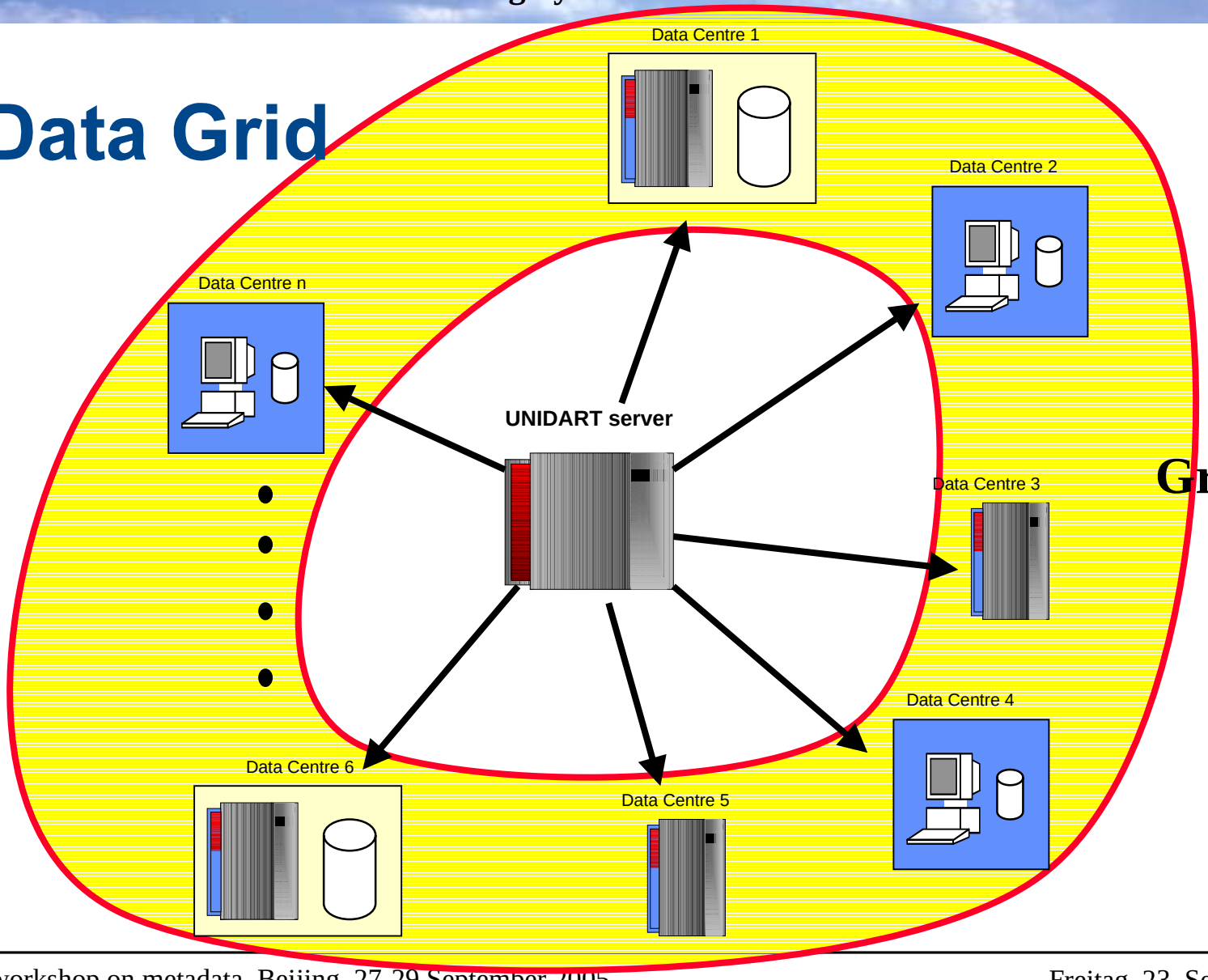
The Metadata Harvesting Procedure in UNIDART

