

# Organizing and accessing the multiple data streams

Floor van Leeuwen

Institute of Astronomy  
Cambridge

# This presentation

- Exploring some concepts of accessing the multitude of data streams
  - Not just Gaia, also other streams
  - Building on experience
    - Gaia photometric reductions QC system
    - Hipparcos data exploration
- Short demo on some of the possibilities
  - Using the Hipparcos data exploration as example



# Aim of the presentation

- To be seen in the context of the ASTRA proposal
  - First steps towards development of a system that provides a dynamic access to the combined data streams
  - Making data available as needed
  - Providing various global overviews of the available data
  - Allowing direct access to details on small scales

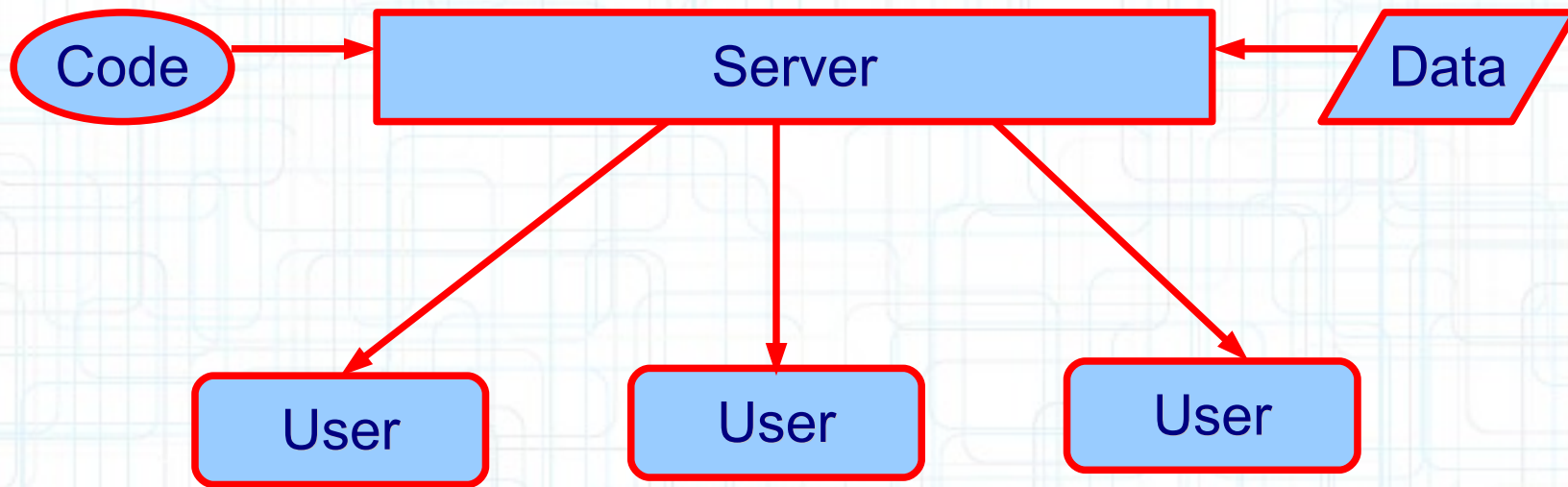
# Structure

- One or more servers providing access to the available data as requested
  - Access is through internet connection
- Resource tree
  - Managing the different layers of the available data and their connectivity
- One or more Java programs
  - Provide the GUI and display facilities for the data, using the resource tree and server for access



# Details

- Gaia Wiki pages, CU5 → meetings
  - CU5M11 → Agenda
    - Video presentation by A.Yoldas on the server interface
    - Presentation by F.Suess on the resource tree
    - Presentation by F.van Leeuwen on GUI



# The user

- The user logs in to the server
  - This provides a “key” for all further communications with the server
  - Download the required .jar file
  - Start interaction
- Based on the requests from the user, data is downloaded
  - The resource tree structure allows for only data that is needed to be downloaded



# The server

- Master server
  - Any updates to the software and the the data go to the master server
  - Other servers update from the master server
    - Could use SVN or equivalent for the software
- Data base provides reference to data files
  - Data base may be structured around HealPix concept
  - Could use Hadoop distributed file system



# Data architecture & services - resources

- Software architecture is described in terms of a set of layered resources
- Web service design relies on resource-based approach
- A resource can describe almost anything: operations as well as data objects
- A resource has a single, abstract definition which may include:
  - data fields: primitive data types, strings, enumerations, arrays
  - resource references
  - a set of functional constraints
- A resource can have more than one representation, for example:
  - Java class definition: QC client & web-service resources
  - Web file format: XML, JSON, etc.
  - QC server entities: JAXB, JPA
  - Resource serialisation: binary file format
  - Database representation
- Each representation supports all of the features of the resource definition
- A representation may support additional features

