



d4SCIENCE

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.candela@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

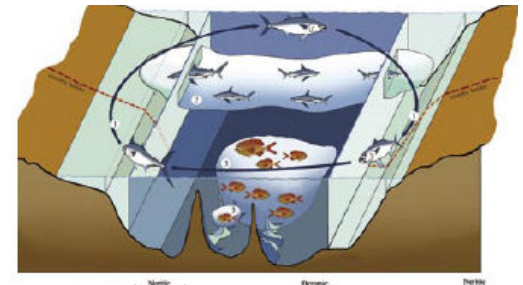
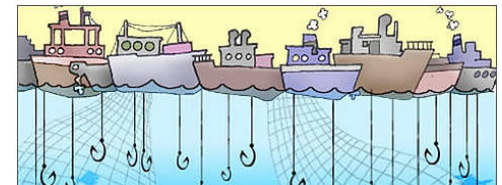
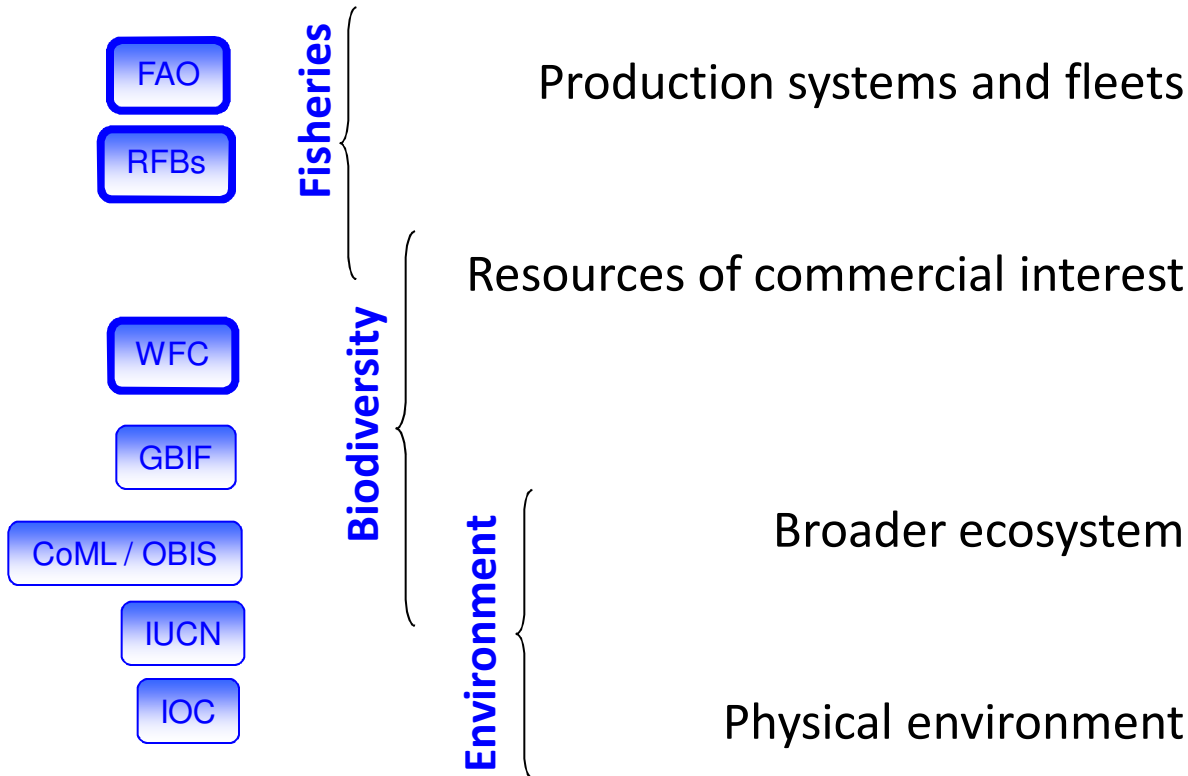
D4Science Infrastructure