Jürgen Seib
Deutscher Wetterdienst
e-mail: juergen.seib@dwd.de

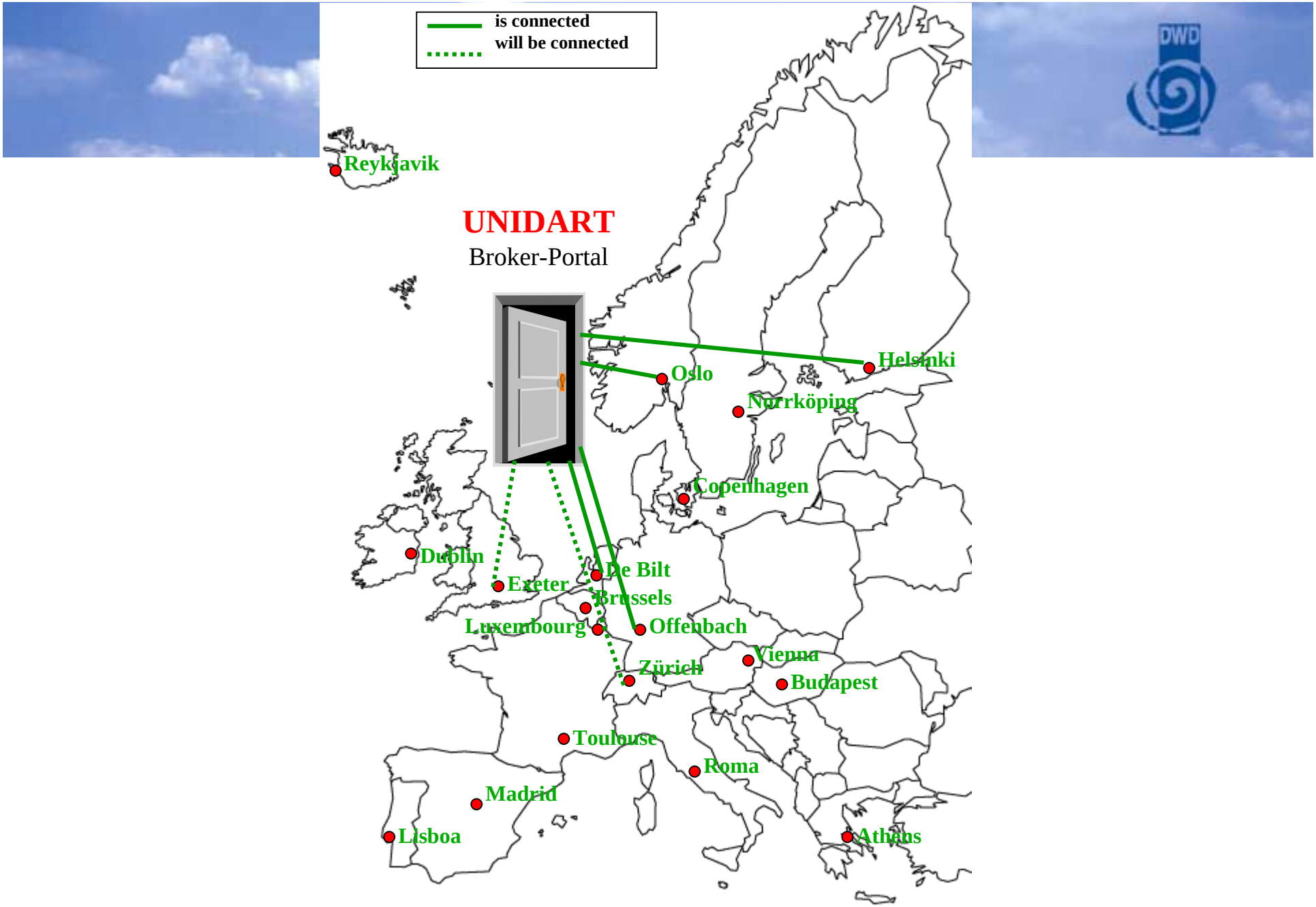
The main goal of UNIDART

Development of a Web-based information system that allows a **uniform and integrated access** to **heterogeneous and distributed data sources** which store **any kind of meteorological data and products**

