

# Publication Working Team: Progress update and future roadmap



ESGF F2F Conference  
Washington, DC, December 2016



Team lead: Sasha Ames

Lukasz Lacinski, Rick Wagner(ANL), Guillaume Levavasseur (IPSL), Ag Stevens, Alan Iwi. (CEDA) Katharina Berger, Hans Rathmun. (DKRZ), Serguei Nikonov (GFDL), Georgi Kostov(NOAA), Luca Cinquini (JPL/ESRL), Claire Trenham (ANU)

This is a coordinated multiagency, international collaboration of institutions that continually develop, deploy, and maintain software needed to facilitate and empower the study of climate



# 2016 progress and accomplishments

---

- 22 projects republished
- “Administrative” merger with QCWT given strong overlap of concerns
- esg-publisher development
  - Project-specific ini files
    - No need to reconfigure for different file system roots
  - Single-dataset mapfiles
  - Publisher test suite (needs final integration)
  - Globus urls
- Release of esg-prep
  - Functionality for mapfile generation, vocabulary check, ini file retrieval
- Ingestion service API deployment and integration into ACME Workbench
- Comprehensive documentation: <http://esgf.github.io/esg-publisher>

# Progress in preparing the publisher for CMIP6

---

- Support for “list” facets
- Automated global attribute extraction
- Integration of CF checks, PrePARE, aka CMOR CMIP6-checker library (Denis Nadeau)
  - Presents additional installation prerequisites
- Retraction of datasets – functionality for unpublishing
- Test CMOR-ized datasets published on test nodes

# 2017 Roadmap

---

- Update installation to deploy via Conda
  - Phase one: Conda integrated into installer for publisher dependencies
  - Phase two: publisher Conda package
- Add “drs” manager to esgprep
- CMIP6 publishing
- ES-DOC generation (via integrated service call)
- Tracking service integration (for dataset version update and retraction)
- Integrate Ingestion API into installation
- Address scalability concerns of publisher database access, current uvcdat

# Collaborations

---

- IWT on installation update
- Replication/versioning on PID integration and automated replica publishing
- PCMDI/WIP on PrePARE integration, sample CMIP6 data

# Node Manager, Tracking and Feedback: Progress update and future roadmap



ESGF F2F Conference  
Washington, DC, December 2016



Team lead: Sasha Ames  
Prashanth, Luca, Katharina, Nicolas

# Progress and Accomplishments for 2016: Node Manager

---

- **Node Manager:** “Core” ESGF module that maintains node status, federated metadata, properties, metrics and registry of services
- Django module integrated into apache for node-based deployment
  - Worker/monitoring daemon process running via control scripts
  - Task queue of requests through file system
- Testing performed with up to 7 nodes in cluster environment; five test-federation nodes
- Dashboard integration tests with served registration.xml
- Refactor and module conversion to SSL tested in virtual environment
- Code merged into github: <https://github.com/ESGF/esgf-node-manager>