- 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



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

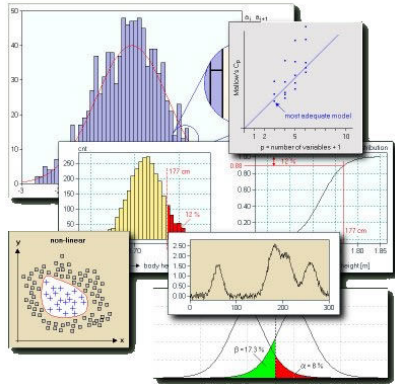
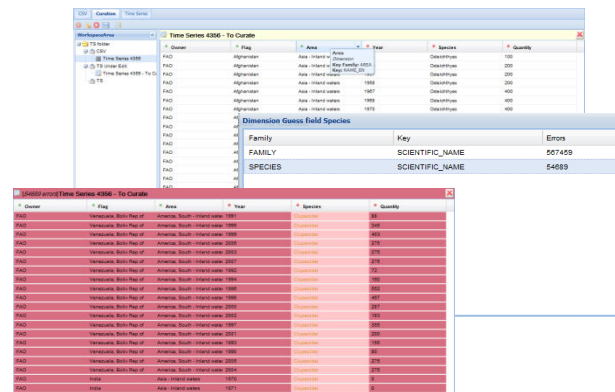
D4Science-II provides an **ecosystem of e-Infrastructure** resources based on the concept of Virtual Research Environment. A VRE can be describe as:

*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.

A VRE supports cooperative activities such as:

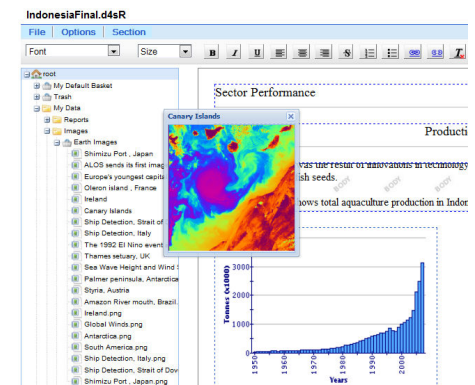
- Data analysis and processing
- Metadata and provenance generation
- Data generation, integration, enrichment, and curation
- Processes execution and optimization
- Registration and execution of application specific tools