Data Grid



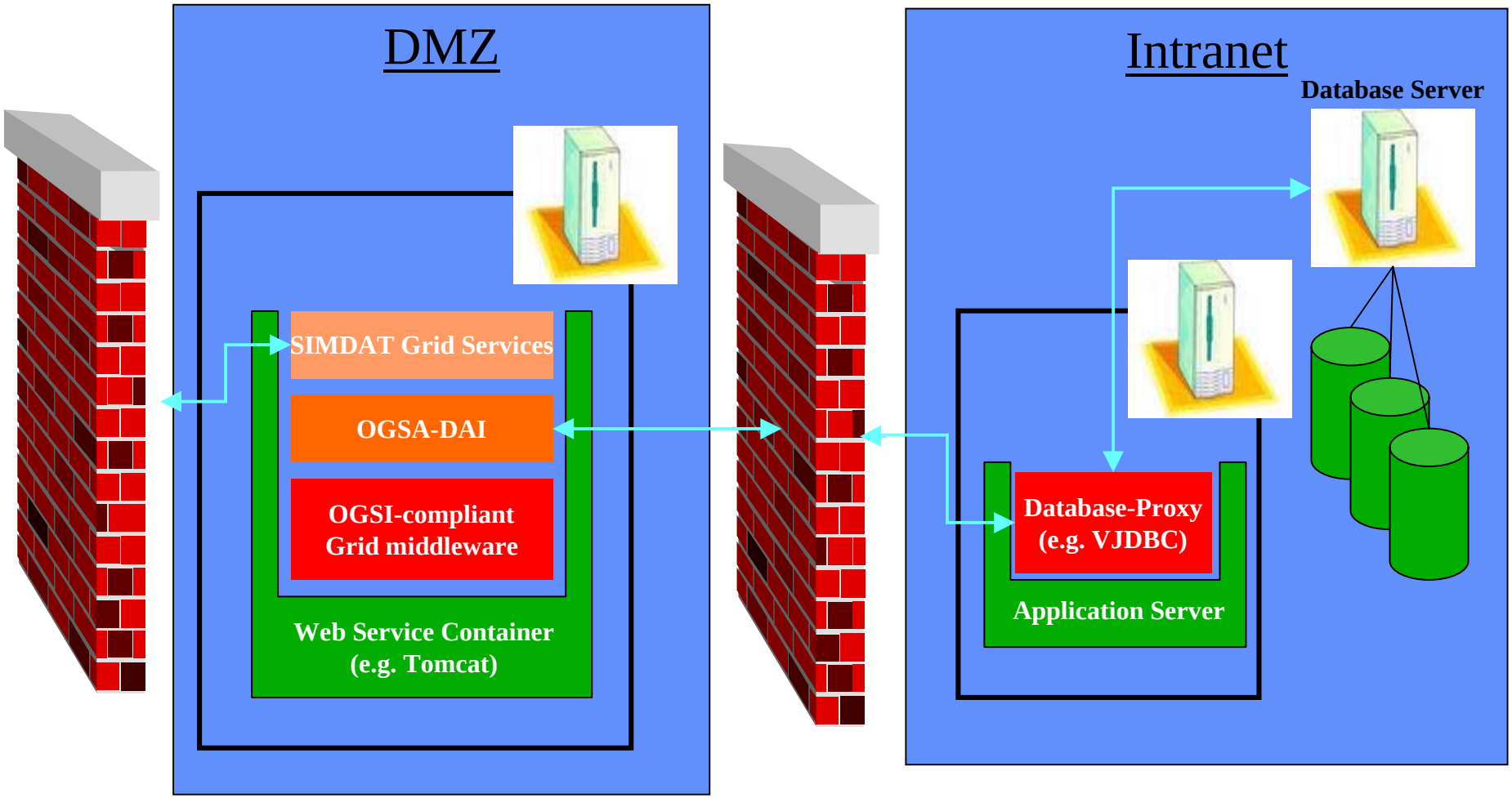
Grid nodes


is connected
will be connected
Reykjavik
UNIDART
Broker-Portal
Helsinki
Oslo
Norrköping
Copenhagen
Dublin
De Bilt
Exeter
Brussels
Luxembourg
Offenbach
Vienna
Zürich
Budapest
Toulouse
Roma
Madrid
Lisboa
Athens

