A Collocation Service for *In Situ* and Remotely Sensed Measurements

Steven J. Worley
National Center for Atmospheric Research
Andrew W. Bingham
PO.DAAC JPL NASA
Zaihua Ji
National Center for Atmospheric Research

Topics

1. Motivation
2. Web Service and Use Cases
3. Infrastructure and Testing Results
4. Record Content for Web Service
5. Next Steps
6. Request to the Community
7. Conclusions

Motivation

Support new research opportunities

- Compare and Contrast Independent Observing Systems (*in situ* .vs. remote)
  - Reveal biases and systematic errors – in both
- Leads to improved algorithms and QC

Community benefits

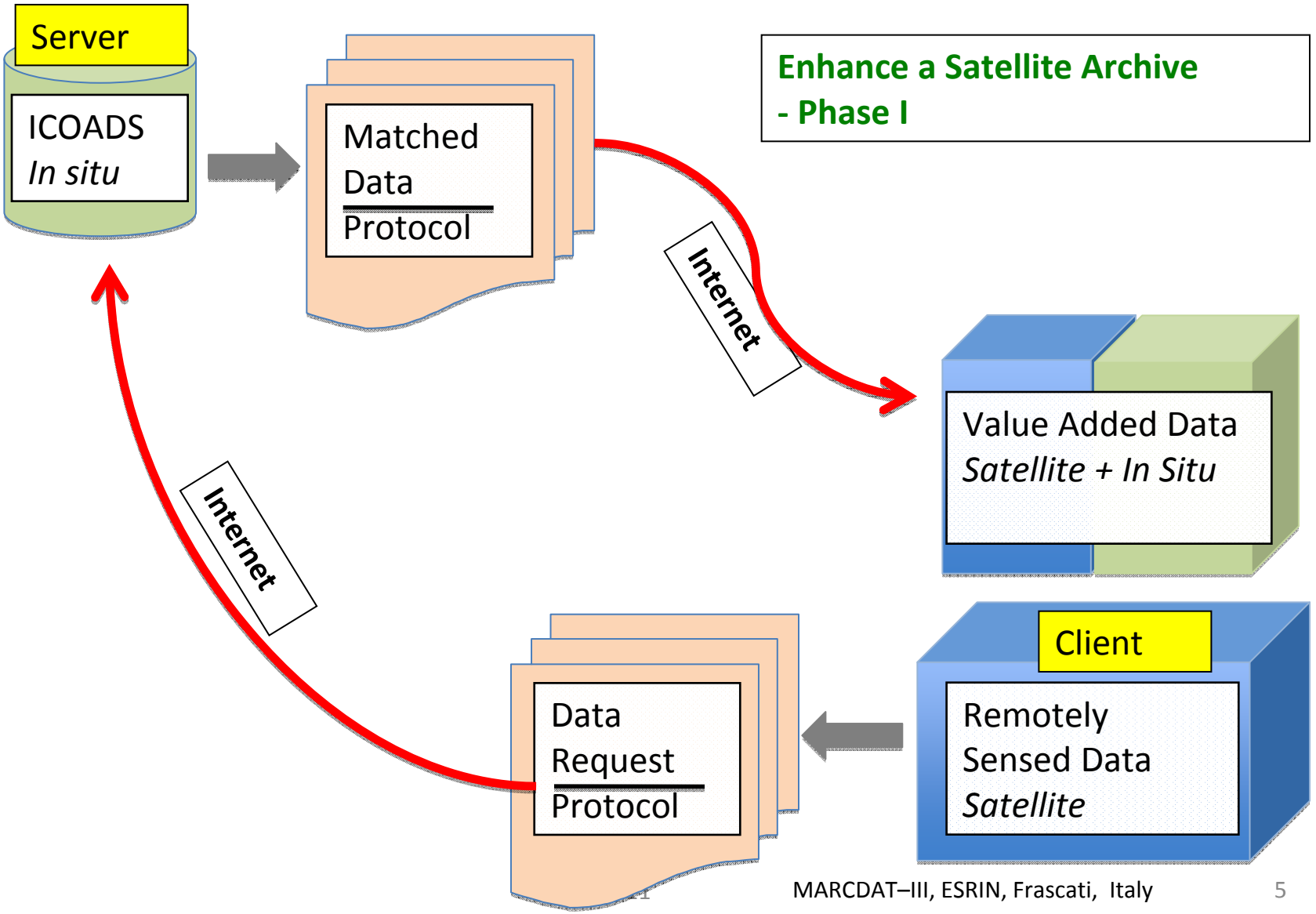
- More accurate independent data products
- More accurate combined synergistic analyses
- Easy to get research started in this domain

