

## Data Sets and Data Resources

Malcolm Atkinson

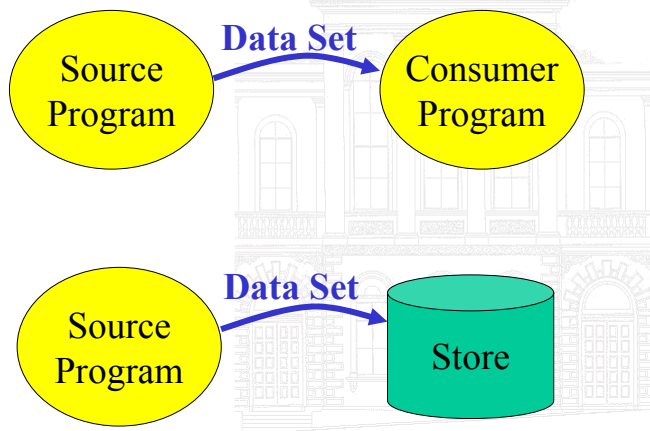
Director of National e-Science Centre  
[www.nesc.ac.uk](http://www.nesc.ac.uk)

25<sup>th</sup> June 2003

GGF8 Seattle  
DAIS WG Second WG Meeting at GGF8

## Generic Idea

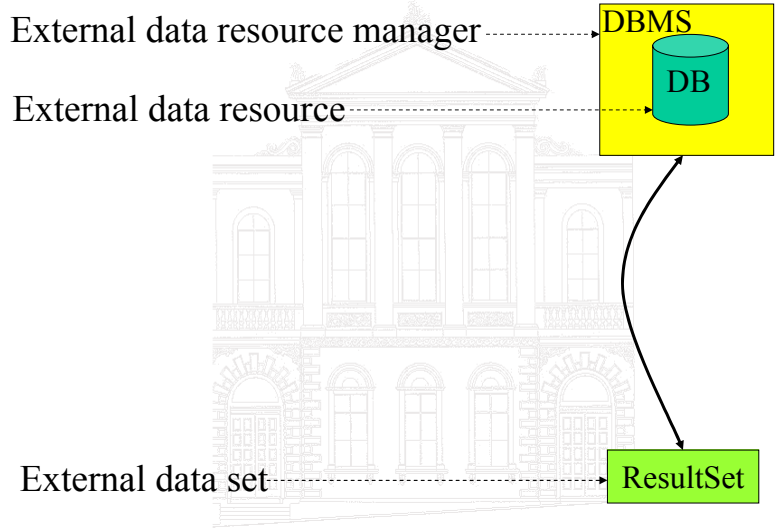
- Abstract / Physical Value Passed or Stored

