5th EGEE User Forum

14 April 2010
Uppsala (Sweden)

Building Scientific Workflows for the Fisheries and Aquaculture Management Community based on Virtual Research Environments

Pedro Andrade (CERN) pedro.andrade@cern.ch

Leonardo Candela (CNR) leonardo.candelia@isti.cnr.it
Andrea Manzi (CERN) andrea.manzi@cern.ch
Pasquale Pagano (CNR) pasquale.pagano@isti.cnr.it

FARM Community

- who is the community
- what are their needs/requirements

D4Science Infrastructure

- management of heterogeneous resources
- creation of virtual research environments

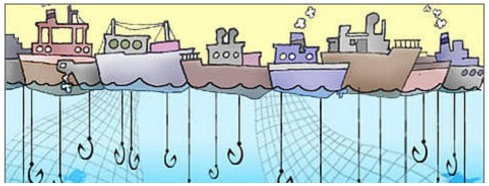
FARM VREs

- multiple environments for different applications
- combining VREs as scientific workflows

Fisheries and **A**quaculture **R**esearch **M**anagement – Experts, scientists and researches from the UN FAO and other related agencies/organizations

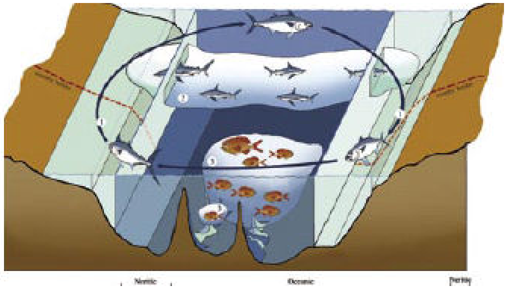
- FAO
- RFBs
- WFC
- GBIF
- CoML / OBIS
- IUCN
- IOC

Fisheries
- Production systems and fleets
- Resources of commercial interest



Biodiversity
- Resources of commercial interest
- Broader ecosystem

Environment
- Broader ecosystem
- Physical environment



The scientific daily activities of this community brought several **challenging requirements** to the D4Science project:

- Collaborative working environments
- Intensive on-demand data processing
- Automatic reporting from templates
- Combine biodiversity information with fisheries time series
- Manipulation of very large time series data (harmonization)
- Spatial dimension and mapping (GIS)
- Consumption of data and derived products through dynamically defined application workflows
- Support for annotations

They need an **infrastructure enabler** system providing **advanced data management** services through dynamically deployed specialized research environments

- Automatically describe and promote harmonization and provenance information
- Promote re-usage of data sets and derived products by different applications by offering rich publication and discovery tools
- Can be easily enriched by community specific applications
- Registration, publication, and discovery of applications as simple as registration, publication, and discovery of data

D4Science-II provides an **ecosystem of e-Infrastructure** resources based on the concept of Virtual Research Environment. A VRE can be describe as:

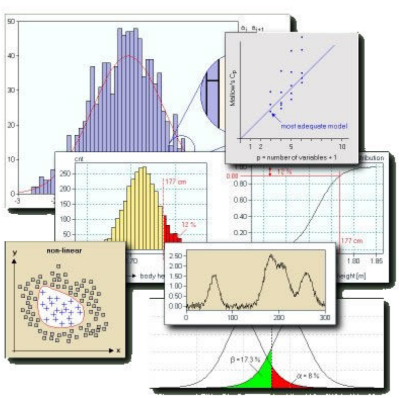
*A distributed and **dynamically** created **environment,***
*where a subset of data, services, and hardware **resources**,*
*are dynamically assigned to a subset of **users,***
*for a **limited timeframe,***
*at little or **no cost** for the providers of the infrastructure.*

