

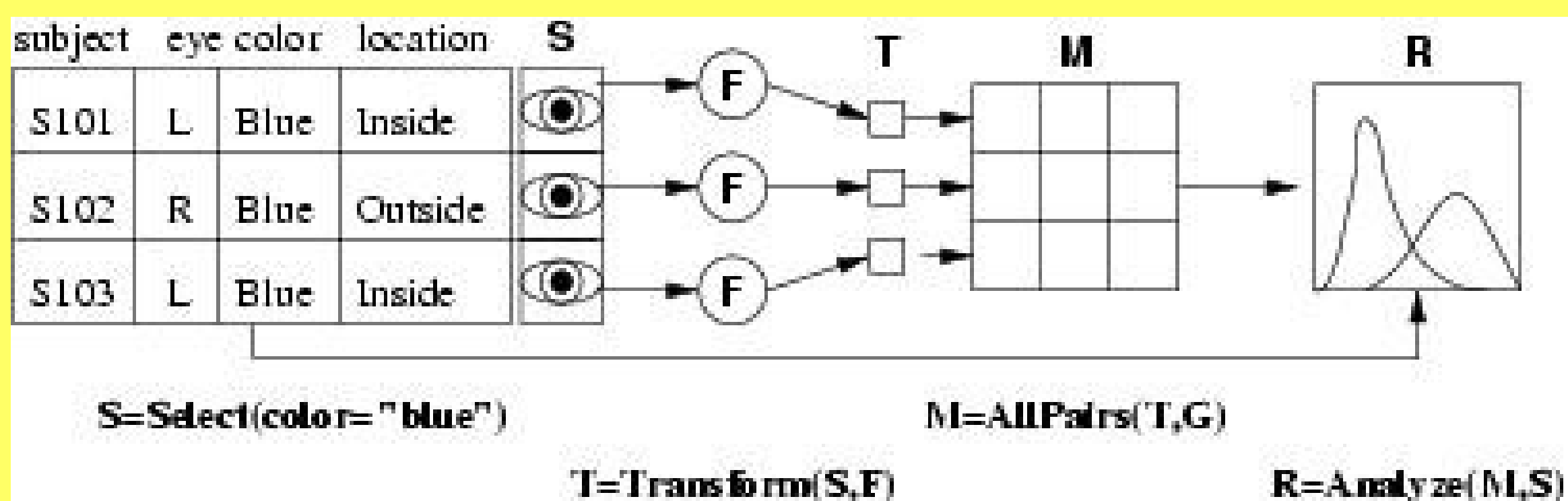
BXGrid: A Data Repository and Workflow Abstraction for Biometrics Research

Hoang Bui, Deborah Thomas, Michael Kelly, Christopher Lyon, Douglas Thain, and Patrick J. Flynn

Biometrics research is often conducted on millions of records which account for terabytes of digital data. The quality of experimental result is highly depend on the effective management and analysis of those data. We have designed and implemented BXGrid, a data repository and workflow abstraction for biometrics research. The system is composed of a relational database, an active storage cluster, and a campus computing grid. End users interact with the system through a high level abstraction of four stages: Select, Transform, AllPairs, and Analyze. A high degree of availability and reliability is achieved through transparent fail over, three phase operations, and independent auditing.

