ECPDS: Acquiring, Storing, and Delivering Data Efficiently Worldwide

Multi-purpose Data Store that acquires data from providers, disseminates it to remote sites, and allows remote sites to push or pull data on demand.

ECPDS boosts the efficiency and productivity of data services by using proven and innovative technologies. It offers a portable, adaptable application for diverse environments, with a user-friendly tool for managing data acquisition, push/pull dissemination, and notifications, all using standard protocols.

Driving Forces

- ECMWF forecasts must reach Member and Co-operating States promptly.
- Mature, in-house solution enhancing efficiency and productivity.
- Enables collaboration and community contributions via **OpenECPDS** on GitHub, using a developer container with step-by-step guides for VS Code and Eclipse users for fast onboarding.
- Container-based: scales from laptops to large deployments with hundreds of systems.

Data Storage and Retrieval

- Works like a search engine: indexes metadata, caches data as needed.
- Data sources:
  - Fetched via Data Acquisition service.
  - Pushed through the Data Portal.
  - Registered asynchronously via ECPDS API.
- Users access data via streaming or cached copies.

Protocols and Connections

- **Outgoing:** FTP, SFTP, FTPS, HTTP/S, Amazon S3,Microsoft Azure, Google Cloud Storage.
- **Incoming:** FTP, HTTPS, S3 (SFTP/SCP via Commercial API).
- Modular design supports adding new protocols easily.

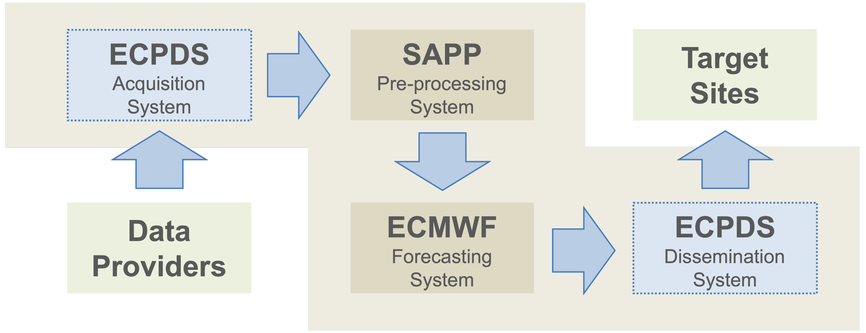
Object Storage

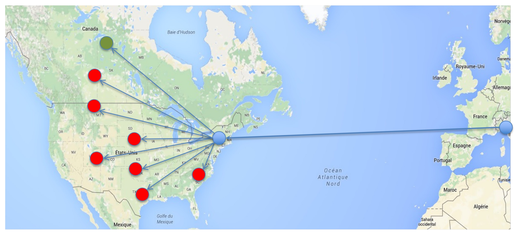
- Stores data as objects: data + metadata + unique ID.
- Replication across multiple locations for high availability.
- Hierarchy-free, emulate directories based on metadata.
- Multiple views depending on user needs.

Additional Features

- **Notification System:** MQTT broker + client.
- **Data Compression:** lzma, zip, gzip, bzip2, lbzip2, lz4, snappy.
- **Data Checksumming:** MD5 (remote), ADLER32 (Data Store).
- **Garbage Collection:** automatically removes expired data.
- **Data Backup:** maps datasets to existing archiving systems.

