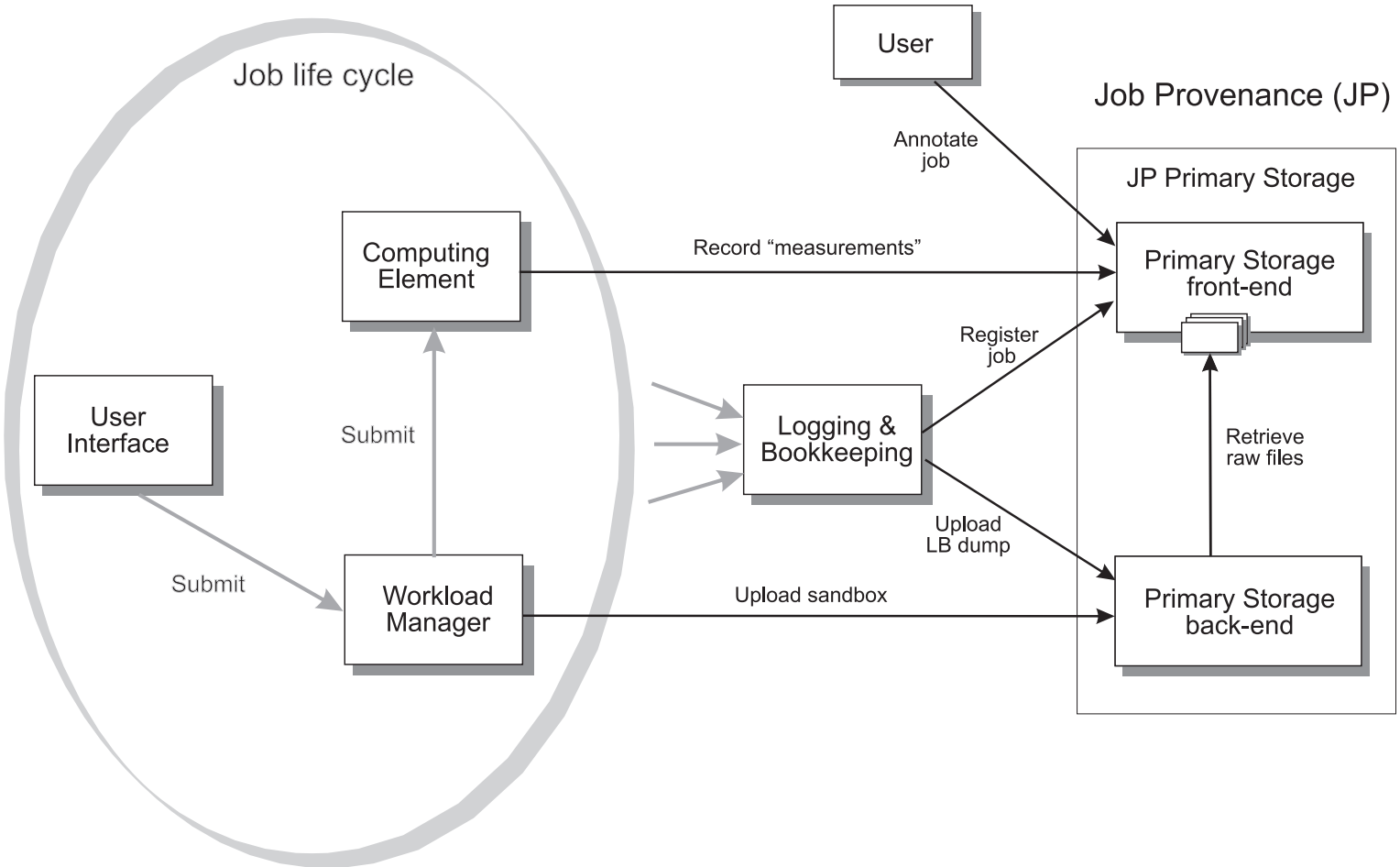
Job Provenance in gLite 3.1

Jiří Sitera – JRA1 CZ cluster, CESNET

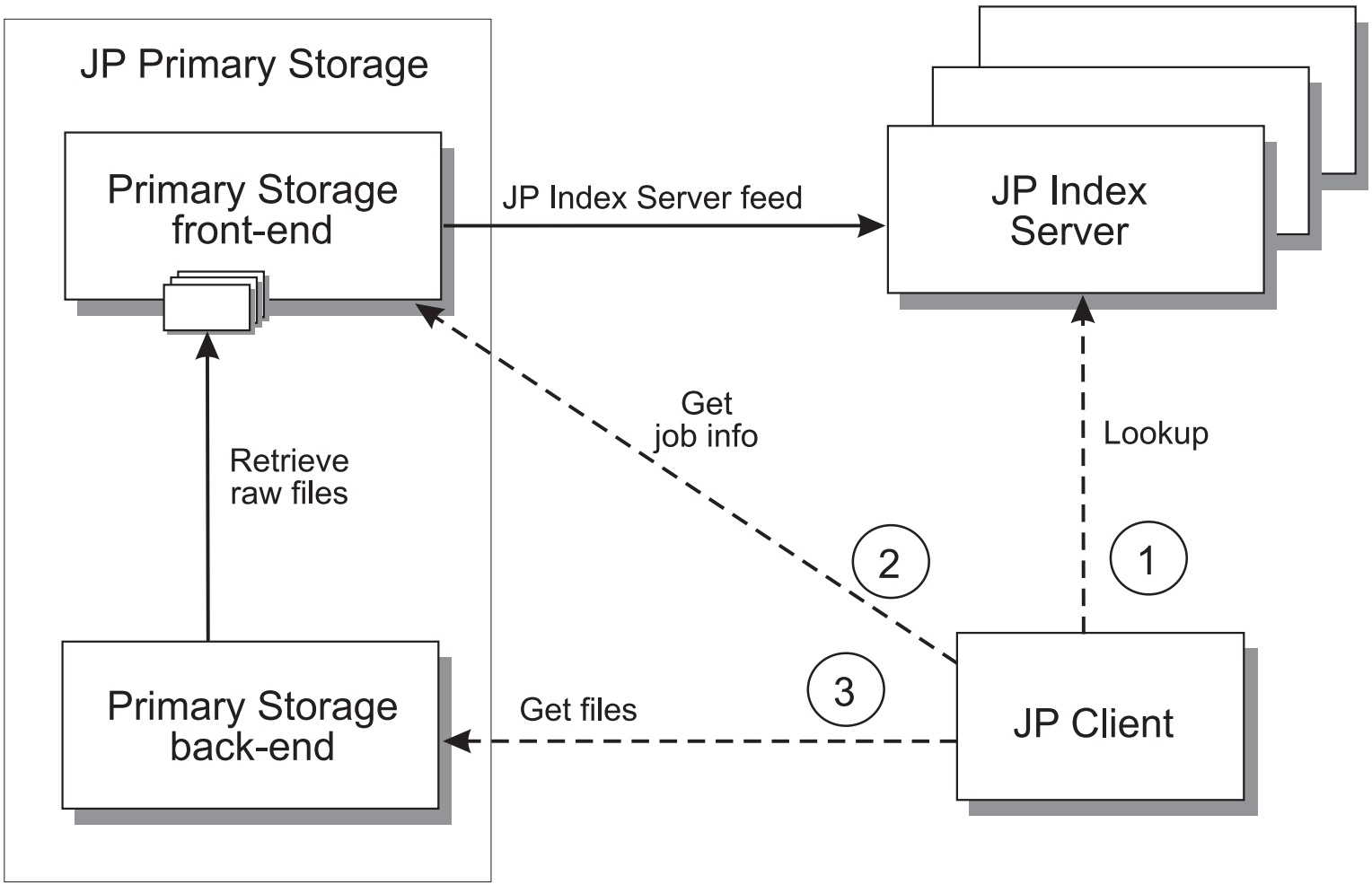
Job Provenance

- Long-term storage of job related data
  - LB data (JDL, job execution track)
  - Miscelaneous input files (input sandbox)
  - "measurements" on CE (installed software, environment)
  - Accounting data (DGAS)
  - User annotations
- Data mining in the stored information
  - Looking up 'provenance' of data
  - Assisted job resubmission

Job Provenance – data gathering part



Job Provenance – user query part



Job Provenance – first release

- Job Provenance as a new component in gLite 3.1
  - Deployed as two components
    - JP Primary Storage
    - JP Index Server
  - The LB deployment contains JP feeding elements
  - New end-user interface (web service, CLI available)
- Limitations in the first release
  - Integrated with LB only
  - Configuration of index servers limited
    - Index server logic already favours the configuration represented internally, currently hardcoded
    - The only missing part is the configuration management (configuration format (file) and its parser)

Summary and future plans

- Job Provenance: Long-term storage of job related data
  - Data-mining service for users
  - Users can annotate ended jobs
  - Access to all job related data (allowing for resubmisson etc.)
- JP is a new gLite component, in 3.1 with limited features
- Future plans
  - Evaluation of the first release deployment
  - Other data sources (sandbox upload)
  - Client integration into UI
  - Configuration & other internal enhancement of JP