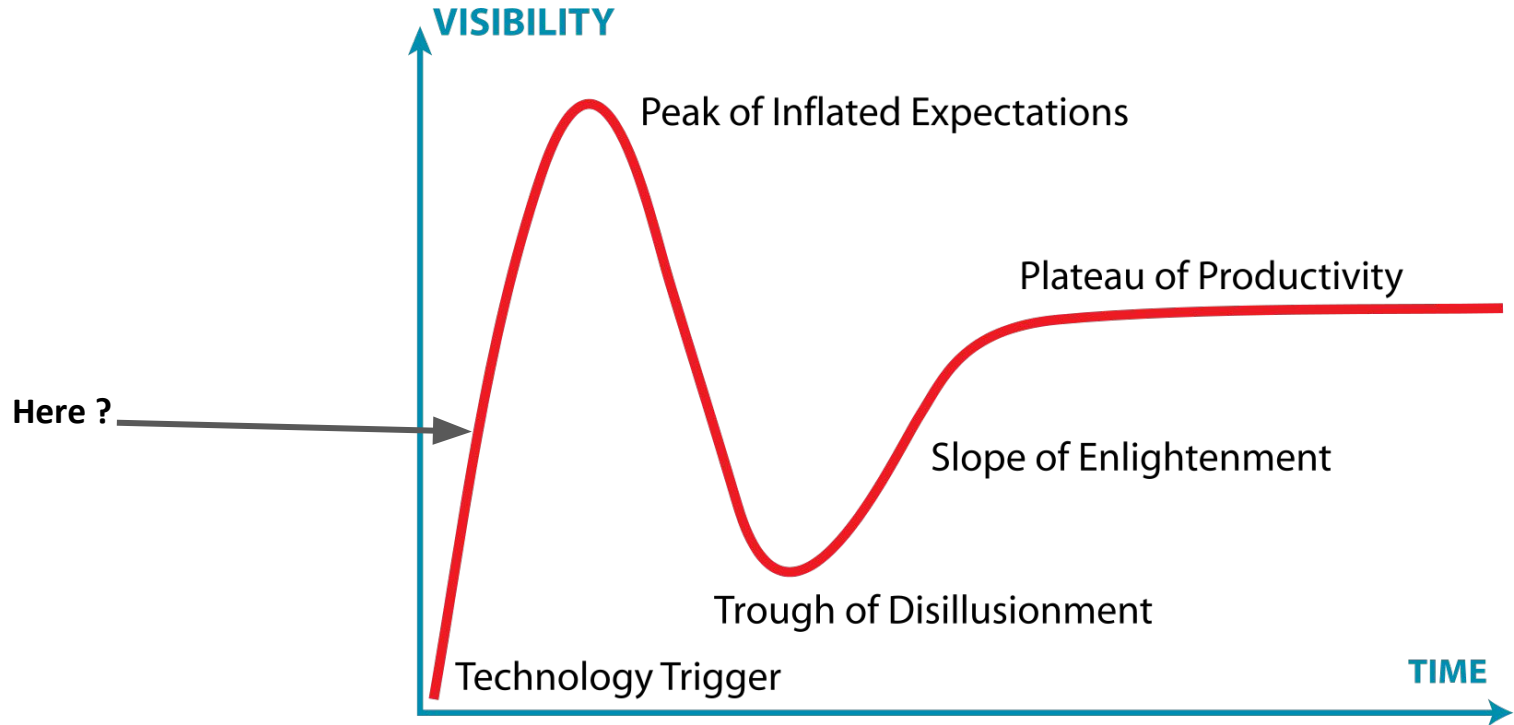




# Discussion

---

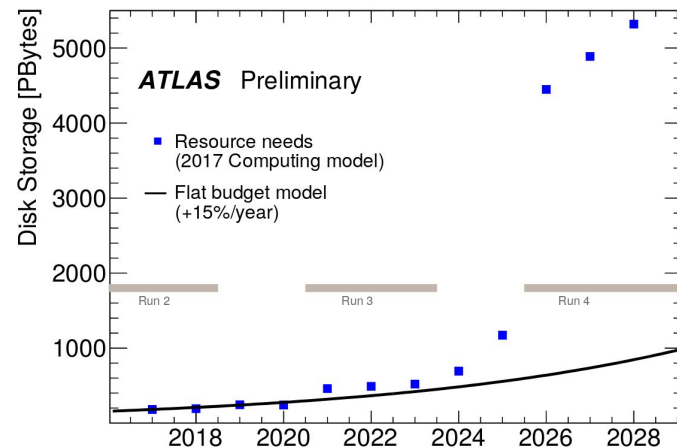