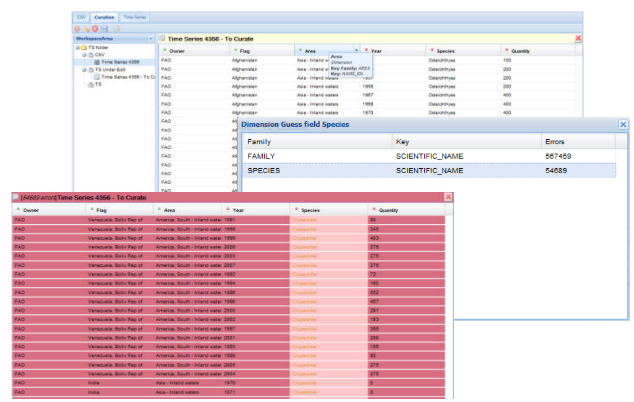
VREs are enriched with facilities for communication, collaboration, and sharing among scientists and researchers.

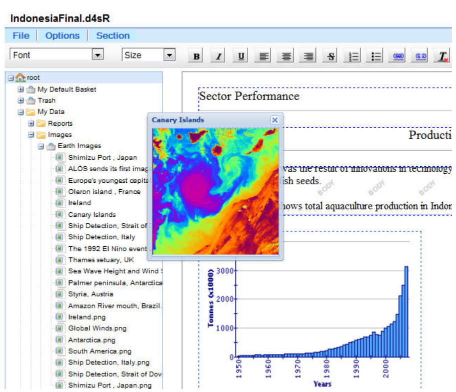
D4Science Infrastructure

A VRE supports cooperative activities such as:

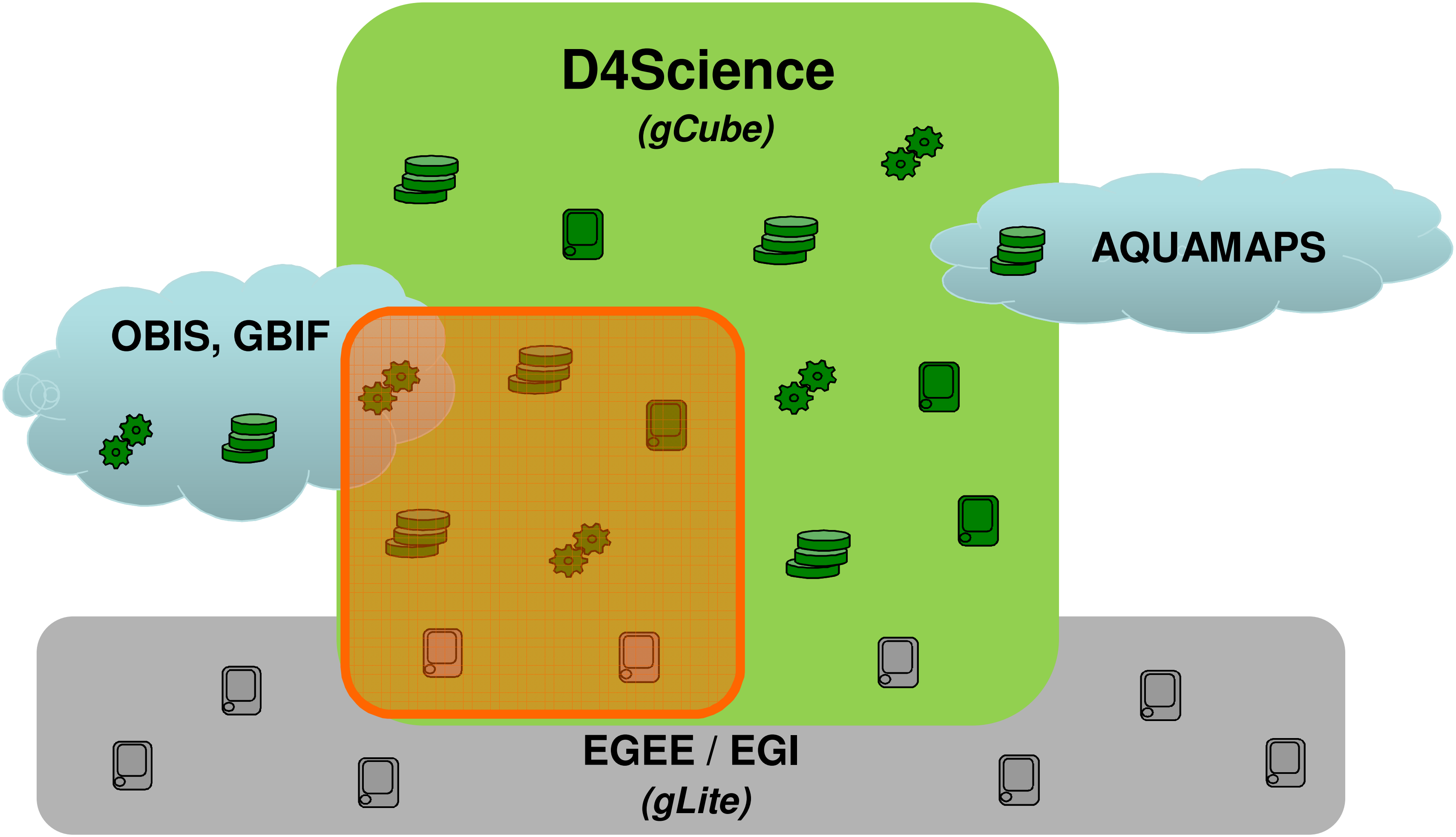
- Data analysis and processing
- Metadata and provenance generation
- Data generation, integration, enrichment, and curation
- Processes execution and optimization
- Registration and execution of application specific tools