Motivation

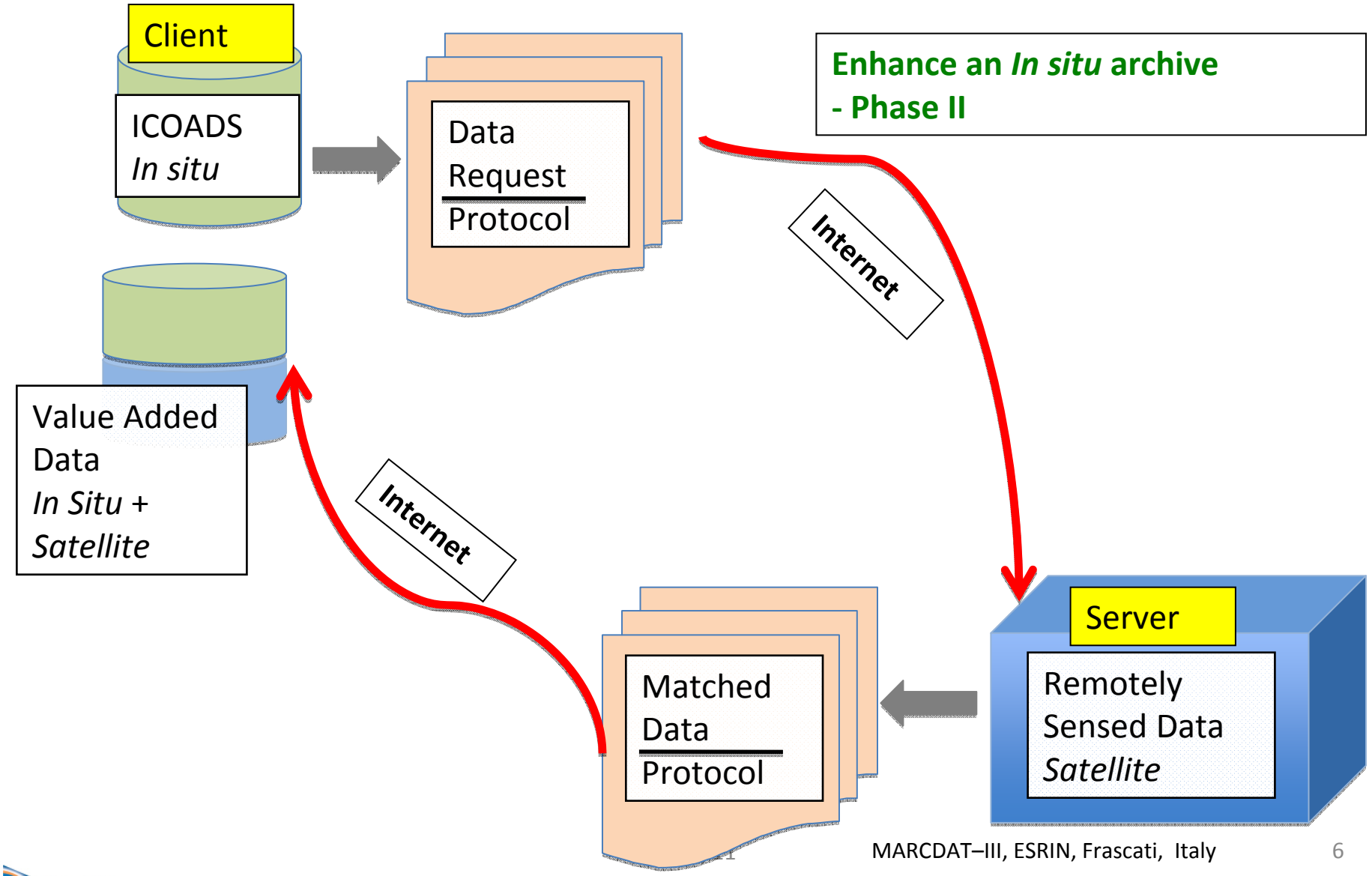
Data Management Benefits

- 24x7 (nearly) availability
- Most current version of data are always available
- Easy to serve a variety & many clients