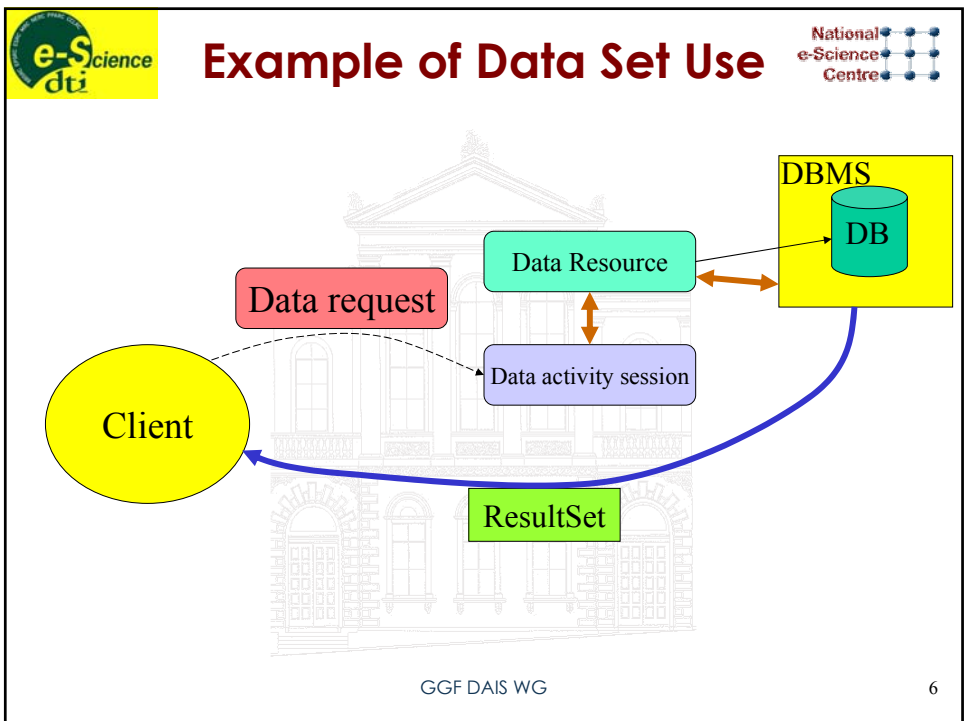
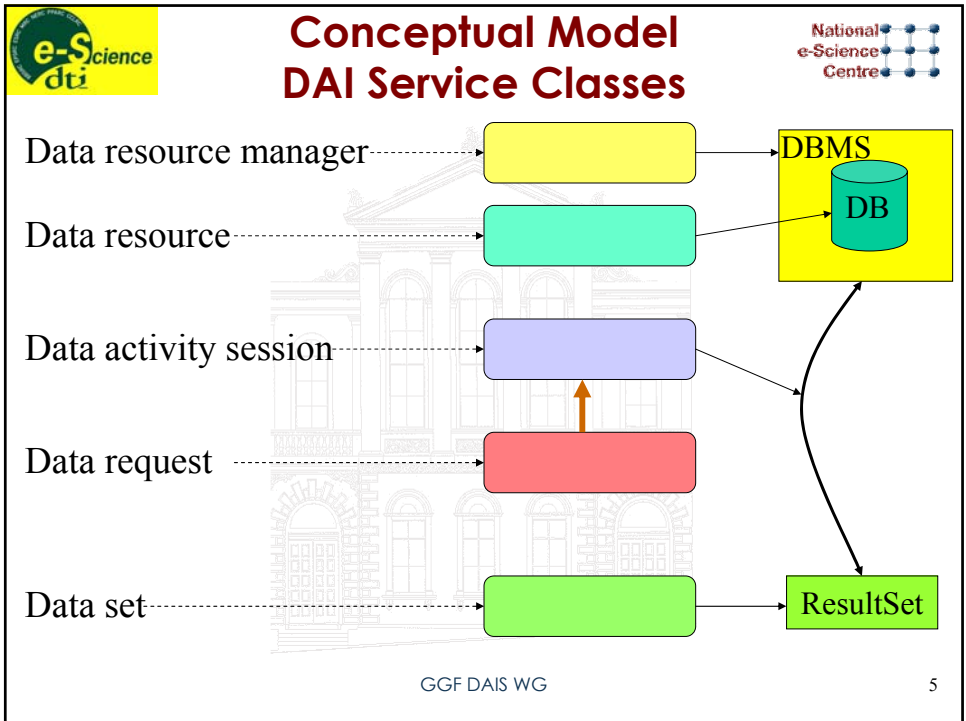