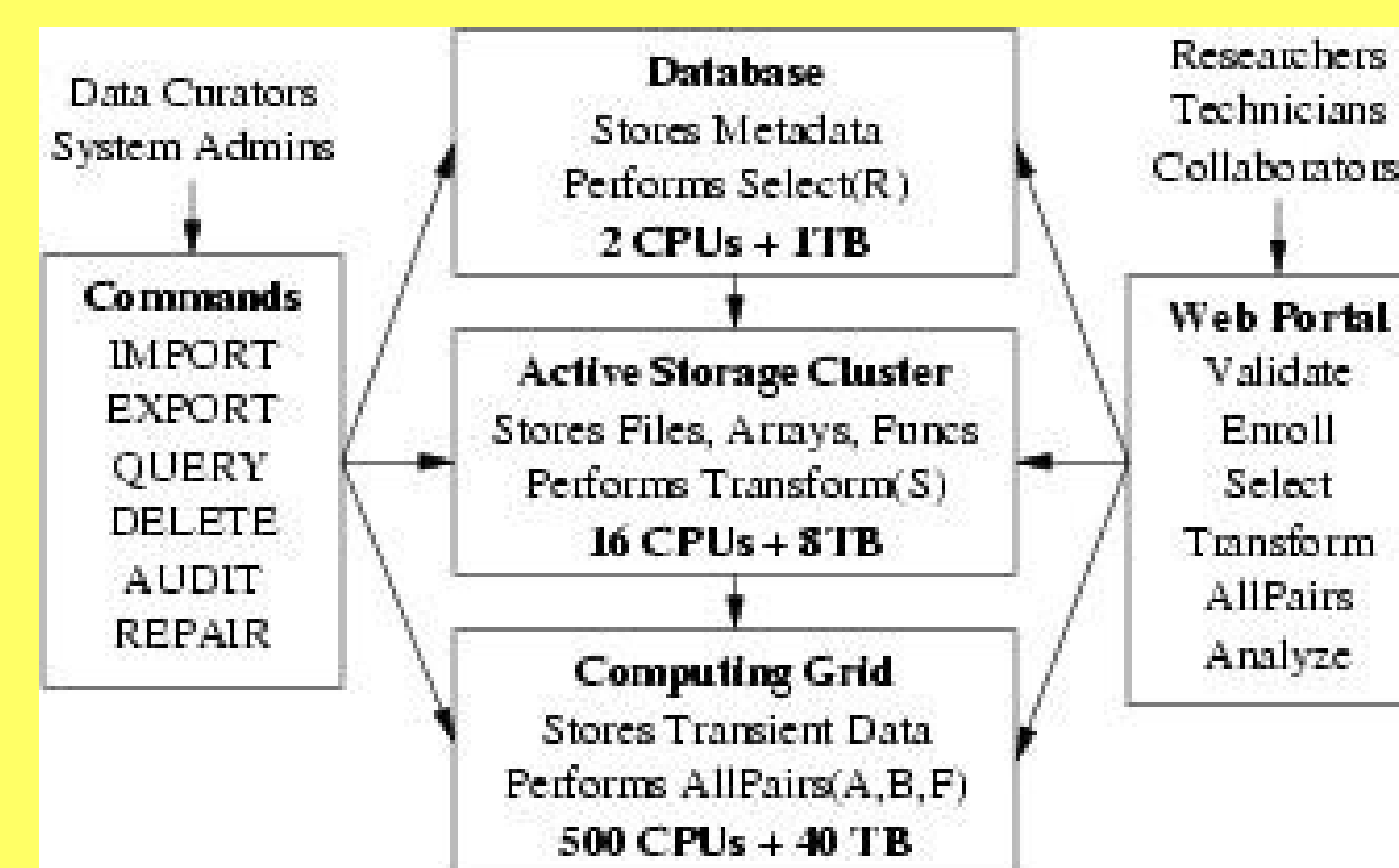
Biometrics experimental abstractions

- **Select(R)** = Select a set of images and metadata from the repository based on requirements R.
- **Transform(S, F)** = Apply function F to all members of set S, yielding the output of F attached to the same metadata as the input.
- **AllPairs(S, F)** = Compare all elements in set S using function, producing a matrix M where each element $M[x][y] = F(S[x], S[y])$.
- **Quality(M, D)** = Reduce matrix M into a metric D that represents the overall quality of the match.

