

# Interactive Computing E-Infrastructure Public Information Event



**FENIX**  
RESEARCH INFRASTRUCTURE

Co-funded by  
the European Union



## R&D topic: Data Mover

### Dorian Krause

Barcelona

15 March 2018

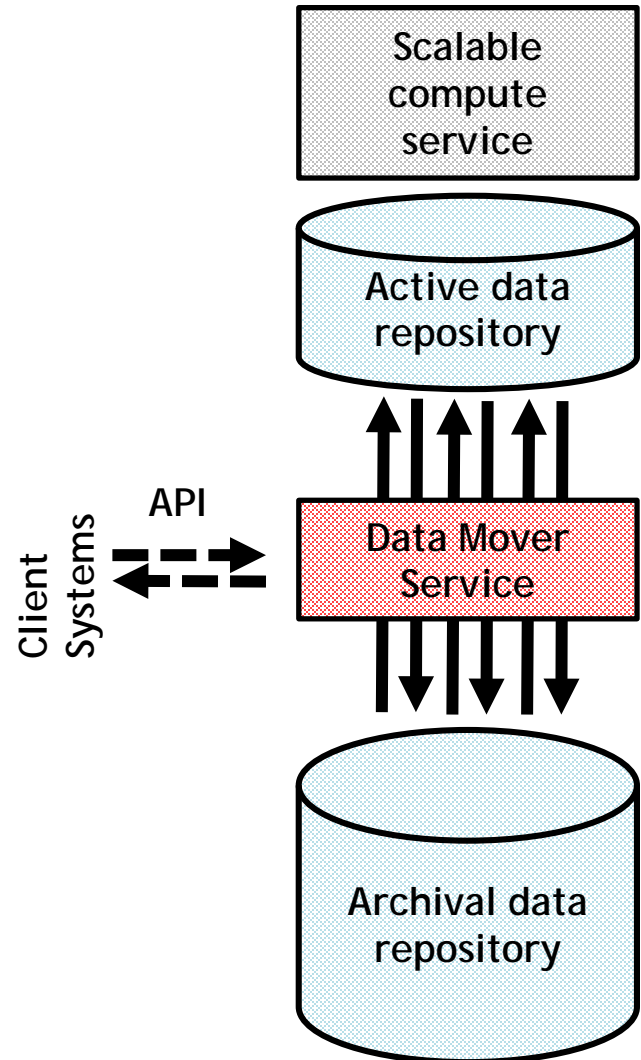


# Disclaimer

*Neither this announcement nor the event itself signifies the beginning of a procurement procedure or constitutes a commitment by the public procurers involved in the presentation to undertake such exercise at a later stage*

# Background

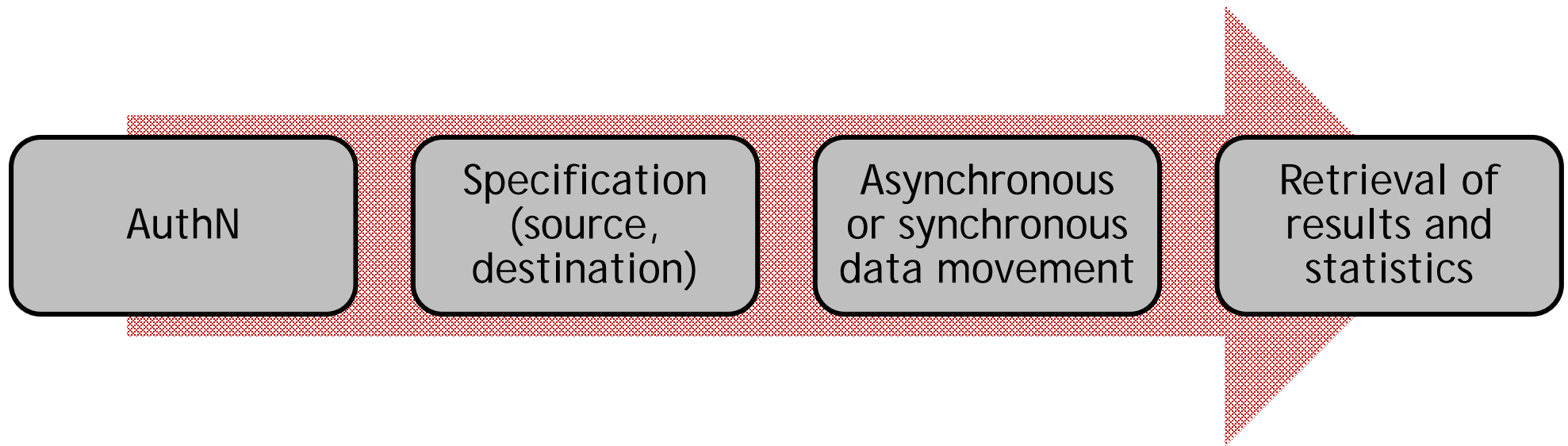
- Scalable compute services access active data repositories (parallel FS)
  - Assumption:  $\approx$  POSIX
- Archival data repositories are federated (SWIFT API)
- Fenix Workflows utilizing scalable compute services require staging step
- Data Mover: High-performance, secure, programmable data movement between archival and active data repositories