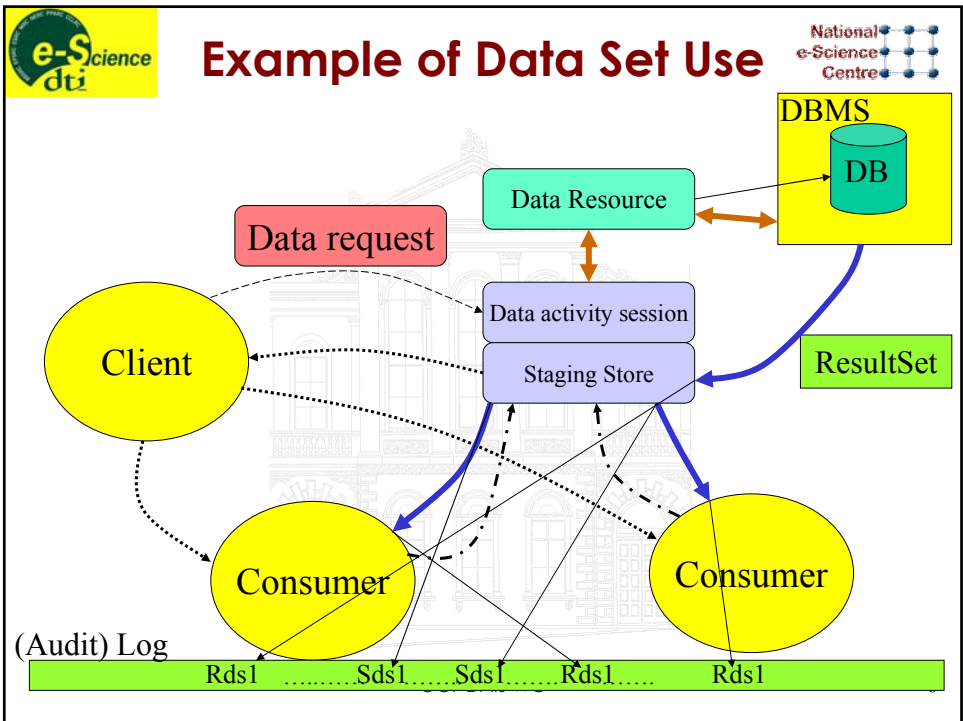
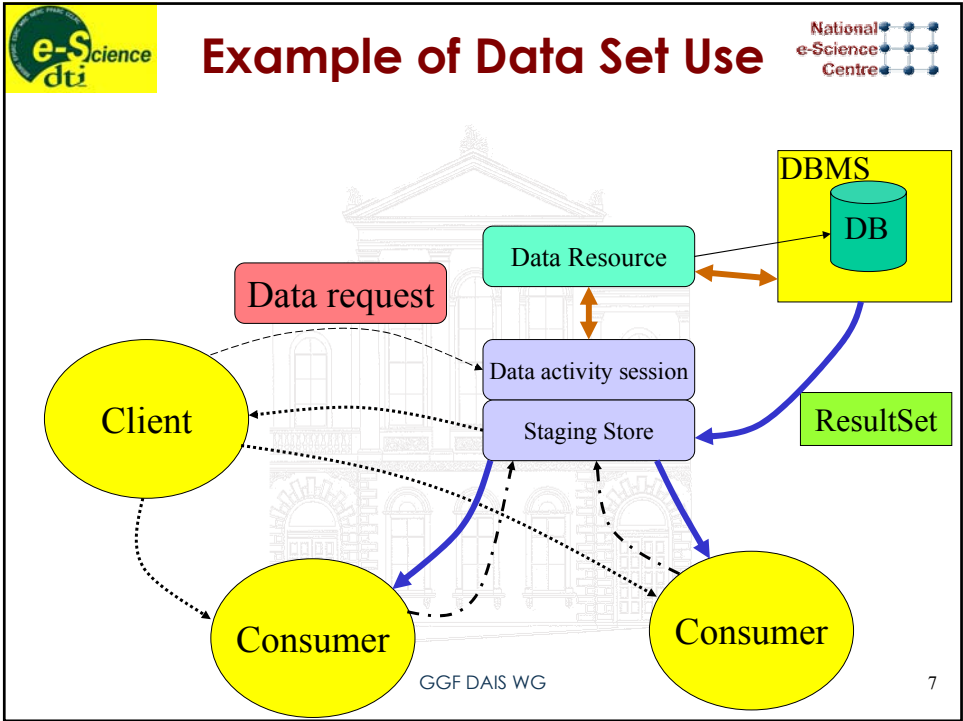
# Questions