Web Service and Use Cases

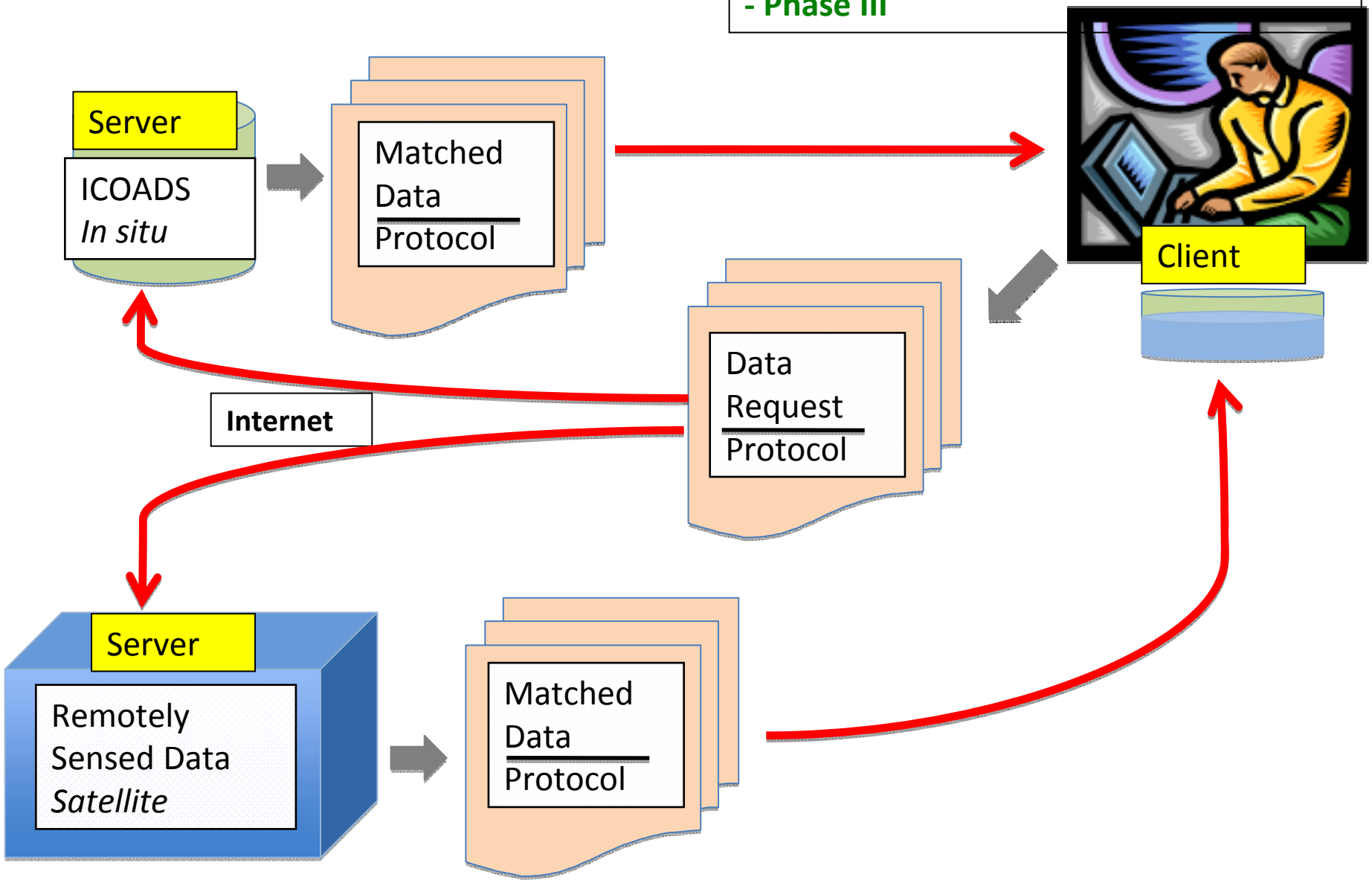


Web Service and Use Cases



Web Service and Use Cases

Enhance an individual's archive
- Phase III



Use Case Development

Phase I system development vision

1. Create an ICOADS Server at NCAR
2. Install Client software and create value-added datasets at PO.DAAC and/oranother agency
3. Tune and ensure system scaling
4. Open service worldwide