Essential role of ECPDS in the ECMWF Numerical Weather Prediction (NWP) system

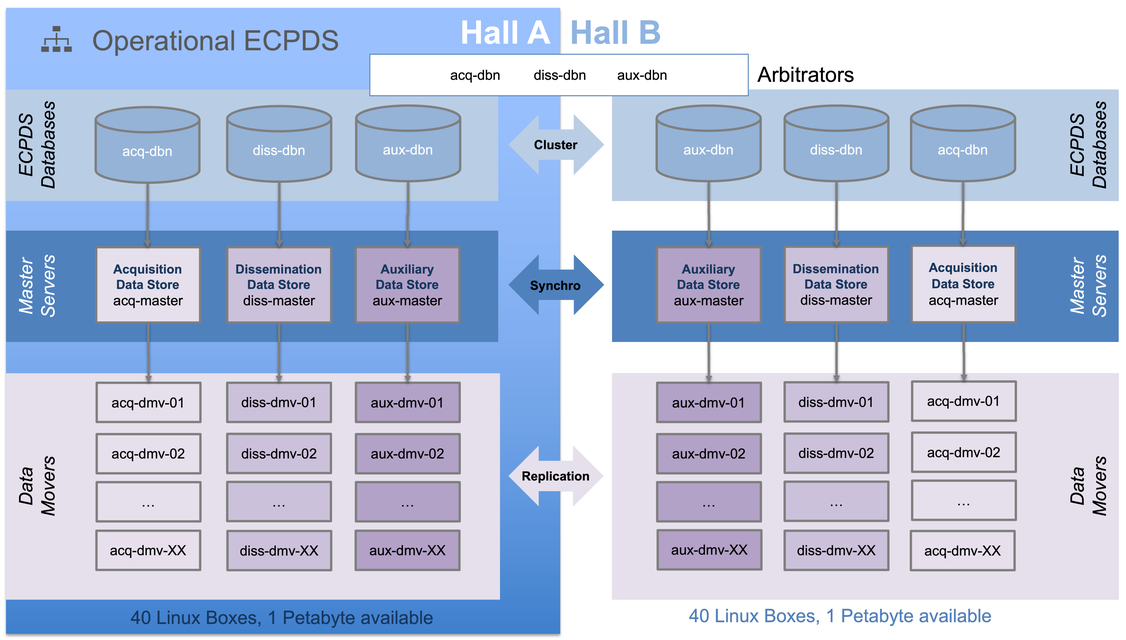




Efficient data dissemination ensures timely access to critical information, even for large datasets and distributed users

Continental Data Movers optimize transfers by reducing latency, balancing load, and improving reliability through strategic deployment, network tuning, and data pre-replication.

High-level overview of the physical infrastructure supporting ECPDS at ECMWF

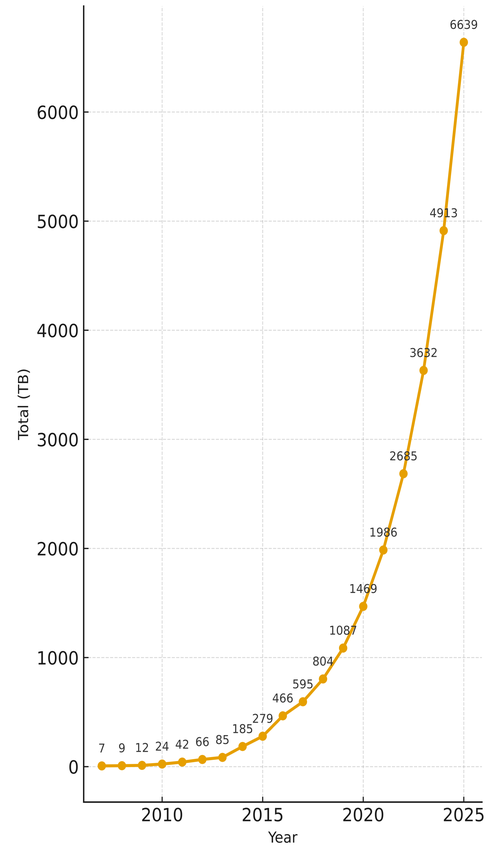


Developed and maintained by the ECPDS Team – ECMWF
GitHub: https://github.com/ecmwf/open-ecpds
Contact: ecpds@ecmwf.int

The **ECPDS infrastructure** supports a vast network of **over 1,000 destinations** for **acquisition and dissemination**, marking a significant milestone in its expansion. Spanning **more than 80 countries**, this network enables **continuous, around-the-clock data exchange**.

Annual Data Volume Growth (TB per month)

Average monthly data volume transferred each year, showing a steady increase over the years.



ECPDS integrates an MQTT Broker and Client to streamline data flows

The **Broker** lets dissemination users subscribe to notifications for new products, enabling automated downstream processing.
The **Client** registers with external providers to detect and retrieve new data automatically, ensuring seamless acquisition and integration.