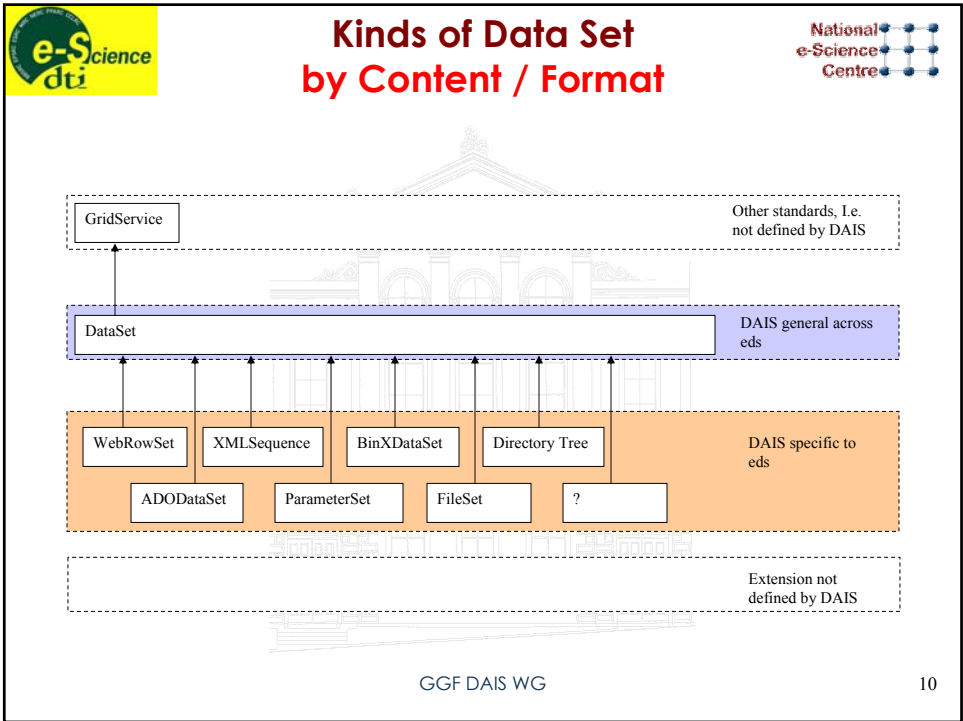
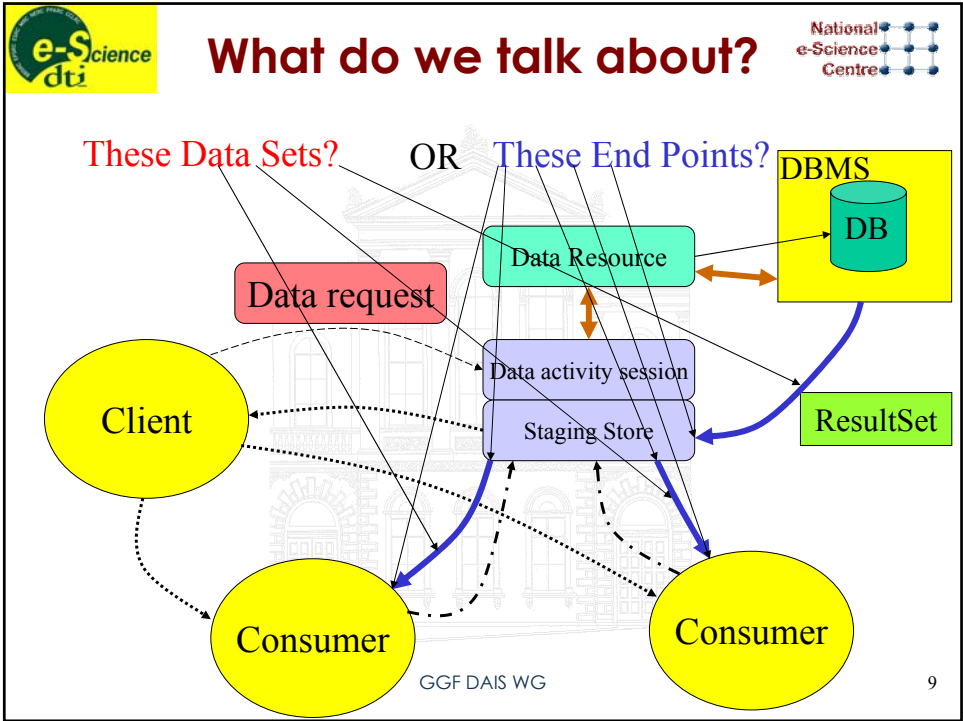
- **Do we need the concept?**
  - **Once Identified remains Identified**
    - ▶ Where ever copied, stored, moved
    - ▶ Value (information) remains constant
    - ▶ Type(s) remain constant
    - ▶ Basis for reference, e.g. parameters pass by reference
- **If yes, is it the right name?**
  - ▶ Data Set
  - ▶ Value Set
- **Is it a Grid Service? If not, what is it?**
- **What do we talk about?**
  - The Data Set or the End Point
- **What are its properties?**
  - Kinds of data set
  - Operations on the whole
  - Creation & access
  - Verification of Identity / Constancy
- **Relationship with Data Resource?**
  - Same, Special case of or Different