Infrastructure and Testing Results

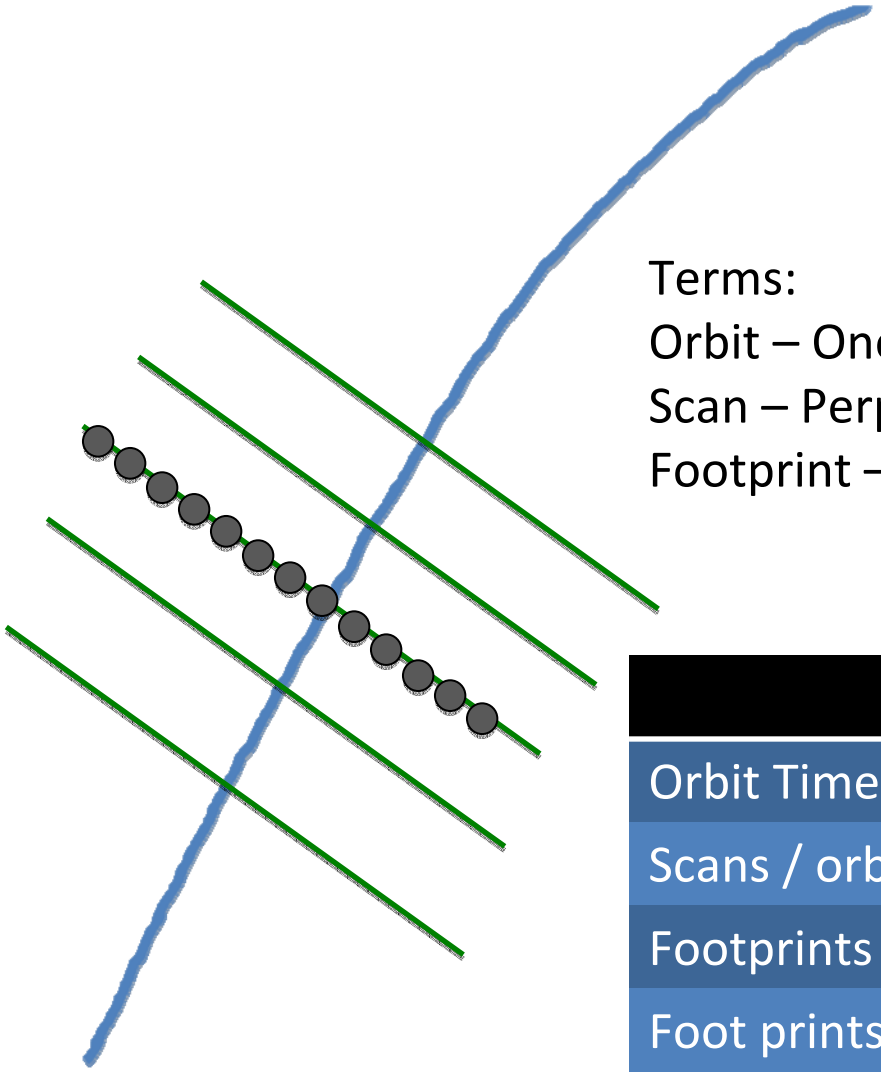
ICOADS prototype server

- MySQL DB, 300M+ IMMA records
  - R2.5 + 2008-April 2011, updated monthly
  - Hierarchical tables enable fast response for L2 data

SatelliteL2 data – orbit swath, from PO.DAAC

- AMSRE, SST
- ASCAT, Wind

Infrastructure and Testing Results



Terms:
Orbit – One Earth Revolution (blue)
Scan – Perpendicular Sampling Line (green)
Footprint – One Earth Surface Sample (grey)

	AMSRE	ASCAT
Orbit Time (min)	~ 95	~ 95
Scans / orbit	4193	3259
Footprints / scan	243	82
Foot prints / orbit	1018899	267238