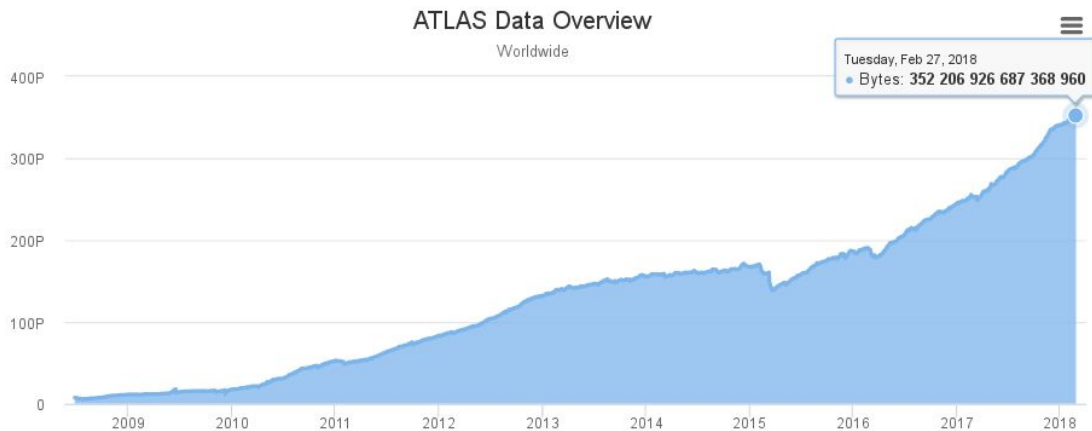
# Where is Rucio now ?





