



# Grid Computing for Integrative and Translational Cancer Research

Tahsin Kurc

Center for Comprehensive Informatics



# Roadmap

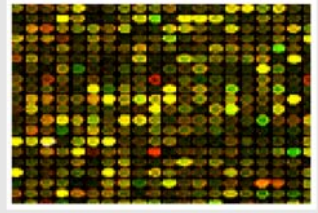
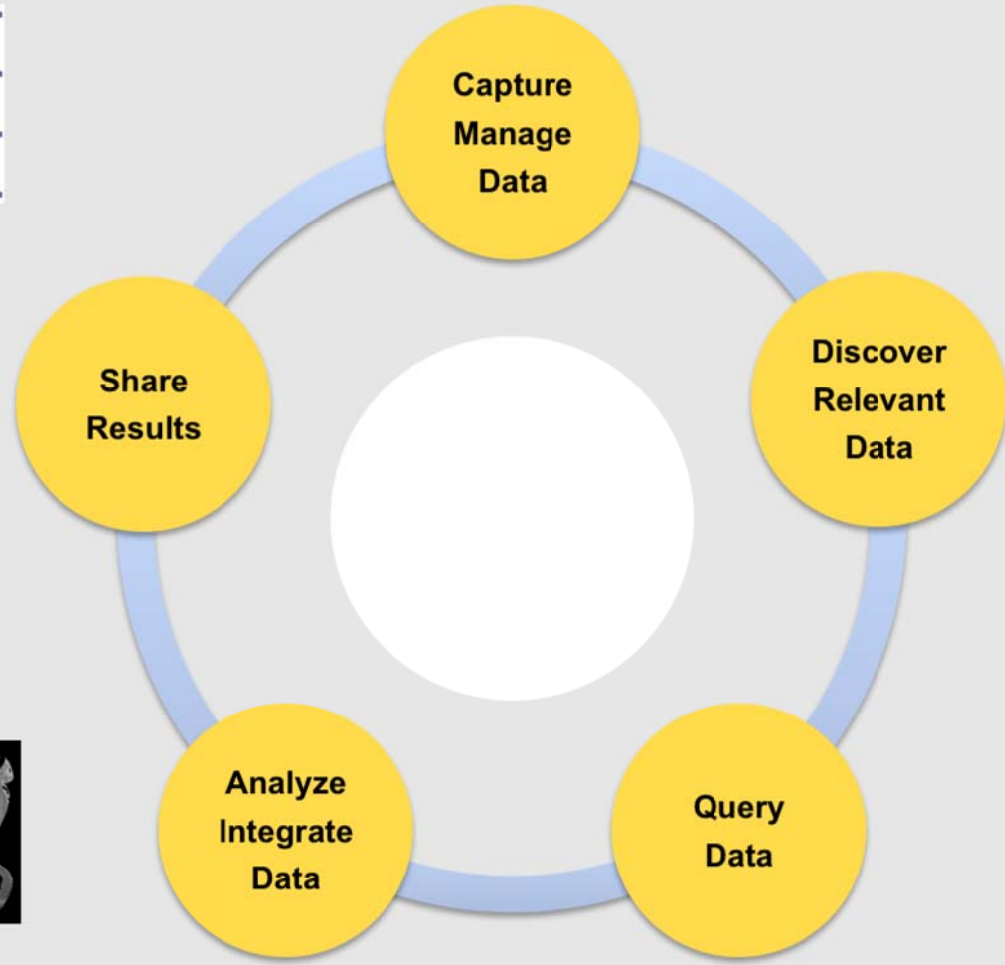
- Example research scenario
- Challenges in multi-institutional settings
- Software system support



