Toward
CEOS Water Portal 2.0

Yoshiyuki Kudo (JAXA/RESTEC)

WGISS-35, INPE, Brazil

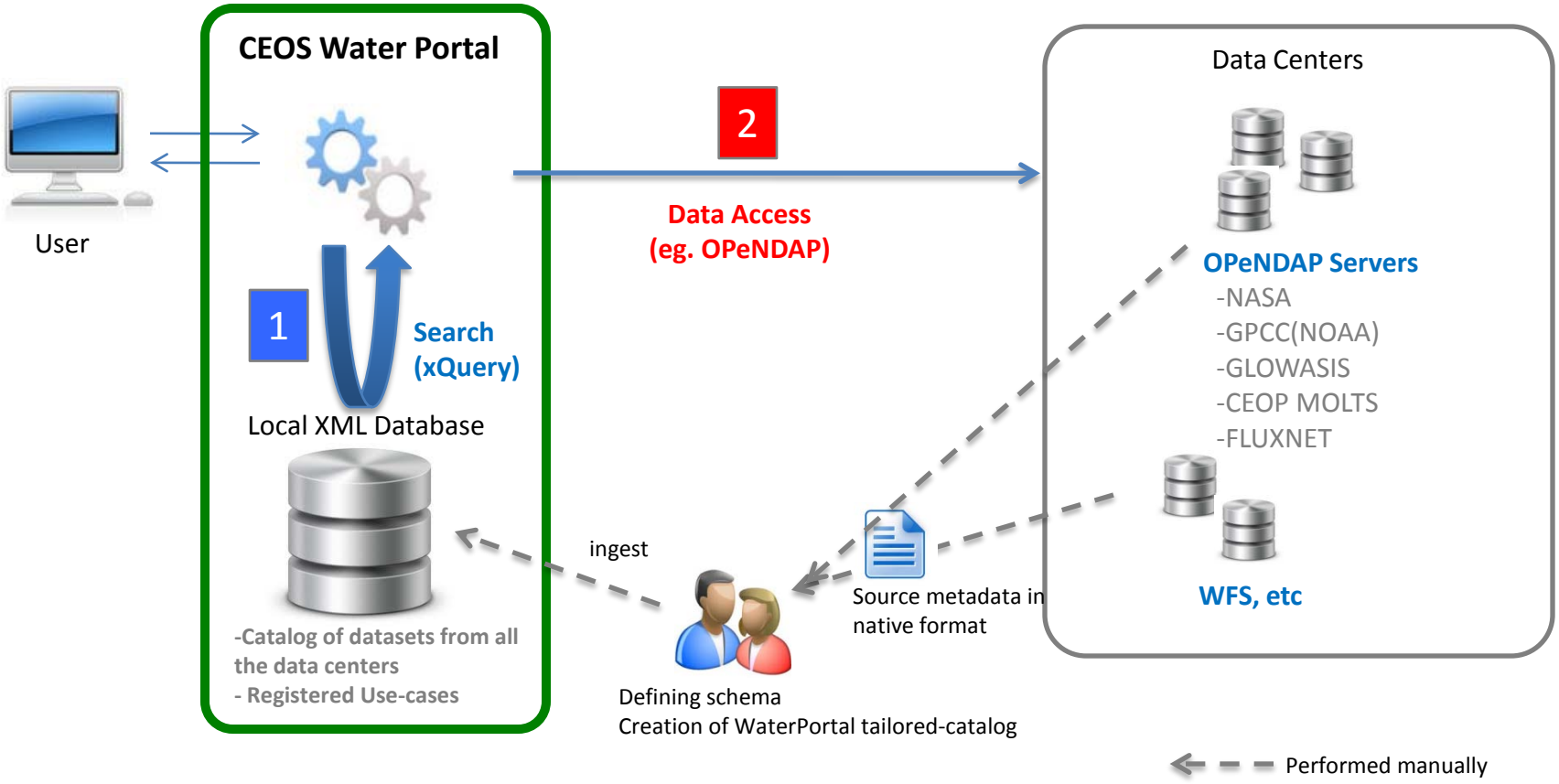
Background

- Catalog creation and management has been labor-intensive
  - Diverse & inconsistent metadata from data partner servers (OPeNDAP, WFS, etc)
  - The portal is planned to become fully operational in another 3 years as part of DIAS[1]
- Need of transition to labor-free/reduced architecture
- Architecture approach based on 2 step search (much like CWIC, ESIP OpenSearch etc.)