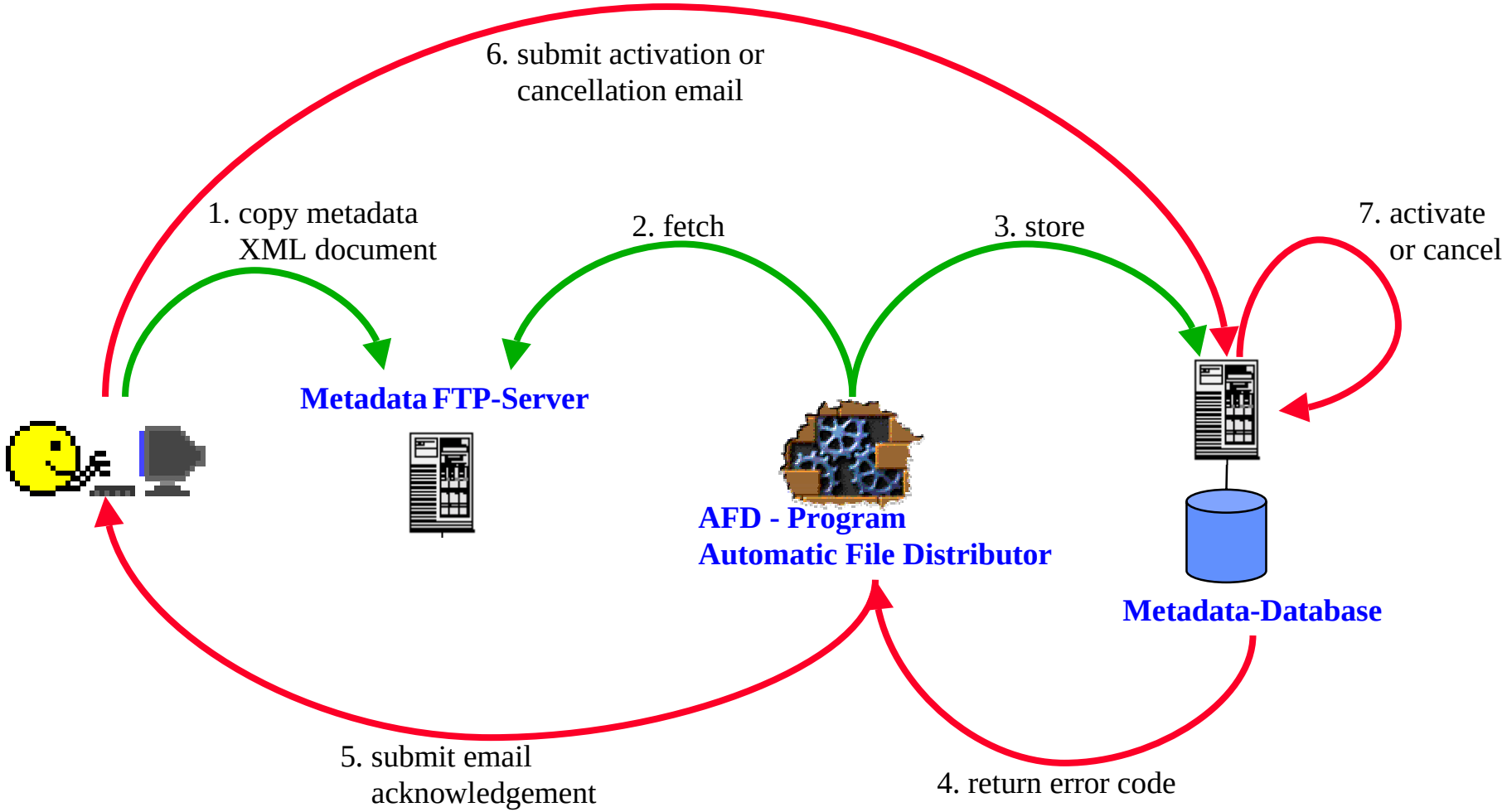
UNIDART server

- connected to each grid node as grid client
- runs the portal application for user interaction
- maintains and manages all metadata in an XML database
- metadata is a set of XML documents
- does XML schema validation
- provides XML query facilities

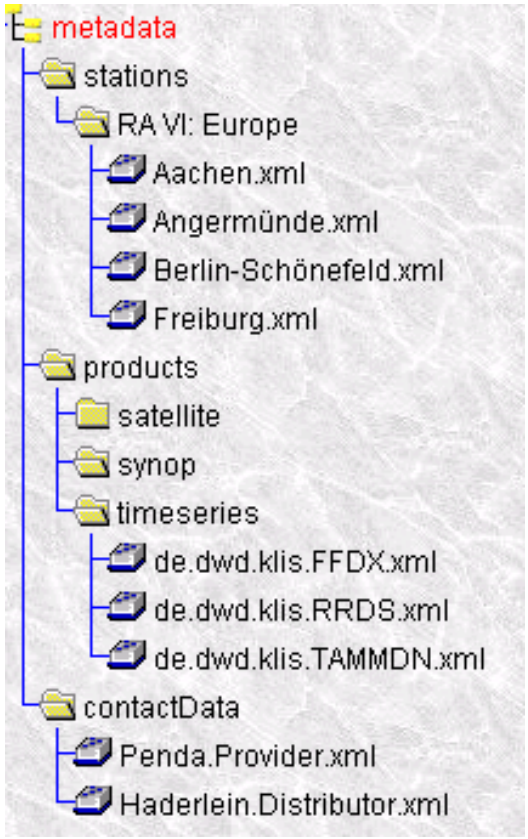
Technical infrastructure of a data centre