# The interface

- The interface is there to provide easy and measured data access facilities
  - In addition, it provides a set of tools to interact dynamically with the data
    - Use the information as obtained from a display to prepare a next display, or to extract selected data records
    - Explore the data in hierarchical form
    - Preserve graphs as obtained in the investigation

# Possibilities for ASTRA

- From Gaia:
  - 3D maps of stellar distribution
  - 3D velocity vectors
  - Reddening as function of distance
    - Up to several kpc
- From Planck:
  - Dust as integrated along the line-of-sight
  - Possible dust-temperature variations



# The demo

- Uses the Hipparcos (FvL, 2007) data
  - Astrometry and broad-band photometry
- Additional photometric data
  - Geneva, Walraven, Stromgren
- External information on orbits
- Java program using JFreeChart and Java Swing components
  - Development done in NetBeans IDE

# Organizing and accessing the multiple data streams

Floor van Leeuwen

Institute of Astronomy  
Cambridge

19 June 2012

Heidelberg

1

# This presentation

- Exploring some concepts of accessing the multitude of data streams
  - Not just Gaia, also other streams
  - Building on experience
    - Gaia photometric reductions QC system
    - Hipparcos data exploration
- Short demo on some of the possibilities
  - Using the Hipparcos data exploration as example



# Aim of the presentation

- To be seen in the context of the ASTRA proposal
  - First steps towards development of a system that provides a dynamic access to the combined data streams
  - Making data available as needed
  - Providing various global overviews of the available data
  - Allowing direct access to details on small scales

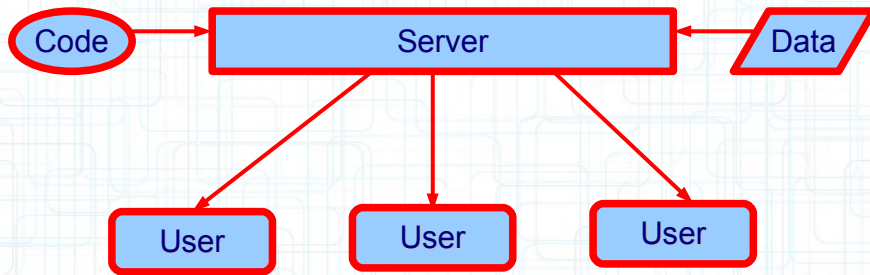
# Structure

- One or more servers providing access to the available data as requested
  - Access is through internet connection
- Resource tree
  - Managing the different layers of the available data and their connectivity
- One or more Java programs
  - Provide the GUI and display facilities for the data, using the resource tree and server for access

## Details

- Gaia Wiki pages, CU5 → meetings
  - CU5M11 → Agenda
    - Video presentation by A.Yoldas on the server interface
    - Presentation by F.Suess on the resource tree
    - Presentation by F.van Leeuwen on GUI





# The user

- The user logs in to the server
  - This provides a “key” for all further communications with the server
  - Download the required .jar file
  - Start interaction
- Based on the requests from the user, data is downloaded
  - The resource tree structure allows for only data that is needed to be downloaded

# The server

- **Master server**
  - Any updates to the software and the the data go to the master server
  - Other servers update from the master server
    - Could use SVN or equivalent for the software
- **Data base provides reference to data files**
  - Data base may be structured around HealPix concept
  - Could use Hadoop distributed file system



## Data architecture & services - resources

- Software architecture is described in terms of a set of layered resources
- Web service design relies on resource-based approach
- A resource can describe almost anything: operations as well as data objects
- A resource has a single, abstract definition which may include:
  - data fields: primitive data types, strings, enumerations, arrays
  - resource references
  - a set of functional constraints
- A resource can have more than one representation, for example:
  - Java class definition: QC client & web-service resources
  - Web file format: XML, JSON, etc.
  - QC server entities: JAXB, JPA
  - Resource serialisation: binary file format
  - Database representation
- Each representation supports all of the features of the resource definition
- A representation may support additional features

# The interface

- The interface is there to provide easy and measured data access facilities
  - In addition, it provides a set of tools to interact dynamically with the data
    - Use the information as obtained from a display to prepare a next display, or to extract selected data records
    - Explore the data in hierarchical form
    - Preserve graphs as obtained in the investigation

# Possibilities for ASTRA

- From Gaia:
  - 3D maps of stellar distribution
  - 3D velocity vectors
  - Reddening as function of distance
    - Up to several kpc
- From Planck:
  - Dust as integrated along the line-of-sight
  - Possible dust-temperature variations



## The demo

- Uses the Hipparcos (FvL, 2007) data
  - Astrometry and broad-band photometry
- Additional photometric data
  - Geneva, Walraven, Stromgren
- External information on orbits
- Java program using JFreeChart and Java Swing components
  - Development done in NetBeans IDE