Infrastructure and Testing Results

Challenges for Phase I

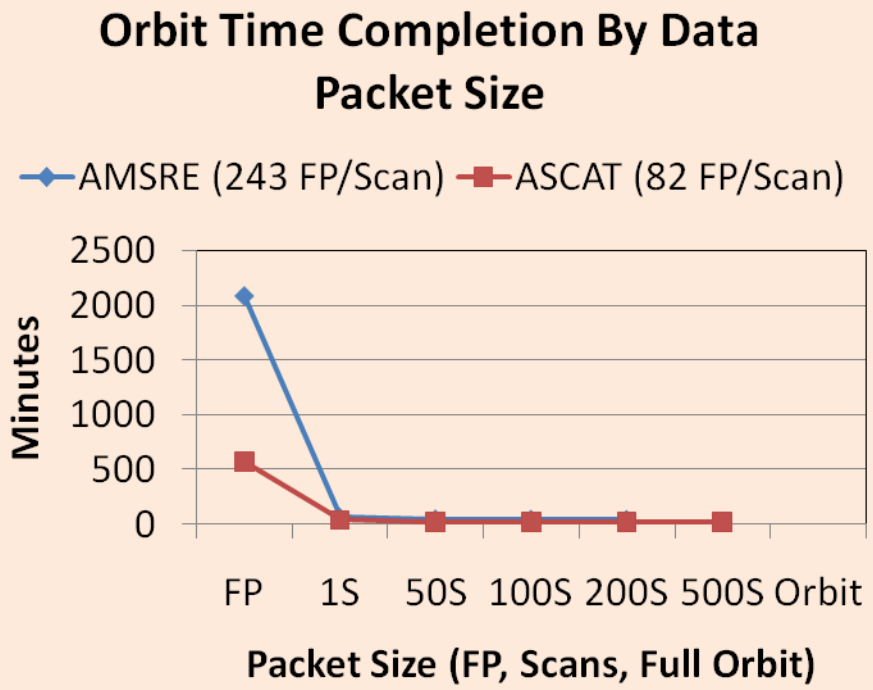
- Create value-added satellite datasets
- Fast Internet-based DB queries

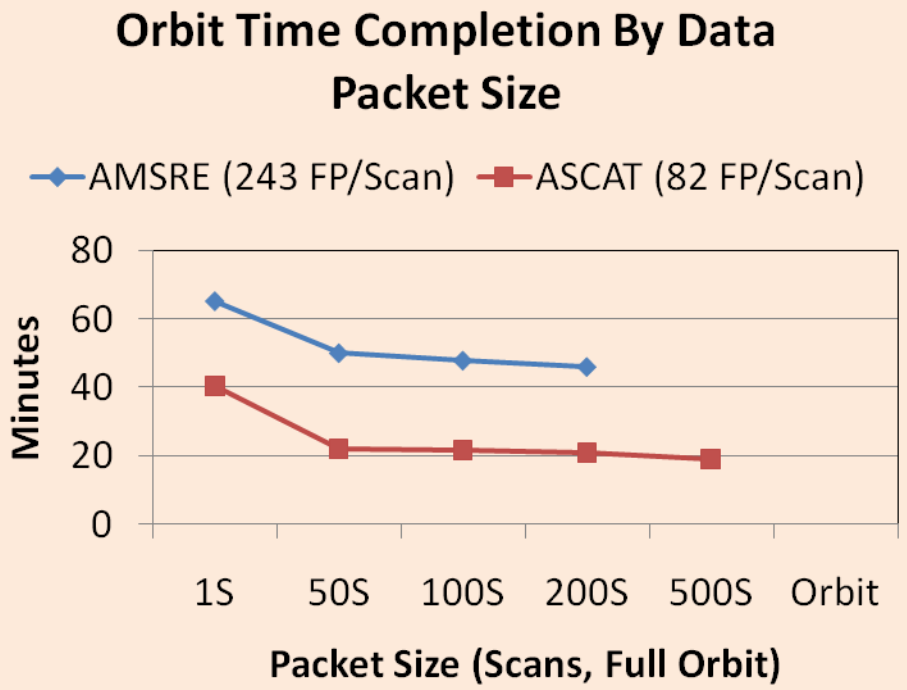
Influencing Factors

1. Data packet size
2. DB response speed
   - Impact of time and space ranges on queries
3. Internet transfer rate using 'wget' with HTTP protocol

Infrastructure and Testing Results

How much data to send in each packet?