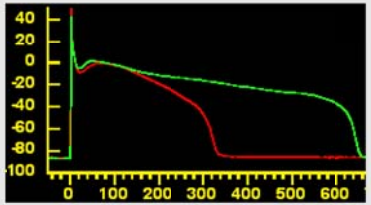
# Application/Research Scenario

- A clinician/researcher is involved in a multi-institutional clinical trial
  - Microarray, Genomic, Proteomic, and Image data are collected from patients participating in the trial at different institutions
  - Researcher wants to carry out a correlative analysis to assess the treatment
    - Query and analyze microarray, image, and protein data from multiple patients to find interesting patterns
    - Look for similar patterns in other microarray, protein, and image databases
- A clinician/researcher wants to compile a view of a patient
  - Patient may have been treated at multiple institutions
  - Lab data, image and genomic data, diagnosis information may have been collected at different institutions

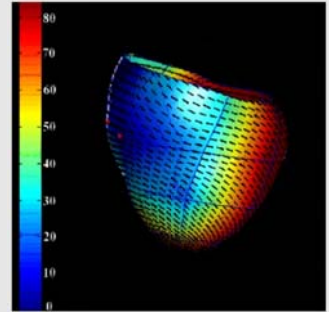
# Application/Research Scenario



Gene Expression Profiling



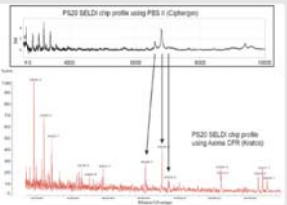
Lab Measurements



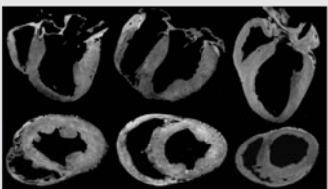
System Biology Modeling



Genetic Variability



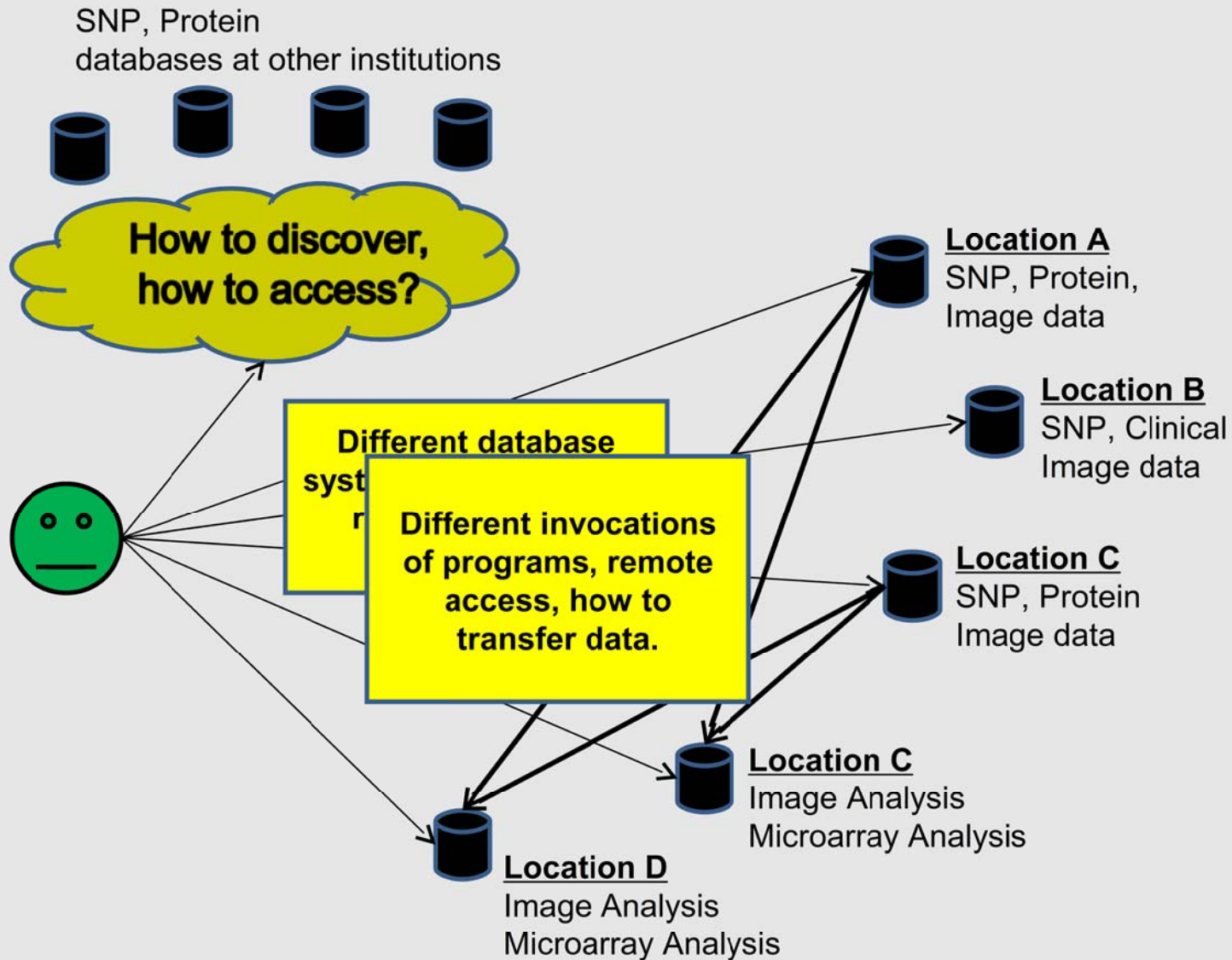
Protein Expression Profiling



Multi-Modal Imaging



# Application/Research Scenario





# Problems and Challenges

- “Legacy” systems, most of which cannot be readily and remotely accessed and shared between institutions
  - Limited application of common data structures and common vocabularies
- Difficulty in identifying and accessing available resources
- Limited infrastructure to securely share data among different institutions
- Large volume of data must be queried, transferred, and processed
- Increased complexity in larger scale environments