# Conceptual Model External Universe









## Kinds of Data Set by Behaviour / Usage

- **Delivery**
  - Push as one Unit
  - Push incrementally
  - Push at most once
  - Push to subscribers
  - Hold until get all
  - Hold until get fragments
  - Hold until any get
- **Materialisation**
  - Materialise on Demand
  - Materialise Eagerly
  - Materialise in Parallel
- **Connection Method**
  - Publish and subscribe
  - Pass handle via client
  - Declarative, ...
- **Longevity**
  - Until eol
  - Until eol or destroy
  - Until used n times
- **Security**
  - Encryption method
- **Unit of Access (items)**
  - Multiples of Row
  - Multiples of Children
  - ...
- **Many other properties**
  - Cursor available
  - Reset available
  - Compression

GGF DAIS WG

11

## Relationship with Data Resource

- **Data Set**
  - No update
  - Copy & Move
  - Retains persistent identity (GUID)
    - ▶ DS description
  - Retains value
  - Retains type
  - Simple(?) access repertoire
  - Lifetime bound to physical eds
  - Minimal SDEs
- **Data Resource**
  - Rich Repertoire of Access & Update Methods
  - Value changes
  - Type (may) Change
  - Has a defined drm
  - Located in edr
  - Lifetime independent of edr
  - Complex SDEs

GGF DAIS WG

12

# Data Movement

Currently Delivery Specification Functions on a das

# Motivation & QoS Matters

- **Results of no value**
    - If not delivered to a consumer
  - **Updates**
    - Need supplied values
  - **These may be large!**
  - **Supplier and Consumer(s)**
    - Same process or different processes
    - Same machine or same cluster
    - Same LAN or Wide Area hop
    - Same architecture, hosting environment or different
- 
- **Are messages (notifications) transmitted from triggers always small? Treated differently?**
  - **Are they in or out of scope?**

## Requirements

- **Information Preservation**
  - A value transmitted arrives as the same value
  - For all values of all types
- **Composability**
  - Any Grid (Data) Service with
  - Any Data Movement Mechanism
- **Performance & Costs**
  - Choices to be supported?

## Requirements 2

- **Integration**
  - Available between all services
  - Available for all Information
    - Data, Code, Processes, State
  - DAIS-WG current focus is Information
    - Higher-level view of data
- **Control**
  - What is moved (Query result | DB | File | set of (files | results), ...)
  - Where From & To\* (Notation in Request document?)
  - What units of transfer (Logical | Physical)
  - What synchronisation (and lifetime)
  - Privacy, e.g. Encryption
  - Dynamic selectivity & Flow control
  - What compression / encoding
  - What transformations en route
  - What monitoring mechanisms



## Fundamental Issues

- **What portType(s)?**
- **What is (are) their functions?**
- **What semantics of the functions?**
  - **Effects on data**
  - **Effects on GD\* state**
  - **Reflection in SDEs**
  - **Failure modes**