1 Foot Print per Data Packet is <u>not viable</u>

- AMSRE 34 hours to complete an orbit
- ASCAT 10 hours

1 Scan (or more) per Data Packet <u>is viable</u>

- AMSRE 65-46 minutes
- ASCAT 40-19 minutes

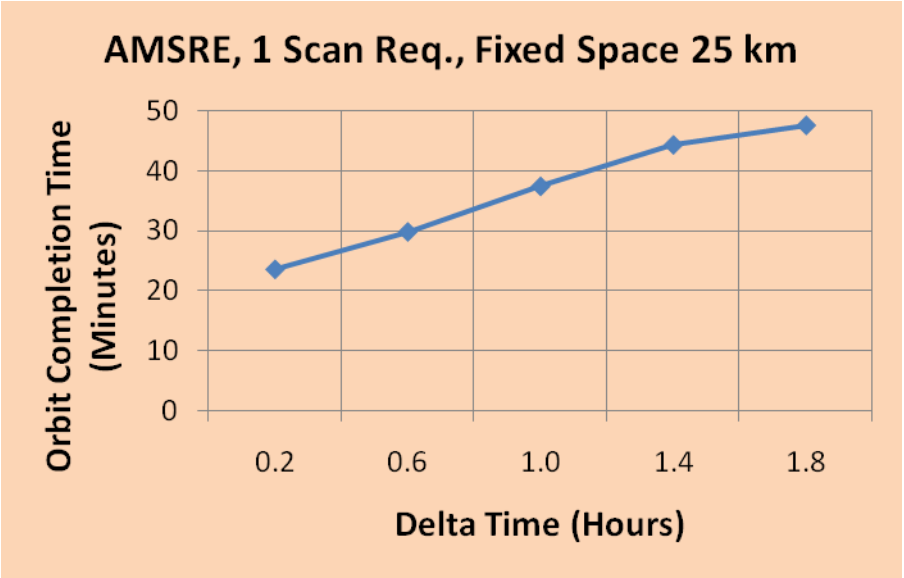
Hit upper limit > 200S per packet, AMSRE

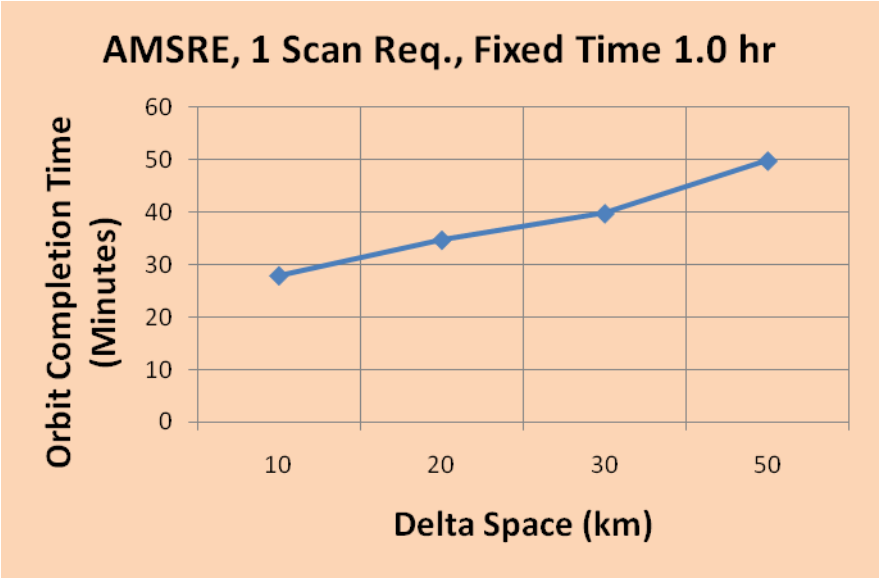
- Server time out (receipt and processing time too long)

