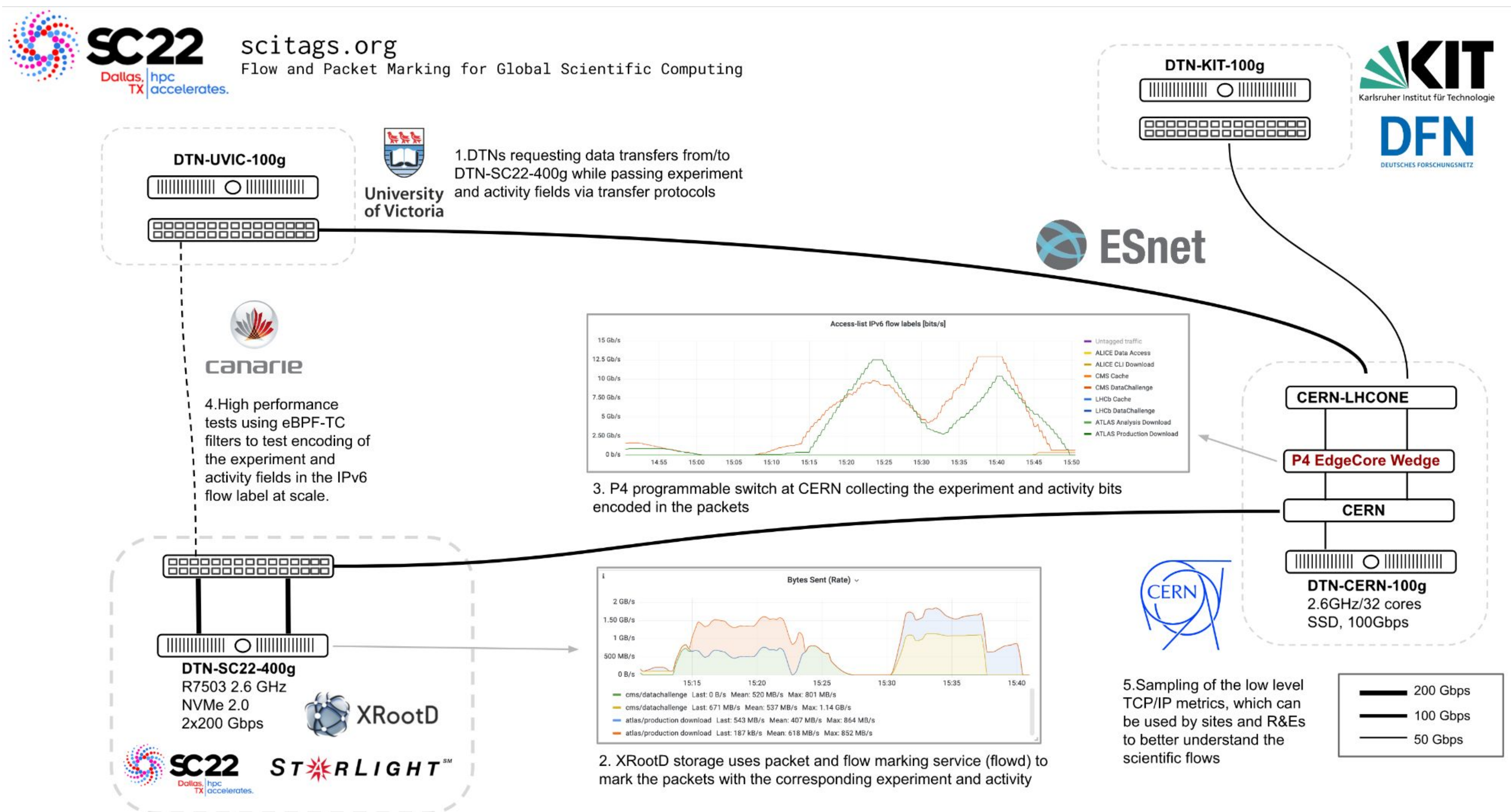


# Scientific Network Tags

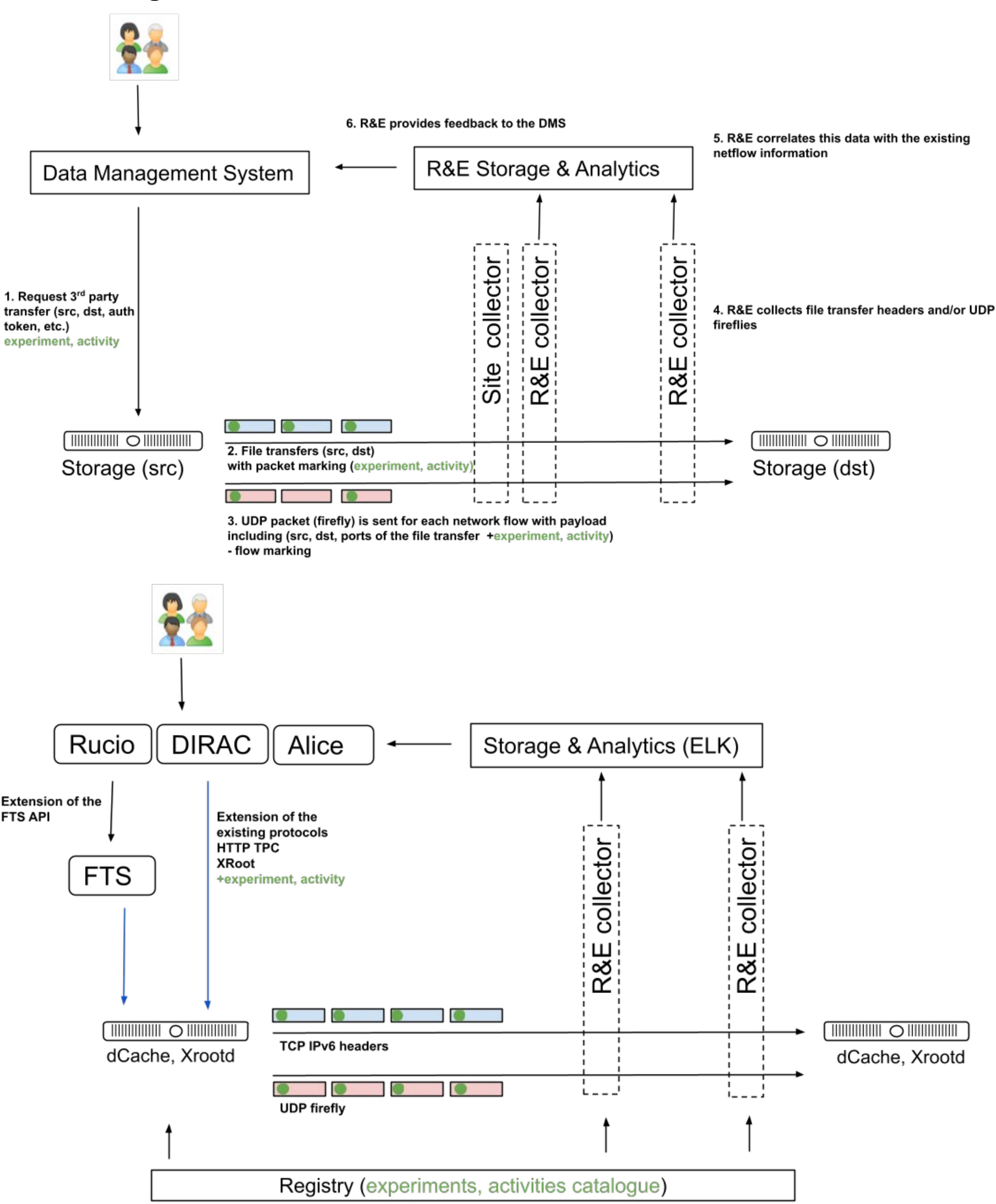
## Packet Marking for Data Intensive Scientific Workflows



Managing large scale scientific workflows over networks is becoming increasingly complex, especially as multiple science projects share the same foundation resources simultaneously yet are governed by multiple divergent variables: requirements, constraints, configurations, technologies etc. A key method to address this issue is to employ techniques that provide high fidelity visibility into exactly how science flows utilize network resources end-to-end. This demonstration showcases one such method, Scientific Network Tags (Scitags), an initiative that is promoting identification of the science domains and their high-level activities at the network level. This open system initiative provides open source technologies to help Research and Education Networks understand resource utilization while providing information to scientific communities on the behavior of their network flows.



