

## **Data Proliferation: STOP THAT**

**Ian D Norman, UK Ministry of Defence, Washington DC**  
**Rod Roderique, Department of Defense, Washington DC**  
**E-mail: k3qii@erols.com**

**Presented at the THIC Meeting at the National Center  
Center for Atmospheric Research**

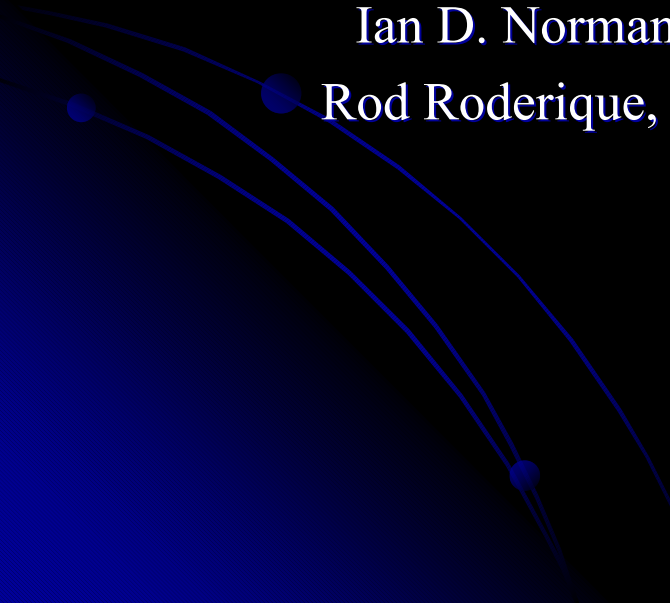
**Boulder CO 80305-5602**

**June 11-12, 2002**

# Data Proliferation STOP THAT

Ian D. Norman, Ministry of Defence, Washington D.C.

Rod Roderique, Department of Defense, Washington D.C.



# Preface

- This presentation *Data Proliferation, STOP THAT* was prepared for the THIC Technical Symposium, Theme: *Towards PetaByte Archives* Boulder, Colorado; 11-12 June 2002.
- The authors wish to thank THIC for providing a forum where technical ideas can be exchanged.

# Why are PetaBytes of Storage Needed?

- The Information Age (Data Explosion)
  - Internet Access Speed Improvements
    - DSL / T1 / Cable Modem
  - Personal Storage Affordable
    - Hard Drive = 80 GByte, 8 msec access, \$105
  - Enterprise Structures Access Data Faster
    - Storage Area Networks (SAN)
    - Full-Fabric Fibre-Channel, 2-GBytes/sec

# Performance Improvement Ideas

- **Technology vs. Discipline**
- **Reduce Data Movement**
- **Improve MetaData Structures**
- **Improve File Transfer Structures**

# Technology vs. Discipline

- Technology

- Storage

- More Density
- Faster Access
- Affordable

- Networks

- Modern Architecture
- Faster Access

- Discipline

- File Management

- Minimize Replication

- Personal

- Trust in the Process
- Move Only What You Need

# Reduce Data Movement

- **Install “processing applications” where the data resides and “display results” (remote processing).**
  - **Applications need to better utilize technology (smarter applications software).**
  - **Don’t move it / Use it.**

# Improved MetaData Structures

- MetaData need not only describe “envelope information”.
  - **Recognize file size, provide a file index to desired portion of file, by size (file partition; e.g., 1/10, 1/20, 1/50, etc.).**
  - **Recognize file content, provide a file index to desired portion of file, by content.**



# Improved File Transfer Structures

- **Permit user to start using “first data” prior to “end of file marker”.**
  - **Recognize file size, smartly subdivide file into smaller files.**
- **Incorporate “loss less data compression” techniques during file transfer.**
  - **Storage access speed and processing speed are now faster than network speed, compress and un-compress as part of the file transfer process.**

# Summary

- Storage technology alone is not the solution.
- Applications must better utilize modern technology.
- Reduction of data movement is needed.
- Improved metadata structures are needed.
- Improved file storage/transfer structures are needed.
- User education/discipline is needed.