Update and Insert Metadata

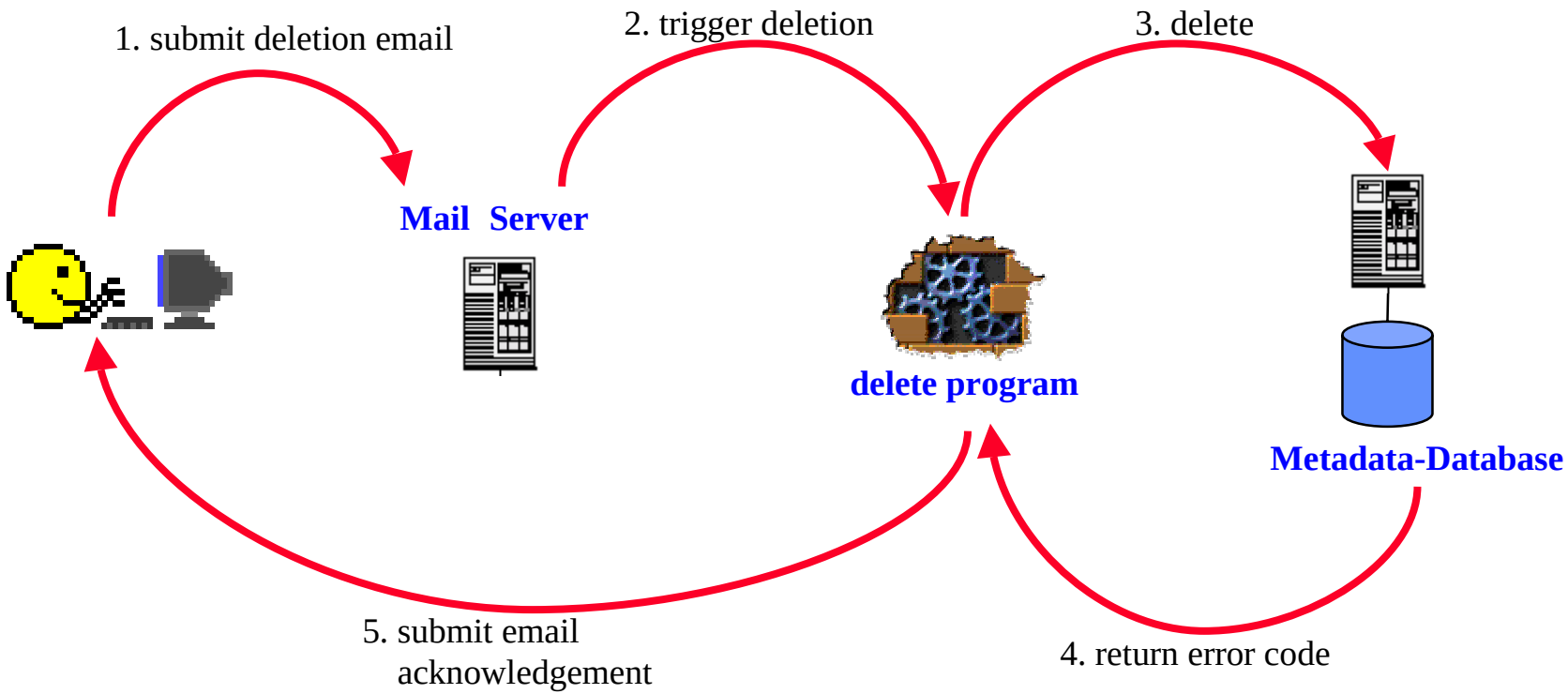


Directory Structure of FTP Account



- a data centre can have stations in more than one WMO region
- the product catalogue is divided into sub-categories
- common agreement is needed for a new product sub-category
- XML document names are unique within each subdirectory
- there is no strict naming convention for XML documents

Delete Metadata



Note: like to unsubscribe from an email list

Summary

- harvesting procedure is
  - simple
  - robust
  - effective
  - safe
- generation and submission of metadata XML documents can be automated at the user side
- possibility to define the date when an update should become effective
- not yet implemented in UNIDART