Infrastructure and Testing Results

Time and Space Range Impact on DB performance

- Locally run – no Internet, LAN

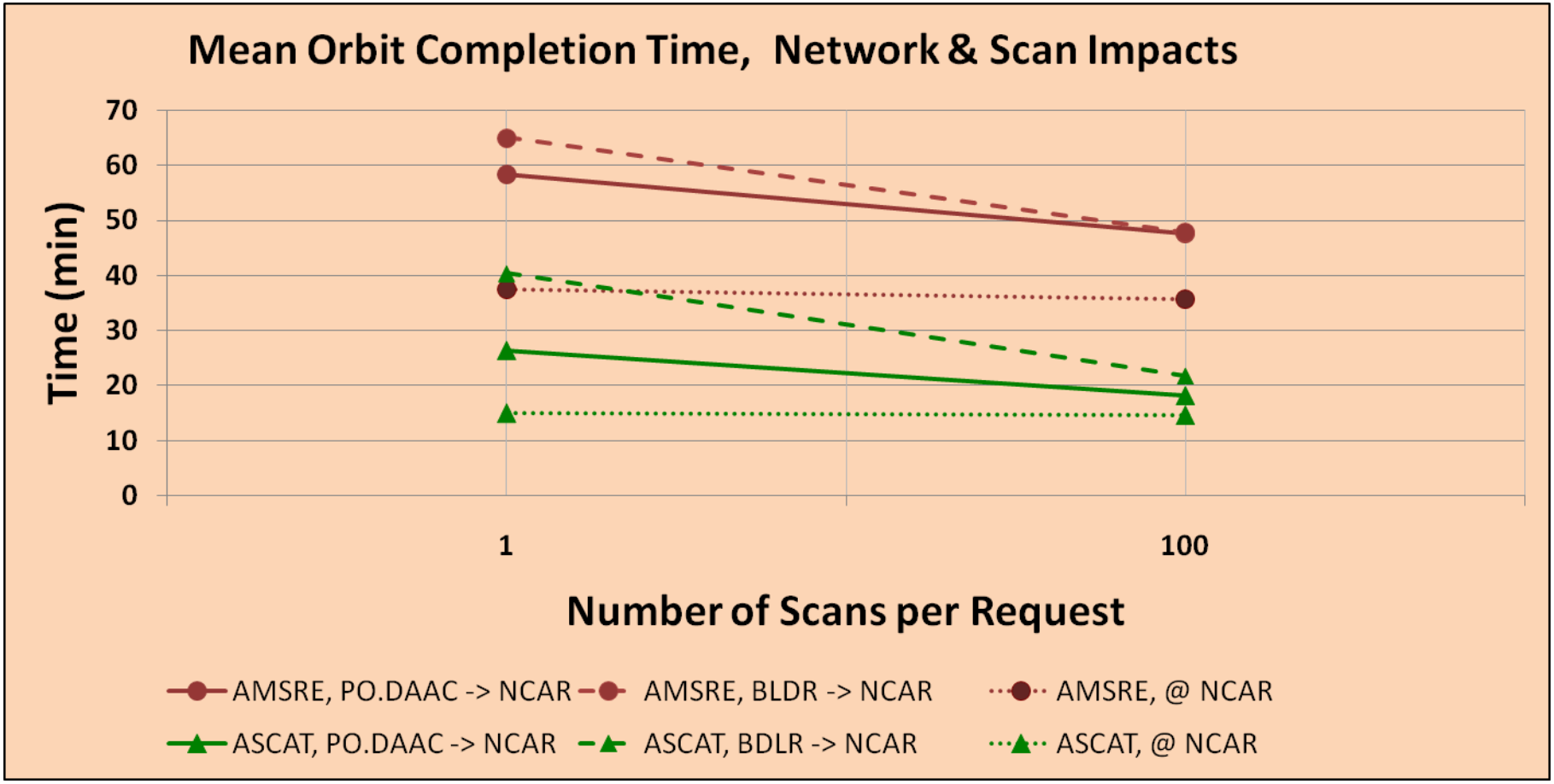




Time to completion can double as ranges grow

Space (km)	Time (hr)	Completion Time (min)
10	0.2	22
20	0.6	29
30	1.4	48
50	1.8	72