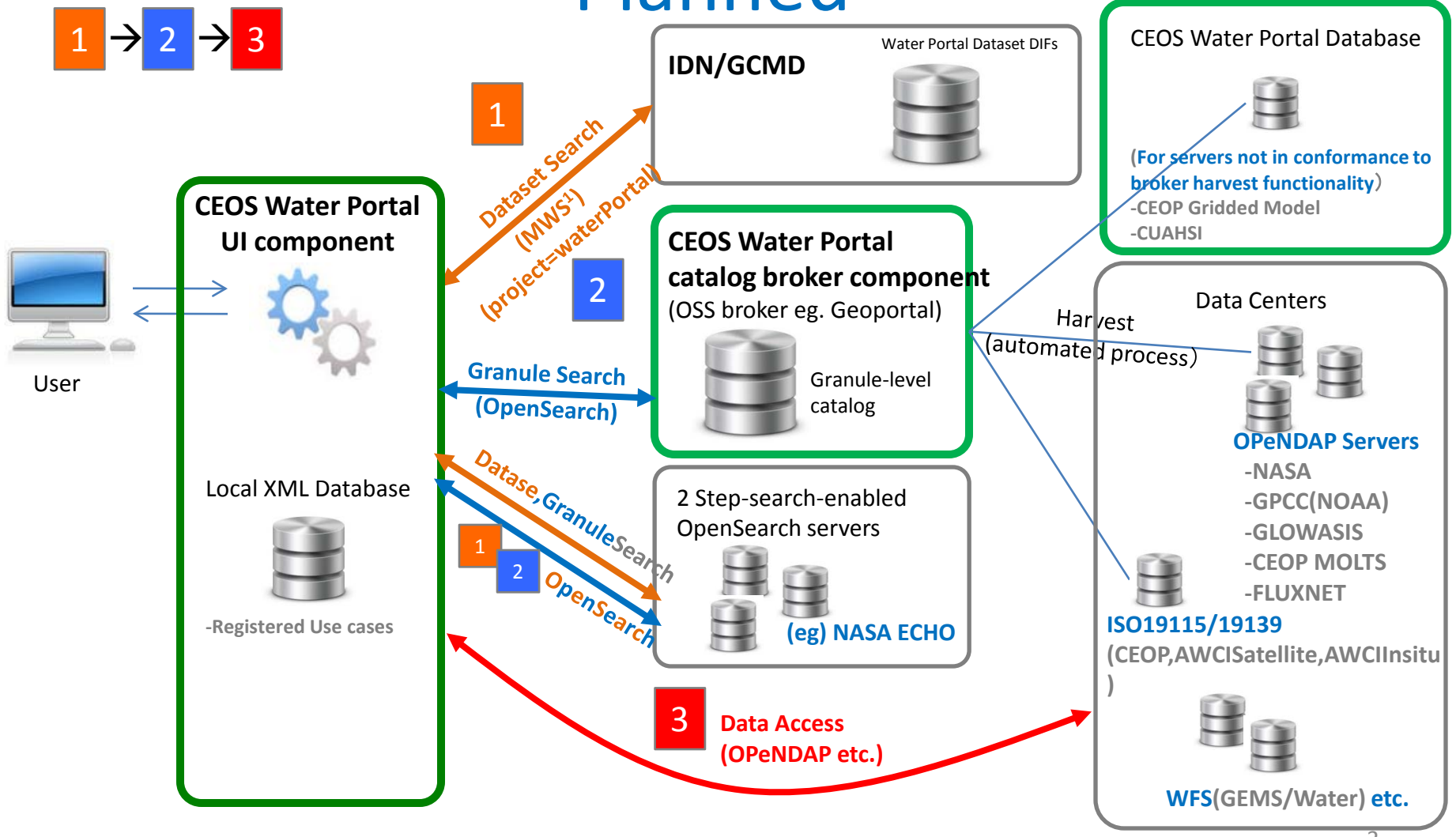
[1] http://www.editoria.u-tokyo.ac.jp/projects/dias/?locale=en

Architecture on Search & Access
-Today-

1 → 2



Architecture on Search & Access -Planned-



*1 MWS: Metadata Web Service by IDN/GCMD

New architecture

- Open, distributed architecture
- Based on 2 step search (Dataset search -> Granule search)
  - Case1 (basic case)
    - Dataset Search : MWS (Metadata Web Service by IDN/GCMD)
    - Granule Search : OpenSearch (CEOS Water Portal catalog broker component)
  - Case 2 (applicable to 2-step OpenSearch-enabled partner servers)
    - Dataset Search : OpenSearch
    - Granule Search : OpenSearch
- For dataset-level catalog, create and ingest DIFs for the entire water portal datasets (except datasets of pattern 2)
- Use OSS (Open Source Software) for brokering the granule level catalog
  - Harvest from each partner servers in an automated fashion
  - Candidate OSS : Esri Geoportal.
- New User Interface
  - Search&access than drill-down
  - Category search by IDN/GCMD Science Keyword as well as ECV variables
  - Support free text search

Mechanism of Dataset->Granule Search Transition for Case 1

- Prearrangements
  - In DIFs, specify project=waterPortal and embed OSDD URL (OpenSearch Description Document) that is applicable to each specific dataset
  - Customize the catalog broker component to be able to recognize the dataset-specific OSDD URL and generate the OSDD.xml dynamically
  - Assure that a search query based on the OSDD URL template returns only the granules of the specific dataset (eg.by using a set of fixed query terms, or insert DIF's EntryID into the broker component database/ response somehow)
- How it works
  - Search IDN/GCMD using MWS with project=WaterPortal
  - Parse the response and get the dataset-specific OSDD
  - Generate and send OpenSearch query for granule-level search

Mechanism of Dataset->Granule Search Transition for Case 2

- Much like Case 1 except it uses OpenSearch for dataset level search
- Much like the way ESIP Discovery Cluster proposes

Hurdles to overcome

1. Performance of catalog broker OSS
   - Search response time
   - Holding and managing tens of millions of granule records
   - Customizable ?
2. DIF creation
   - Can we (portal) make and ingest the partners' dataset DIFs on our own ? (IDN/GCMD policy ?)
   - Effective methodology for creating the large number of DIFs
   - Granularity of a dataset (DIF)
     (eg. CEOP MOLTS ECMWF 2007 Precipitation Dataset, or CEOP MOLTS ECMWF Dataset)
3. Sustainability of current portal's services
   - Subsetting by variable / time (ESIP Discovery group seems to be working on OpenSearch -> OPeNDAP transition mechanizm)

Feasibility study first

- We will do feasibility study through prototyping the new architecture this fiscal year.
- Transition to the new architecture will happen the following year, if it yields satisfying results.

Need support from WGISS

- Endorsement from IDN/GCMD team about leveraging IDN/GCMD in the presented way.
- We would like to solicit the involvement and support of Dr. Chris Lynnes, CEOS Water Portal NASA representative, for helping us establish good communication with Geoportal development team on this ambitious challenge as well as for comments on any aspect.

- Comments ?