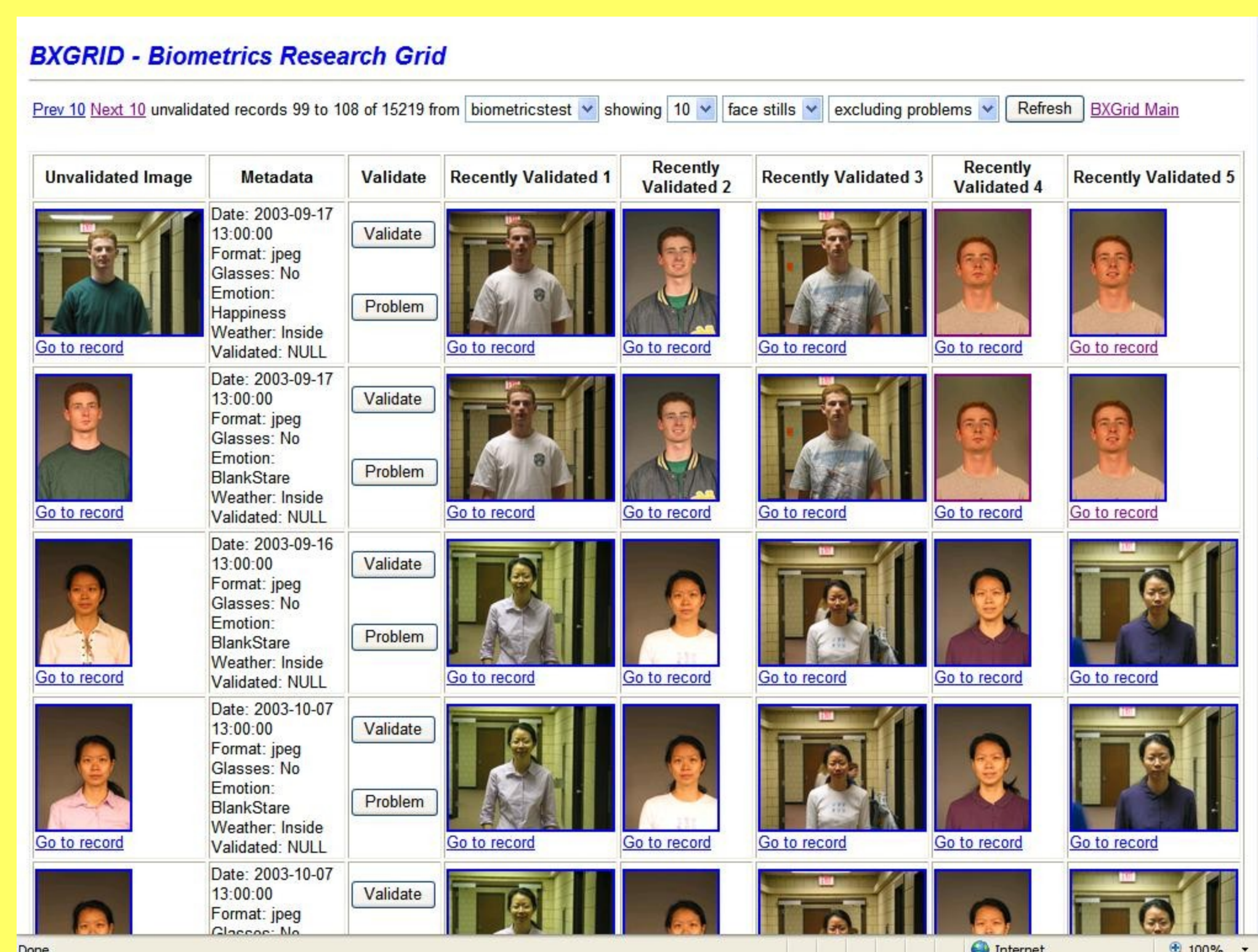


BXGrid Overview

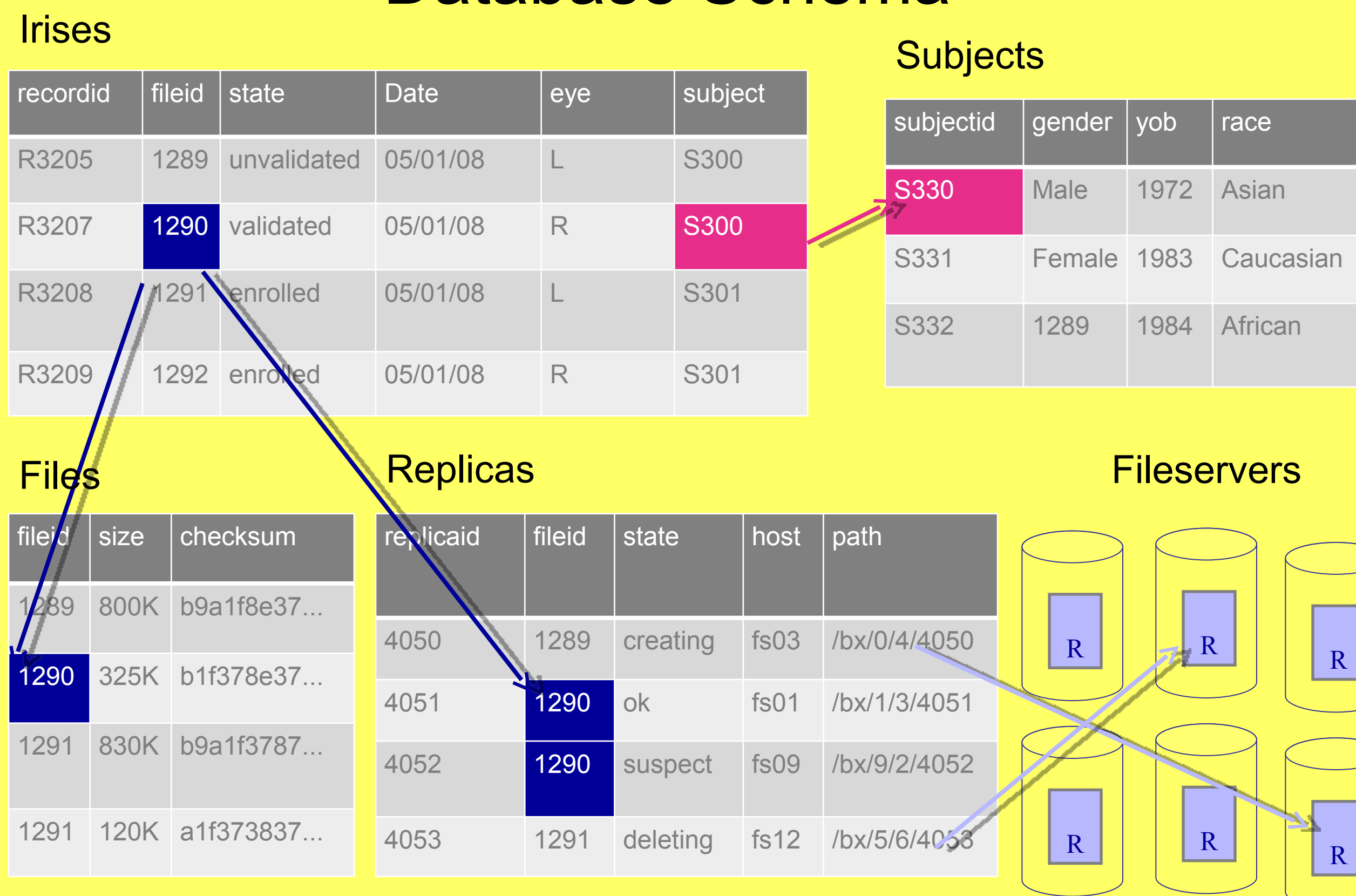
- **Database:** A conventional relational database is used to manage the metadata of all biometrics data: iris image, iris video, face image, face video, etc.
- **Active Storage Cluster:** A clusters of machines, each running Chirp light weight fileserver, stores the actual images, videos, and other large data files.
- **Computing Grid:** A campus computing grid, which consists of 500 CPUs, carries out CPU intensive analysis experiments such as AllPairs.
- **Web Portal:** A user friendly web interface let users navigate the system, validate metadata, and perform other operations.