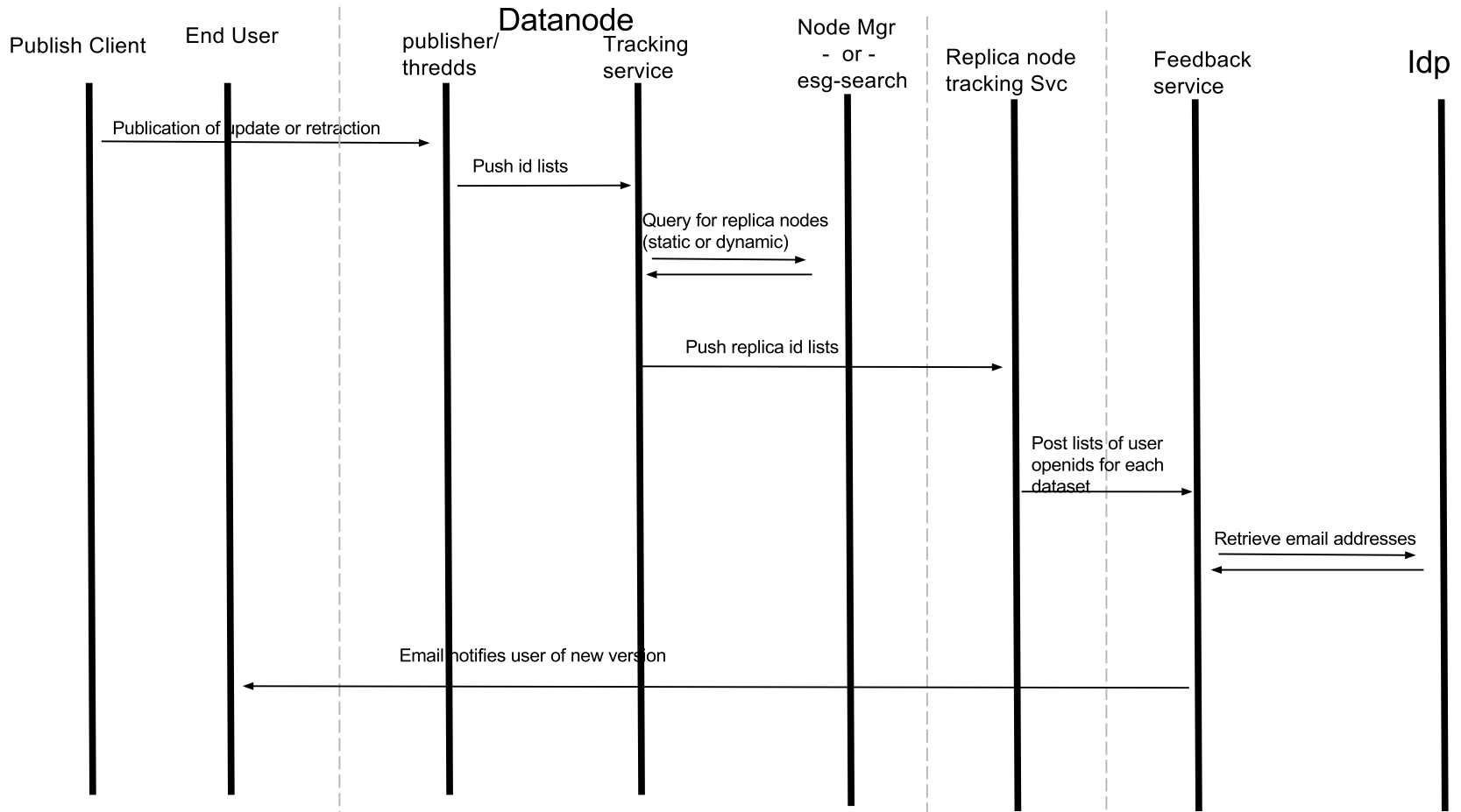
# Tracking and Feedback: scope and status

---

- Provide ESGF data users with notification emails when certain “changes” occur based on previously downloaded content
  - Publication of updated dataset version
  - Retraction of dataset
- Status: architecture documented

# Tracking and Feedback architecture

## Replica Node Tracking and Feedback Workflow



# Missed milestones

---

- Node Manager: production deployment
- Tracking and Feedback: prototype implementation

# 2017 Roadmap

---

- Node Manager
  - Convert control scripts to esg-node
  - Integrate setup into installer
  - Federation production deployment
  - Implement secret key deployment based on signed certificates so we can identify “trusted” nodes
  - Determine if esgf\_\*.xml (non-static) files should be managed automatically as done by previous Java esgf\_node\_manager
- Tracking/Feedback
  - Begin services implementation
  - Integrate publisher notifications to tracking service
  - Feedback service: first version read user email address direct from PGSQL db
  - Query email address from IdP API

# Working teams collaborations

---

- Node Manager
  - IWT – integrate setup into installer
  - CWT – resource availability
  - Stats – determine more metrics
  - Others – use of Node Manager for node status monitoring, sharing of metadata, etc.
    - Eg. make an IdP “not selectable” if determined to be down
- Tracking/Feedback
  - IdEA – query IdP for user emails
  - PWT, Errata – integration of tracking
  - CoG, search – Saved search UI

# Additional Resources Needed

---

- Full/part time developer(s) would help out a lot