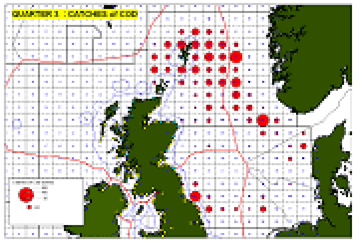
D4Science Infrastructure

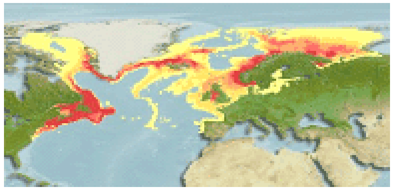


FARM VREs

Three end user environments modelled as D4Science VREs

- Species Prediction Modeling – **AquaMaps** VRE
- Fishery Country Profiles – **FCPPS** VRE
- Integrated Catch IS – **ICIS** VRE

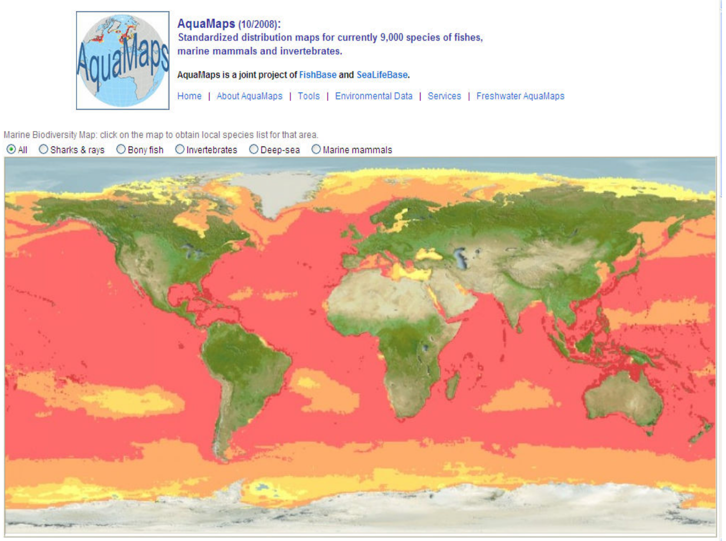




comprehensive profiles	Fishing activity / Catch	Species occurrence maps
Fisheries	**Fisheries**	**Biodiversity**
Fishery Country Profiles	Integrated Catch Information System	Species Prediction Modeling

AquaMaps VRE

Objective: Improve the production of species distribution and biodiversity maps for on-line presentation



Features:

- Parametric generation of species distribution and biodiversity maps
- Improved generation time