# A Software Architecture Approach

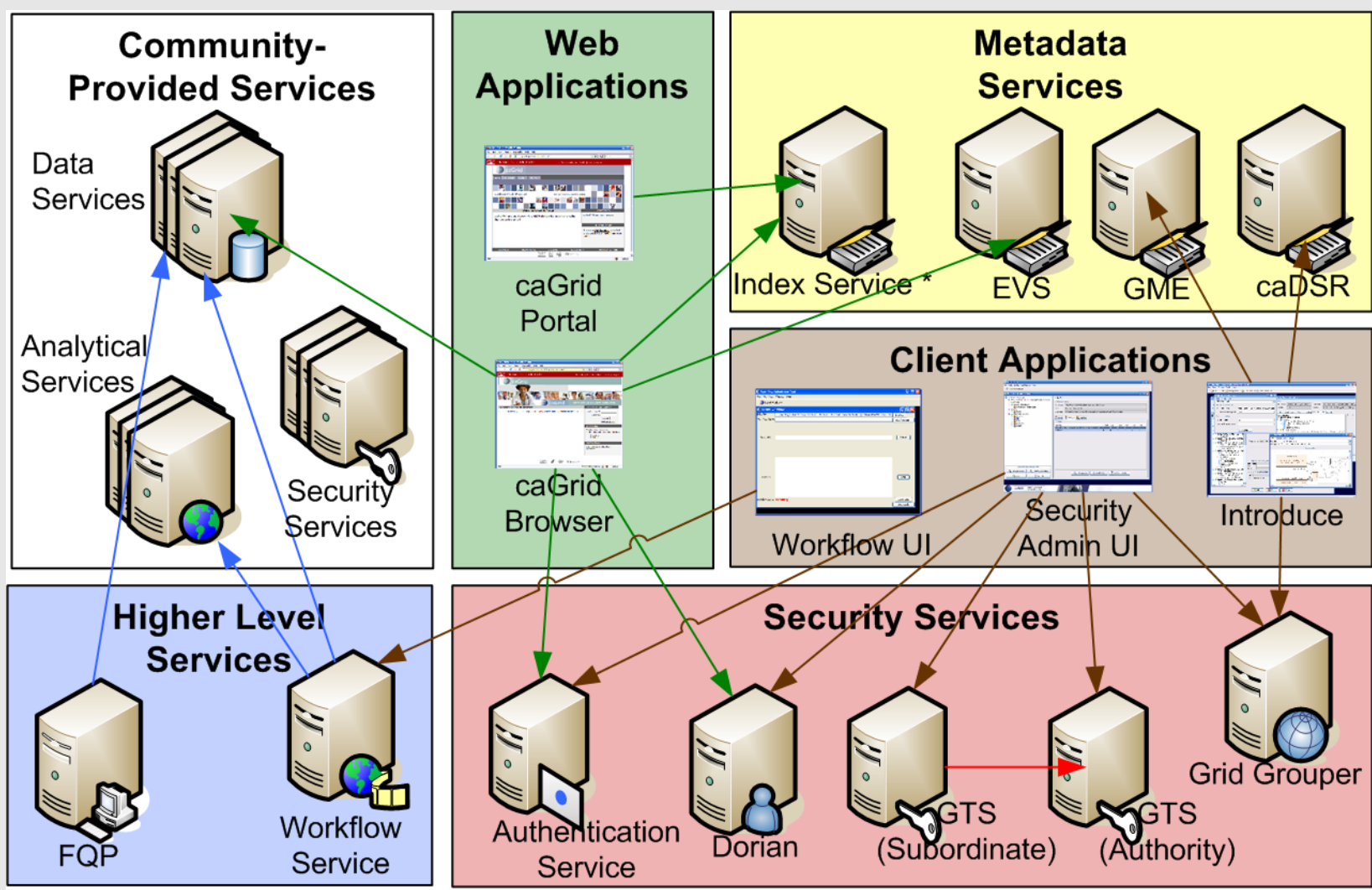
- Model Driven Architecture
  - Shared, harmonized set of data elements, controlled vocabularies, and published data models that facilitate information exchange
- Service Oriented Architecture
  - Resources and applications exposed as services
- Collection of interoperable applications developed to common standards
- A common Grid infrastructure that links applications and resources and enables enforcement of security policies
- Cancer Biomedical Informatics Grid
- CardioVascular Research Grid



# caGrid Infrastructure

- A Grid software middleware infrastructure consisting of services, toolkits, APIs, and runtime environment
  - Standards Based, Open Source
  - Building blocks to create interoperable, Grid-enabled systems
- Service Oriented Architecture
  - Web Services Resource Framework standards
- Model Driven Architecture
  - Object oriented view, published information models, strongly-typed services
  - Rich metadata
- A production Grid deployment of the core services provided by that infrastructure
  - Security, Workflow, Data Services Infrastructure, Service Development/Deployment, Metadata/Advertisement/Discovery