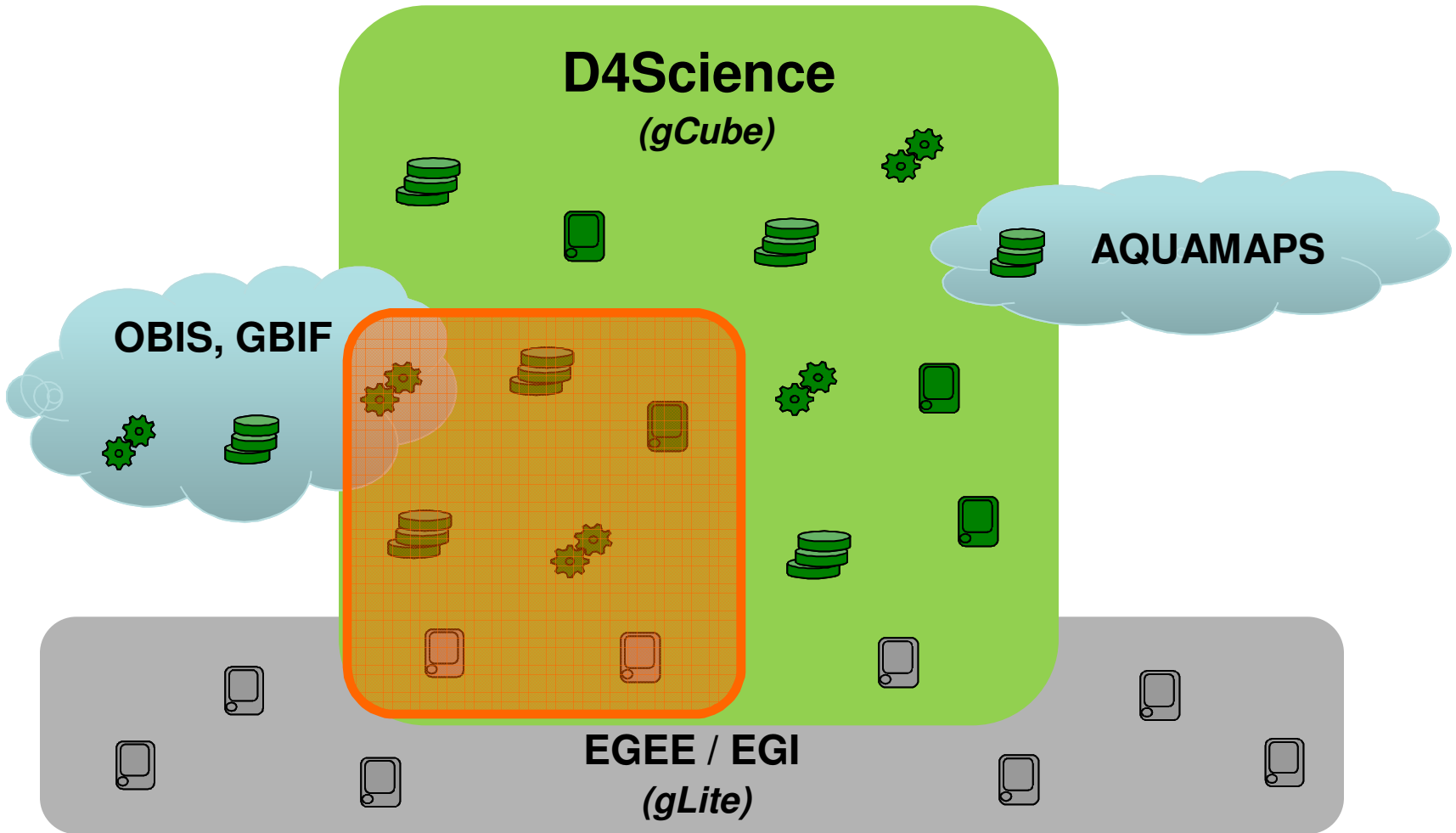



Dimension Guess field Species

Family	Key	Errors
FAMILY	SCIENTIFIC_NAME	567459
SPECIES	SCIENTIFIC_NAME	54889

Owner	Flag	Year	Species	Quantity
FAD	Venezuela, Boliv. Rep. of	1981	Chlorophyll a	80
FAD	Venezuela, Boliv. Rep. of	1982	Chlorophyll a	356
FAD	Venezuela, Boliv. Rep. of	1983	Chlorophyll a	463
FAD	Venezuela, Boliv. Rep. of	1984	Chlorophyll a	278
FAD	Venezuela, Boliv. Rep. of	1985	Chlorophyll a	278
FAD	Venezuela, Boliv. Rep. of	1986	Chlorophyll a	172
FAD	Venezuela, Boliv. Rep. of	1987	Chlorophyll a	190
FAD	Venezuela, Boliv. Rep. of	1988	Chlorophyll a	387
FAD	Venezuela, Boliv. Rep. of	1989	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	1990	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	1991	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	1992	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	1993	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	1994	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	1995	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	1996	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	1997	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	1998	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	1999	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2000	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2001	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2002	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2003	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2004	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2005	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2006	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2007	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2008	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2009	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2010	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2011	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2012	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2013	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2014	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2015	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2016	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2017	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2018	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2019	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2020	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2021	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2022	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2023	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2024	Chlorophyll a	397
FAD	Venezuela, Boliv. Rep. of	2025	Chlorophyll a	397





Three end user environments modelled as D4Science VREs

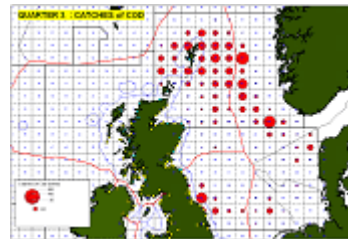
- Species Prediction Modeling – **AquaMaps** VRE
- Fishery Country Profiles – **FCPPS** VRE
- Integrated Catch IS – **ICIS** VRE



comprehensive profiles

Fisheries

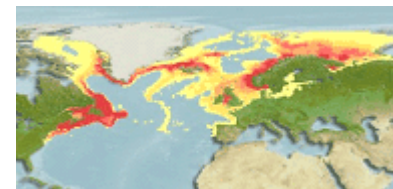
Fishery
Country
Profiles



Fishing activity / Catch

Fisheries

Integrated Catch
Information
System



Species occurrence maps

Biodiversity

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



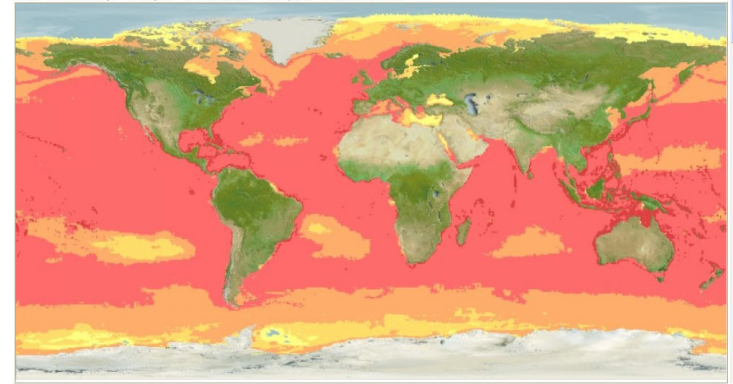
AquaMaps (10/2008):
Standardized distribution maps for currently 9,000 species of fishes,
marine mammals and invertebrates.