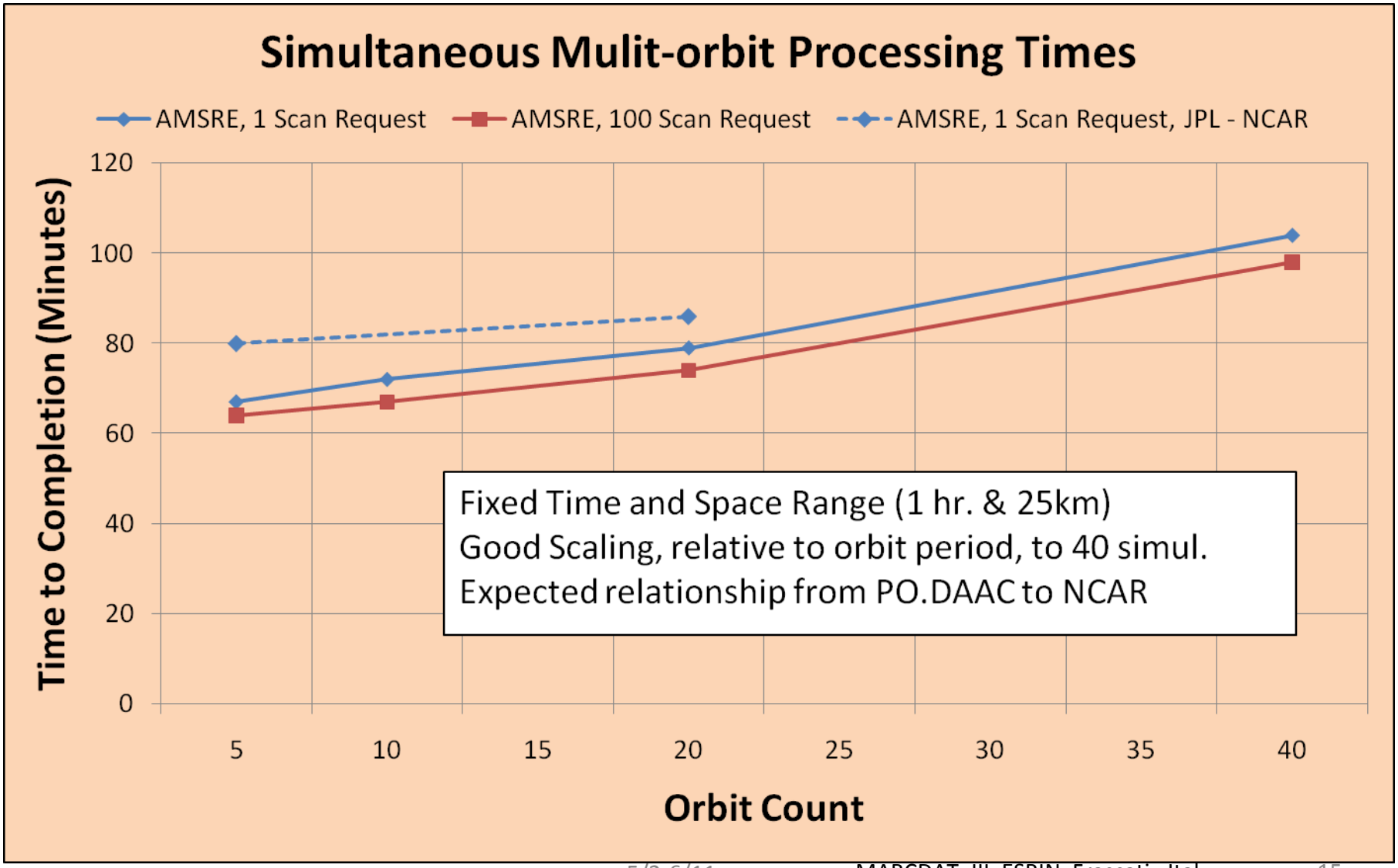
Infrastructure and Testing Results



- ASCAT faster than AMSRE – less data
- Number of scans per request – little impact at NCAR, LAN
- 100 Scans better than 1 for Internet, PO.DAAC to NCAR
  ~ 20% improvement
- Internet PO.DAAC to NCAR better than local Boulder

Infrastructure and Testing Results

Internet-based Tests



Infrastructure and Testing Results

How to manage performance?

- Plan to pair Perl client software with data process software at the client location
  - Via system calls
  - Client functions: form packets, send to server, receive results from server
- Use dedicated web and DB server
  - Optimized for the problem

Record Content for Web Service

Client provided data elements:

- Observed Satellite data
  - Date & time, latitude & longitude
  - Treat as point measurements
- Selectable specifications
  - Time range
  - Space range
  - Parameter to be matched
  - ICOADS trimming (standard, enhanced, none)
  - Code Indicators of additional variables (TBD)

    OR
  - Code to request IMMA Core + ICOADS attm – standard data

Record Content for Web Service

Server data returned to client

- All records, with the requested parameter, in time and space range
- Record content (TBD):
  - Resend client record for identification
  - Selected data fields from ICOADS IMMA

    OR
  - ICOADS Core and ICOADS attm
    - All the standard data fields
    - All the record provenance information (DCK, SID, PT, etc)

Next Steps

Write a white paper based these tests, ideas, and your comments & recommendations

Seek funding to support:

- Development, implementation, testing, and staffing at Server and Client locations
- Dedicated DB and web server
  - Expand capacity to handle more simultaneous orbits
- Design a user web interface for manually specified Client requests

Requests to the Community

- Should we develop this service?
- Are there technological or system requirements we have overlooked?
- Are there additional data requirements?
- Would you use such a service?
- Would you use datasets produced from this service?

Conclusions

- Demonstrated the potential to use a web service to bring *in situ* and remotely sensed data together
- The Phase I development would instantiate an ICOADS DB server and client software package
- Phase I would lead to a matched *in situ* and satellite archive dataset(s)
- There is more work to be done formulating the data to be exchanged