Web Portal

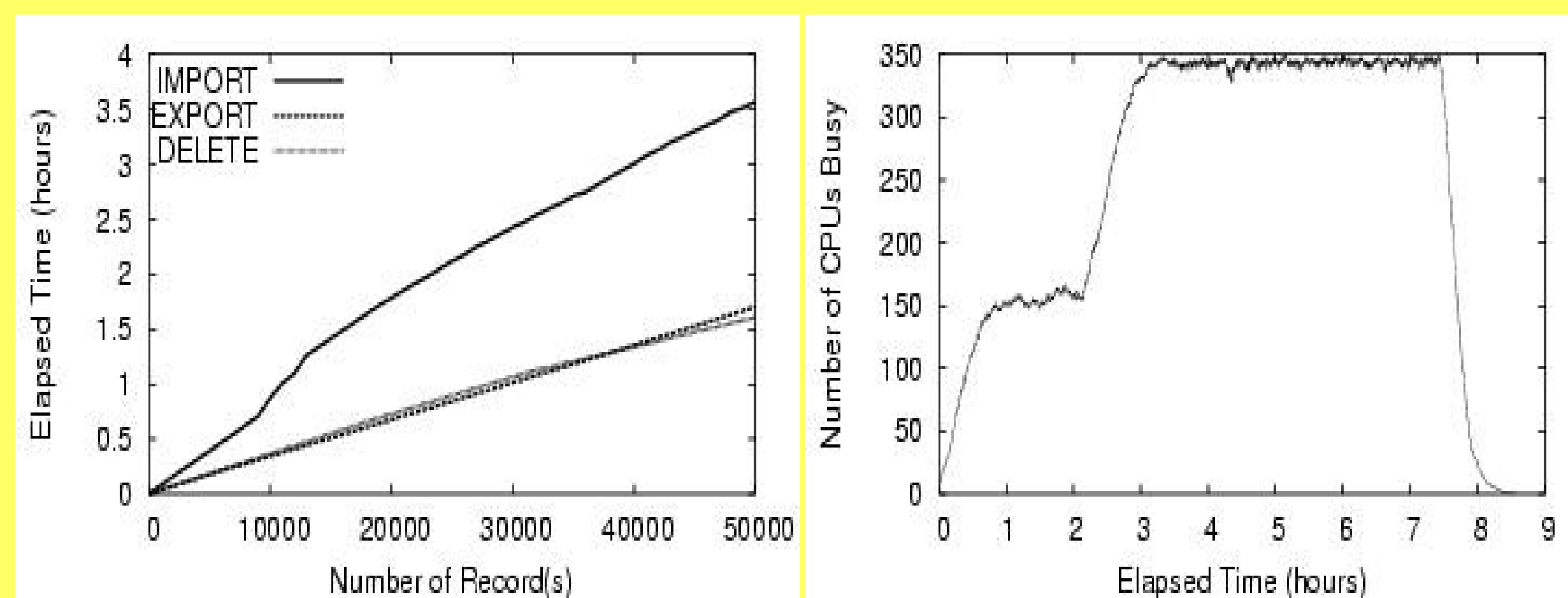


Database Schema



Operational Benchmarks & All-Pairs Example

Perform Import/Export/Delete on 50,000 records



This work was supported by NSF Grant CCF-0621434 "Deconstructing Clusters for High End Biometric Applications".
Web address: <http://www.cse.nd.edu/~ccl>

