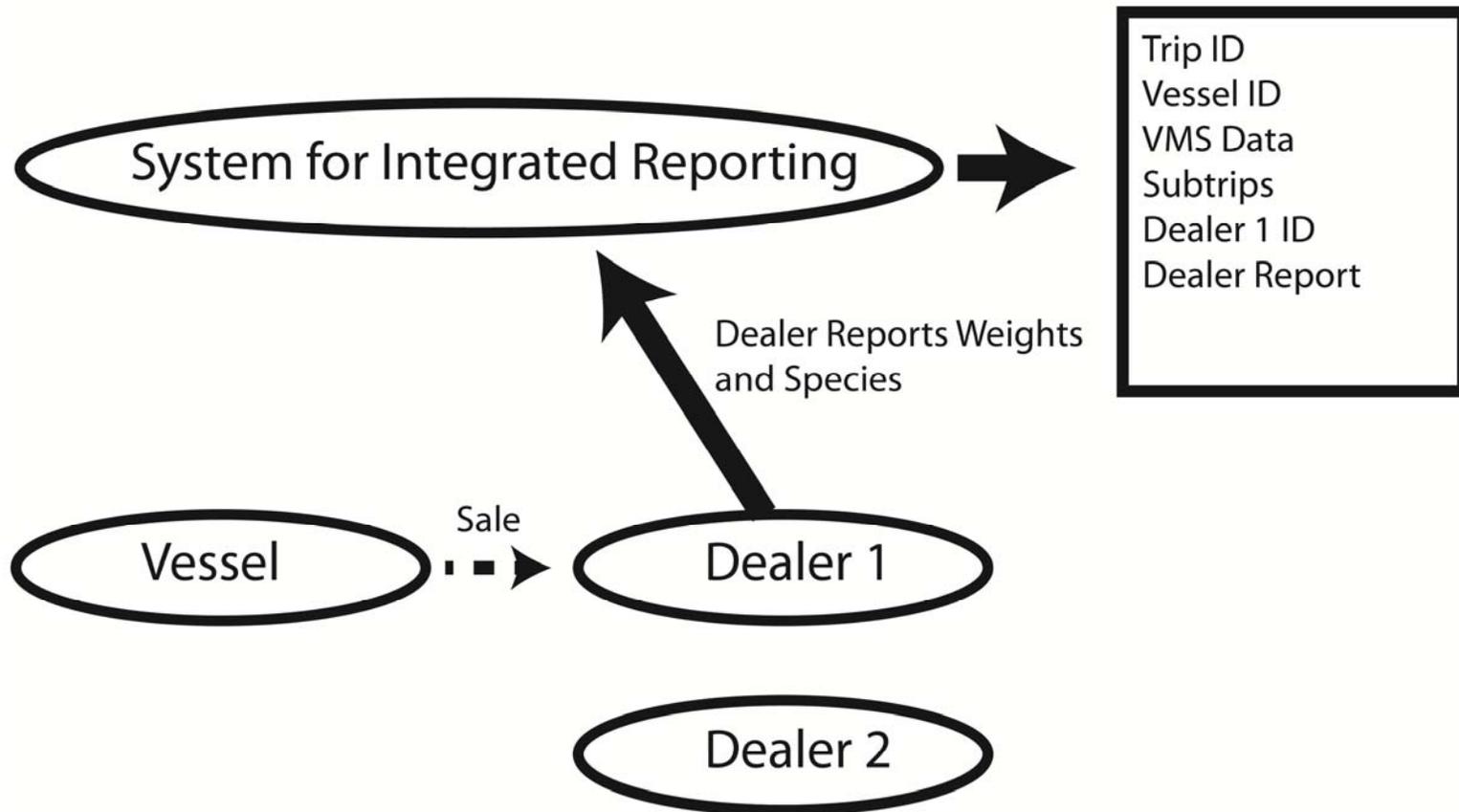
Step 4



Step 5



Final Step



Conclusions

- Most human data entry replaced by machine
- Remaining human data entry highly constrained
- Matching process facilitated at time of sale
- QA replaces QC