### How Scitags Work ?



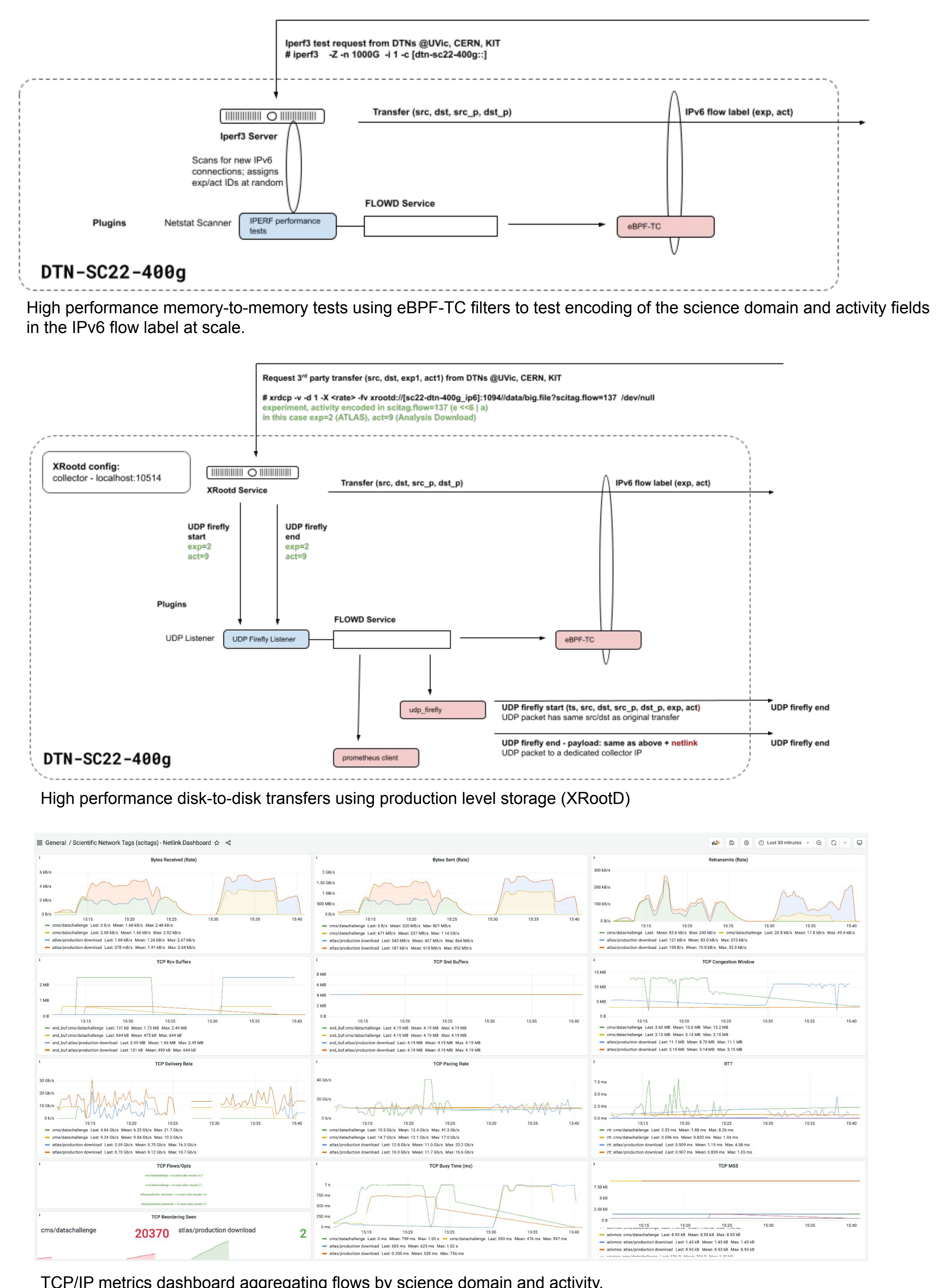
### What's next ?

- Near-term objectives
  - Start rollout and testing of Xrootd implementation
    - Detect flow identifiers from storage path/url, activities from user role mapping
    - Test proxies, cached proxies, private networks (K8s)
  - Finalise development and deploy network of receivers
  - Instrument Rucio/FTS to pass flow identifiers to the storages
  - Involve other storage systems (dCache, etc.); discuss possible design/implementation
- Engage other R&Es and explore available technologies for collectors
  - Deploy additional collectors and perform R&D in the packet collectors
  - Improve existing data collection and analytics
- Test and validate ways to propagate flow identifiers
  - Engage experiments and data management systems
  - Validate, test protocol extensions and FTS integration
  - Explore other possibilities for flow identifier propagation, e.g. tokens
- Plan to submit IETF Informational RFC

• **Scientific Network Tags** (scitags) is an initiative promoting identification of the science domains and their high-level activities at the network level.

- Enables tracking and correlation of the data transfers with the Research and Education Network Providers (R&Es) network flow monitoring
- Helps science domains to better understand how their network flows perform along the path
  - Improves visibility into network flows performance (per activity) within R&E segments
  - Helps to get insights into how science domains are using the networks, get additional data from R&Es on behaviour of the transfers (traffic, paths, etc.)
- Sites can get visibility into how different network flows perform
  - Network monitoring per flow (with experiment/activity information)
    - E.g. RTT, retransmits, segment size, congestion window, all per flow

### SC22 Demonstrations



Acknowledgements: We would like to thank the WLCG, HEPiX and OSG organizations for their work on the topics presented. In addition we want to explicitly acknowledge the support of the National Science Foundation which supported this work via: OSG: NSF MPS-1148698 and IRIS-HEP: NSF OAC-1836650