AquaMaps is a joint project of FishBase and SeaLifeBase.

[Home](#) | [About AquaMaps](#) | [Tools](#) | [Environmental Data](#) | [Services](#) | [Freshwater AquaMaps](#)

Marine Biodiversity Map: click on the map to obtain local species list for that area.

All Sharks & rays Bony fish Invertebrates Deep-sea Marine mammals



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

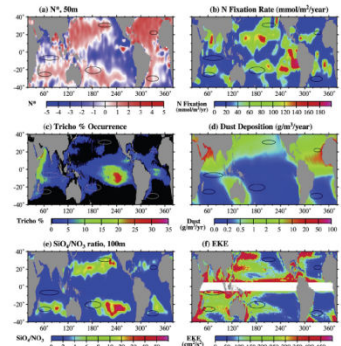


Fig. 11. Global distribution of (a) N fixation rate, (b) N fixation rate, (c) Trichloro-14C occurrence, (d) Dust deposition, (e) SO₄/NO₃ ratio, and (f) EKE. The maps are calculated using remote data from the NOAA/NC World Ocean Atlas 2005 and 15° only to calculate using geographic reference data from AOD50. The maps are schematic, but areas where chlorophyll bloom develop.

and diatomic nanoplankton are of the same magnitude (and as variable as those from Trichodesmium (Mallory et al., 2001). Diatoms with diatom-like symbionts are capable of rapid growth and forming localized blooms (Villalón, 1992), and they commonly develop near the NE Pacific chlorophyll blooms (Wilson et al., 2008). These other types of diatoms they have not been studied as much as Trichodesmium, but their distribution appears fairly ubiquitous. Diatoms containing *Rhodospira* have been observed in the NE Pacific (Chisholm, 1980; Mague et al., 1974; Yocum, 1974; Villalón, 1992), the western Pacific off of Japan (Gonzalez et al., 2005), the Gulf of California (White et al., 2007a), the Indian Ocean (Dreyfus, 2001) and the western tropical Atlantic (Carapane et al., 1999; Foster et al., 2007; Villalón, 1994). Unicellular diatoms have been observed in the Atlantic Ocean (Chaurv et al., 2002; Langlois et al., 2005), the NE Pacific Ocean (Chaurv et al., 2002; Montoya et al., 2004; Zehr et al., 2001, 2007) and the SW Pacific (Carpenter et al., 2005; Garcia et al., 2007).

It is common to think that the N fixed from these different sources of N still remain about the fate of N fixed by Trichodesmium well understood in the diatom, and even less is the fate of N fixed by other diatoms (Mallory et al., 2001).

5.1.4. Vertical migration. Vertically migrating could provide a new source of N into the euphotic chlorophyll blooms. Many of Rhizosolenia diatoms die to descend below the nitricline to acquire nitrate, the surface for photosynthesis (Richardson et al., 1991; Richardson et al., 1995, 1996, 1999), in the North Pacific the surface for photosynthesis (Richardson et al., 1991; Richardson et al., 1995, 1996, 1999), which is common low end of estimated rates of nitrogen (500 $\mu\text{mol N m}^{-2} \text{ day}^{-1}$ for the same area (Dove et al., 2004; Zehr et al., 2001, 2007) and the SW Pacific (Carpenter et al., 2005; Garcia et al., 2007).

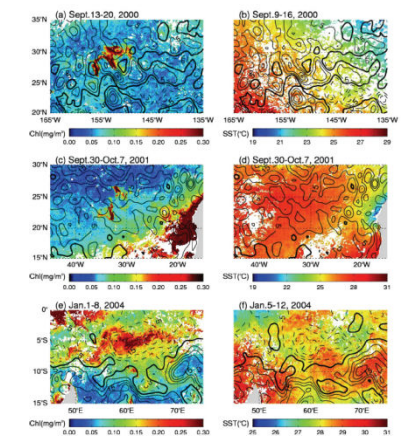


Fig. 12. Contours of weekly SST (red) over weekly composite images of chlorophyll (chl) (left) and SST (right) from bloom periods in the (a) and NE Pacific, (c) and (d) NE Pacific, (e) off NE Atlantic, (f) off NE Atlantic, and (f) off NE Atlantic. Positive (negative) contours of SST are solid (dashed), and the contours are in 1°C.