One Multispecies map computed on 6,188 half degree cells (over 170k) and 2,540 species requires 125 millions computations

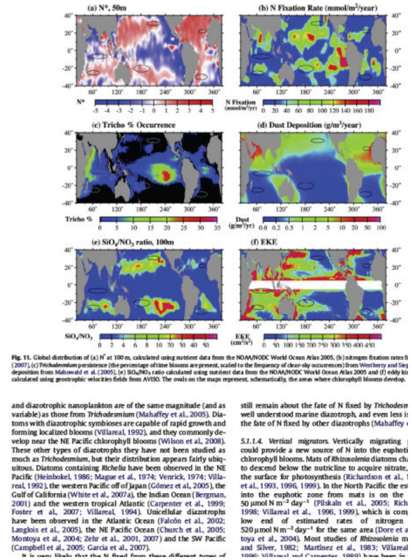
One global map (extended to all species and cells around the world) requires about 400 billions computations

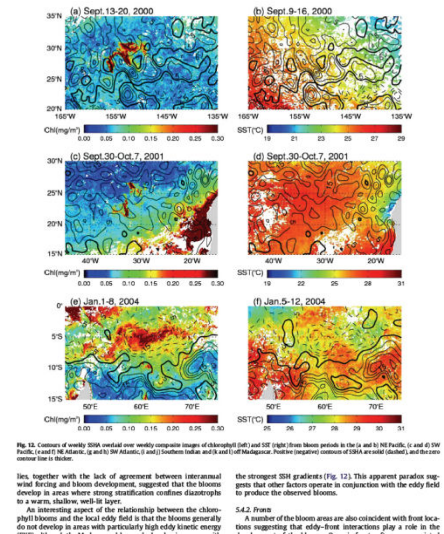
FCPPS VRE

Objective: Increase visibility and accuracy of country profiles

Features:

- Improve timeliness
- Life-cycle support (draft, review)
- Reuse data and report-templates
- Data integration across boundaries
- Document production system





Reports include maps and graphs computed on-demand and updated directly in the report upon request

Country-specific templates integrating 10 years of time series catch observations

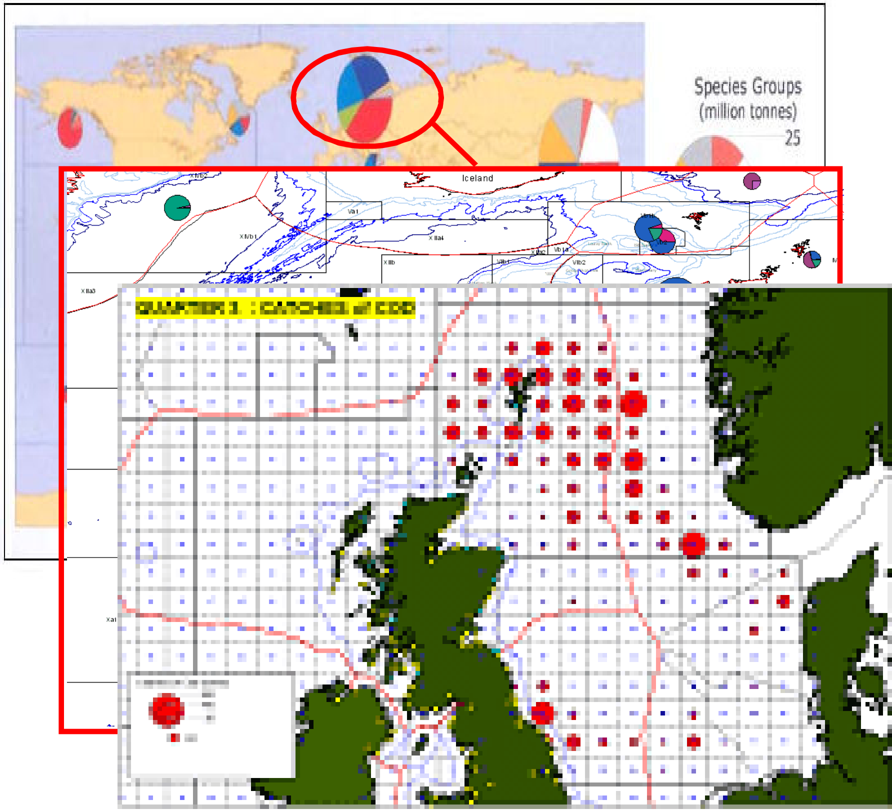
ICIS VRE

Objective: Improve the management of capture statistics data

Features:

- Share time series
- Filter, merge, join data
- Facilitate import and curation
- Harmonized and reallocated catch statistics

Figure A1.5 - World marine catches, main species groups by major marine fishing areas in 2002



10 years catch data containing millions of observation points described by ~ 100 attributes

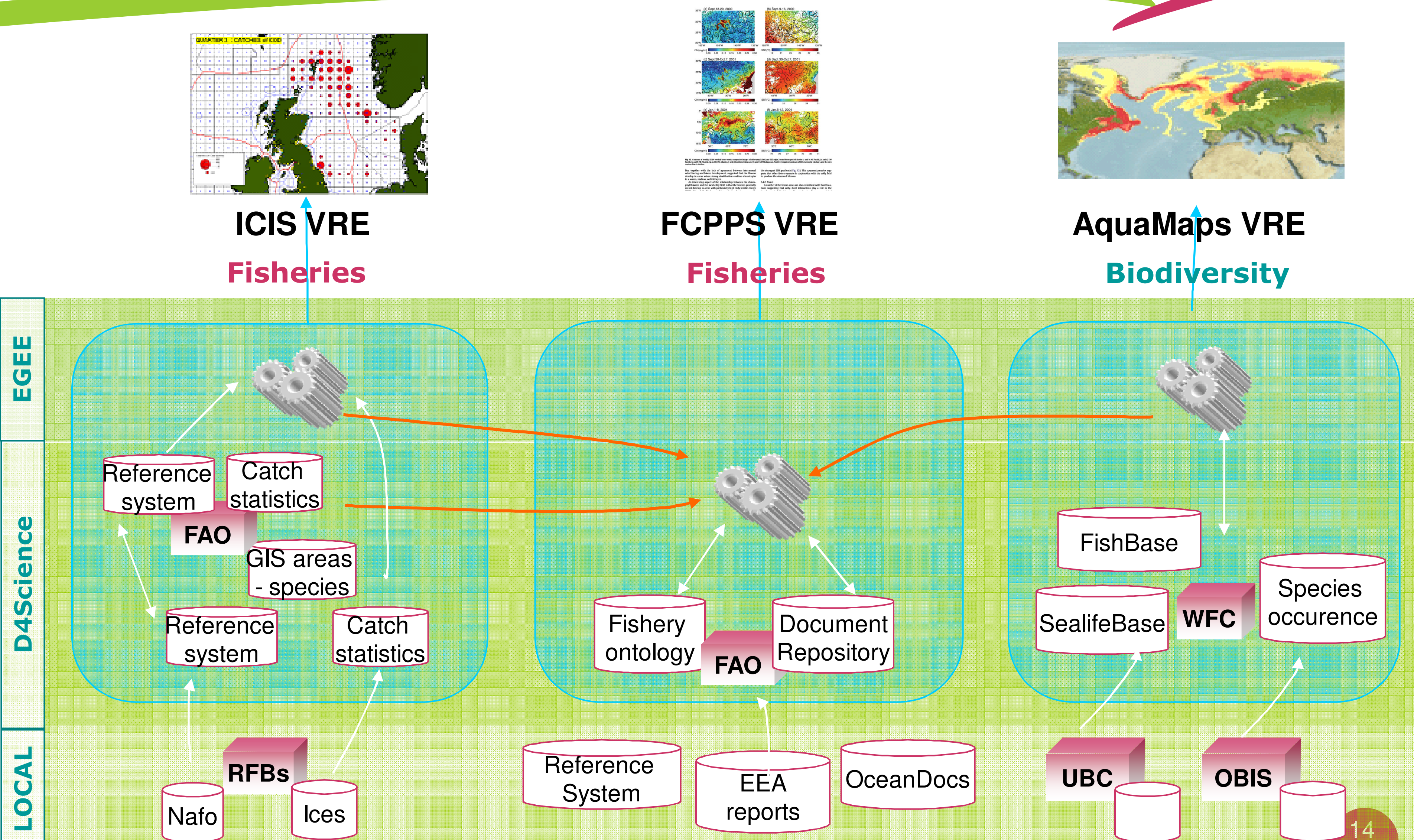
These VREs satisfy **different end-user scenarios**:

- Requested by different FARM groups/users
- Accessed from distributed locations at the same time

However they are based on a **common pool of resources**:

- Data collections are shared, enhanced, and re-used by multiple applications
- Same distributed computing and storage facilities are dynamically allocated to different tasks
- The set of base services enhance the VREs with full text search, content browsing, annotation, metadata management, etc

FARM VREs


ICIS VRE
Fisheries
FCPPS VRE
Fisheries
AquaMaps VRE
Biodiversity
EGEE
D4Science
LOCAL
Reference system
Catch statistics
FAO
GIS areas - species
Reference system
Catch statistics
RFBs
Nafo
Ices
Fishery ontology
FAO
Document Repository
Reference System
EEA reports
OceanDocs
FishBase
SealifeBase
WFC
Species occurence
UBC
OBIS

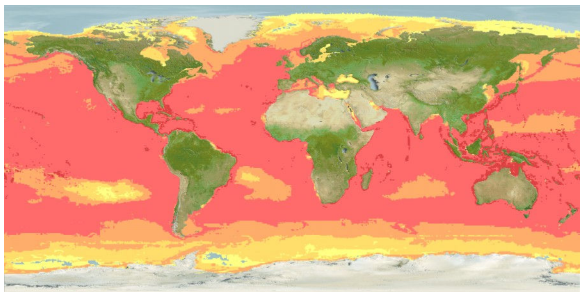
Conclusions

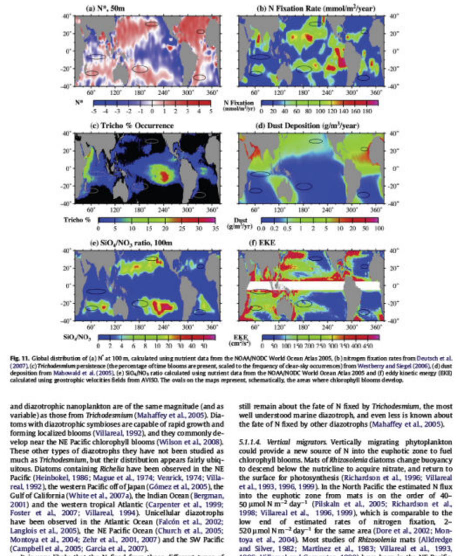
D4Science provides VREs to the FARM community where different users **integrate** and **process** a common set of heterogeneous data collections using distinct applications

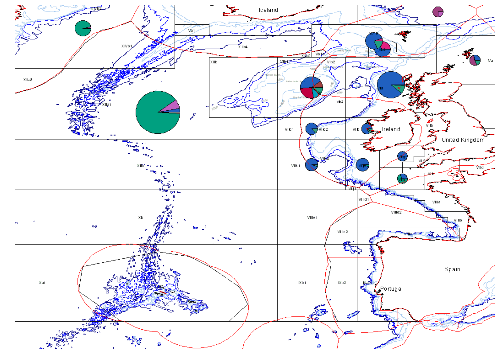
D4Science fosters the collaboration among FARM community members by promoting the **sharing** of **workflows** and data

FARM VREs are currently running in the D4Science infrastructure and **being exploited** by several FAO bodies

Questions ?







www.d4science.eu

www.gcube-system.org