**Special acknowledgement to the weather !**

# Long-term development strategy



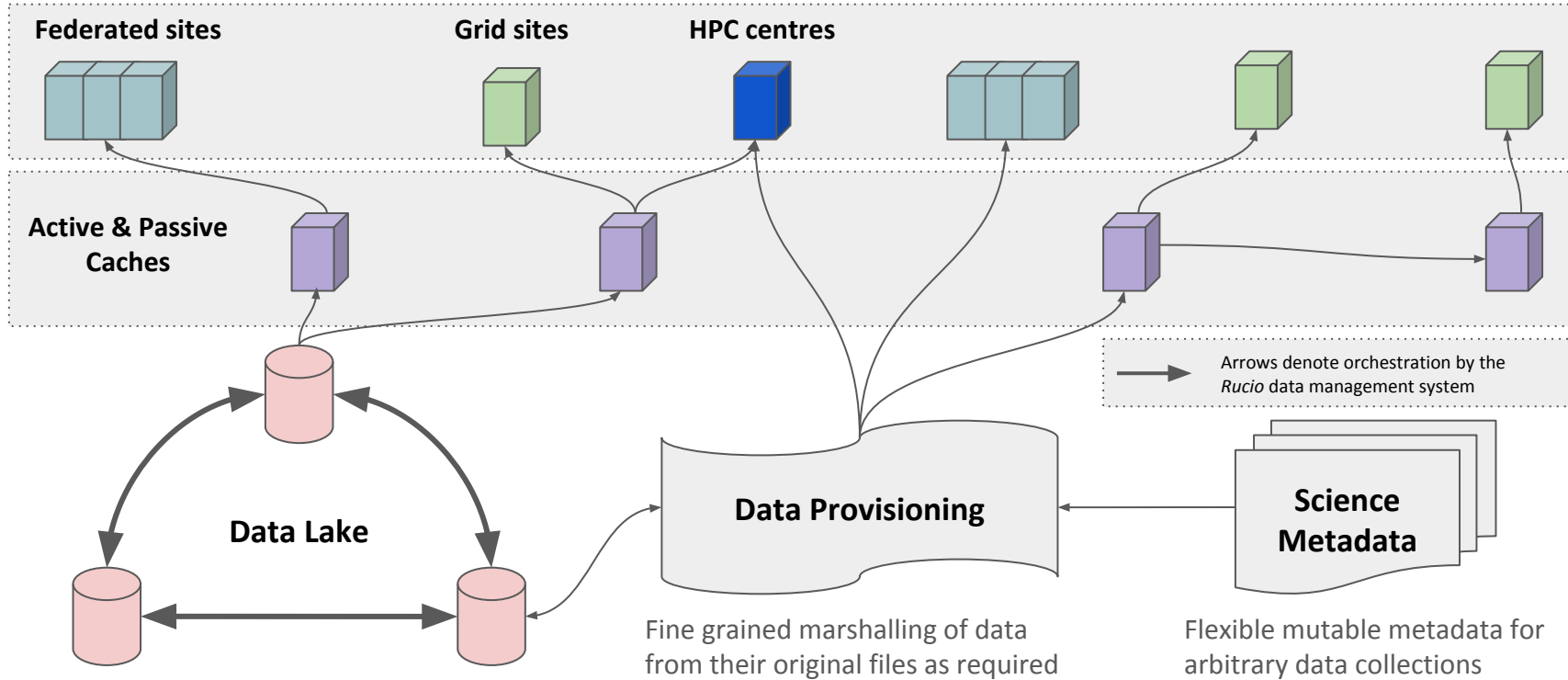
- **Major challenges that will drive HEP data management R&D in the next 5-10 years**

- Factor 10+ increase in storage requirements (>5 Exabytes)
- Factor 100+ increase in number of objects to manage (>100 billion)

- **Long-term data management R&D strategy**

- Explore the usage of large reliable data centres ("lakes") with wide area cache control
- Explore sub-file data object processing to better utilise heterogeneous compute and associated networks
- Stronger integration of software (file formats, new IO patterns, interactive analysis) with distributed computing

# Fine-grained workflows with streaming delivery



# Question box

---

Q: Rucio is mainly Python-based (IIRC). What are the performance bottlenecks this language induce? Are there plans to migrate (in piecewise, micro-service way) to more performant/resources (RSS, CPU, latency, ...) friendly languages?

Q: "\* sub-file support being explored" — why? What's the use-case/user? Costs/Risks/benefits?

Q: "Support for arbitrary metadata being explored" — more info? Semantics/API?

Q: Can specific permissions be applied to individual (and/or arbitrary) metadata attributes? (E.g., only community admins can modify the "isProduction?" bit on a dataset?)

# Potential topics (1)

---

- Based on the [Live minutes](#)
- Flexible metadata component
  - Mentioned many times in the sessions, e.g., OAIS, OID
- More authentication and authorisation support
  - EDUGAIN, SciTokens, Macaroons, OpenID (connect), etc.
- Storage protocols & Data access
  - Native object stores without S3, e.g., ceph
  - Direct access of files within archives (e.g., zipped files)
  - Integration of more transfer tools for third party copy

# Potential topics (2)

---

- Web/Sync&Share interface to Rucio, a la AMS/ASGC ?
  - Userclient data synchronisation à la Dropbox/Owncloud
  - Transparent access to data via WebUI
- Support of different (open source) databases with multi-billion entries
  - PostgreSQL sounds like a solid candidate
- Workflow management system interfaces
  - E.g., HTCondor, Pegasus, etc.
  - DIRAC mentioned few times
  - Already integrated with PanDA, ARC



# Potential topics (3)

---

- Deployment model: Rucio as a Service
  - Current model for OSG, ASGC
  - Other institutes interested
    - FNAL, Oslo, BNL, NeIC
- And many many more ?