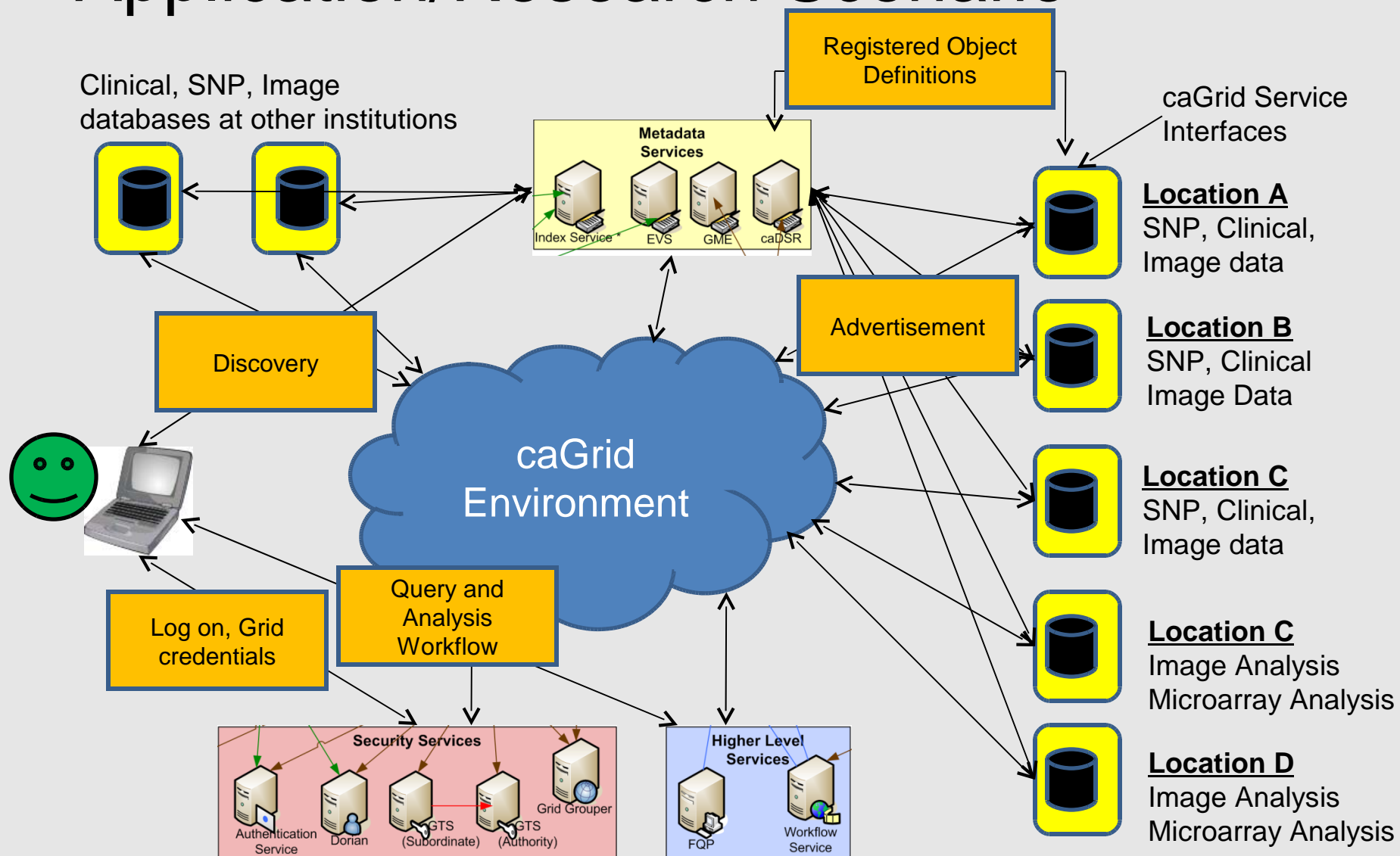
# caGrid Infrastructure



\*All Services Register with the Index Service



# Application/Research Scenario





# More Information

- Emory Center for Comprehensive Informatics
  - <http://www.cci.emory.edu>
- cancer Biomedical Informatics Grid
  - <https://cabig.nci.nih.gov/>
- caGrid Infrastructure
  - <http://cagrid.org>
  - <https://cabig-kc.nci.nih.gov/CaGrid/KC>

## Funding

- The National Cancer Institute, National Institutes of Health under Contract No. HHSN261200800001E; by the NCI through Contracts 28X193, N01-CO-12400, 94995NBS23, and 85983CBS43; by Grant R01LM009239 from the National Library of Medicine; by Grant R24HL085343 from the National Heart, Lung, and Blood Institute; and by PHS Grant UL1RR025008 from the Clinical and Translational Science Award Program, NIH.
- The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.



Thanks!