# Current thinking on use cases (1/2)

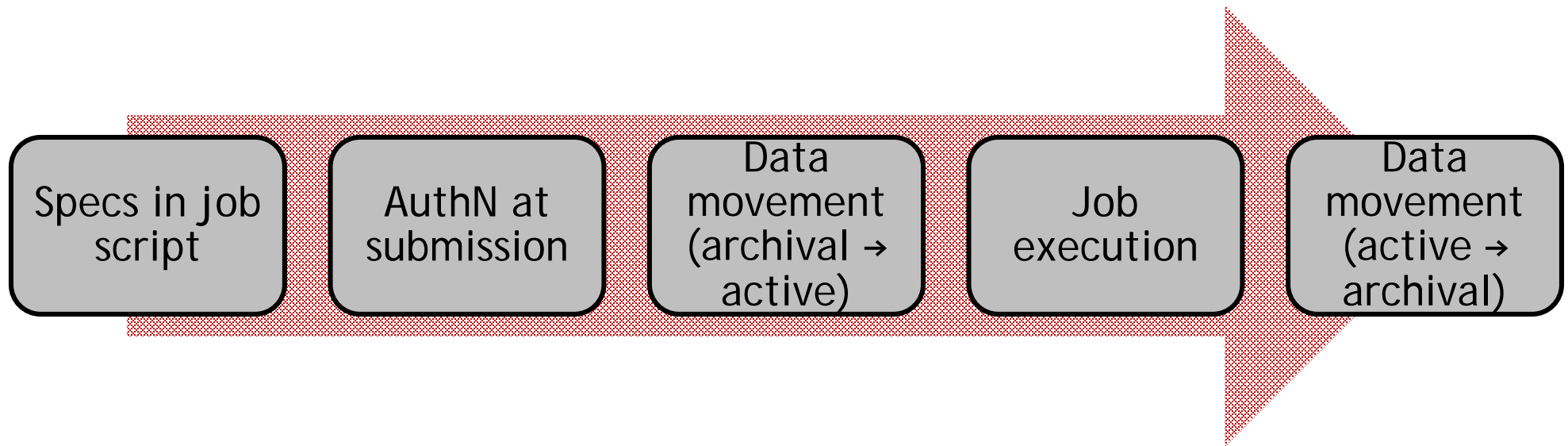
- **Manual data movement**

- Data movement triggered by external client commands via CLI or through workflow management system
- Usage independent of scalable compute resources



# Current thinking on use cases (2/2)

- **Automatic data movement by scalable computing services WLM**
  - Data movement specifications as part of batch job
  - Integration with scheduler required



# Need for R&D on data mover

- Data movement management technologies widely utilized (e.g., hierarchical storage management solutions)
- **But:** Existing solutions on market insufficient for ICEI use-cases (but potentially good basis)
- Missing key characteristics:
  - Lack of support for archival data repository object-storage interface
  - Scalability and performance targets not met
  - Lack of integration with user workflows
  - Federation requirements not met (e.g., auth\*)

# Expected outcome

1. Specification of a vendor-agnostic, scalable hardware platform for data mover services
  - Separate hardware procurements
2. Software for high-performance and secure data movement from the archival data repository to a sufficiently broad class of active data repository technologies (POSIX)
  - In particular: IBM Spectrum Scale, Lustre
  - Optimizations for specific setups and file systems where applicable
3. Software for the workflow integration, in particular, scheduling of data movement in conjunction with (batch) jobs in the federated infrastructure.
4. Maintenance/support services for project duration
  - Sustainability plan will need to be provided and implemented



Co-funded by  
the European Union



# Thank You

[www.fenix-ri.eu](http://www.fenix-ri.eu)