ies, together with the lack of agreement between interannual wind forcing and bloom development, suggested that the blooms develop in areas where strong stratification confines diatoms to a warm, shallow, well-lit layer. An interesting aspect of the relationship between the chlorophyll (chl) blooms and the local eddy field is that the blooms generally do not develop in areas with particularly high eddy kinetic energy (EKE). The strongest EKE gradients (Fig. 12). This apparent paradox suggests that other factors operate in conjunction with the eddy field to produce the observed blooms. A number of the bloom areas are also coincident with front locations, suggesting that eddy-front interactions play a role in the

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

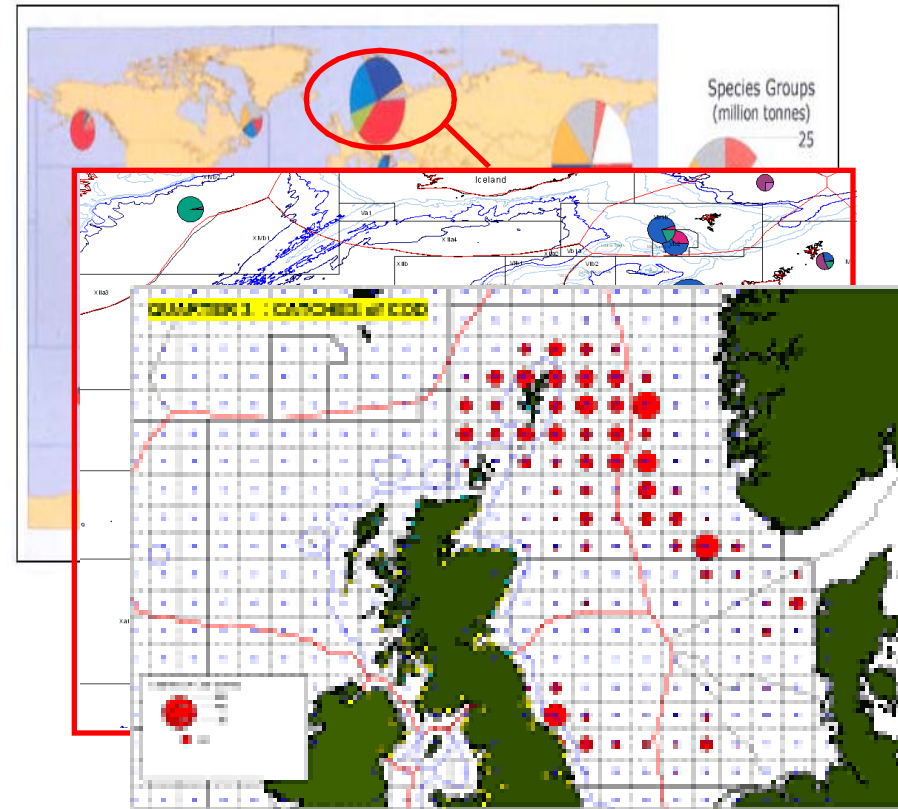
ICIS VRE

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



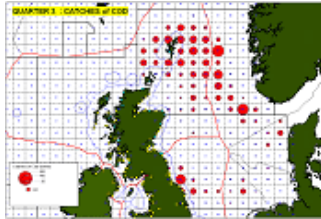
10 years catch data containing millions of observation points described by ~ 100 attributes

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

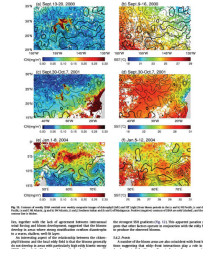
- Requested by different FARM groups/users
- Accessed from distributed locations at the same time

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

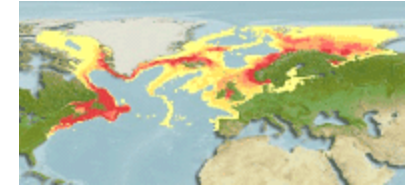
- Data collections are shared, enhanced, and re-used by multiple applications
- Same distributed computing and storage facilities are dynamically allocated to different tasks
- The set of base services enhance the VREs with full text search, content browsing, annotation, metadata management, etc



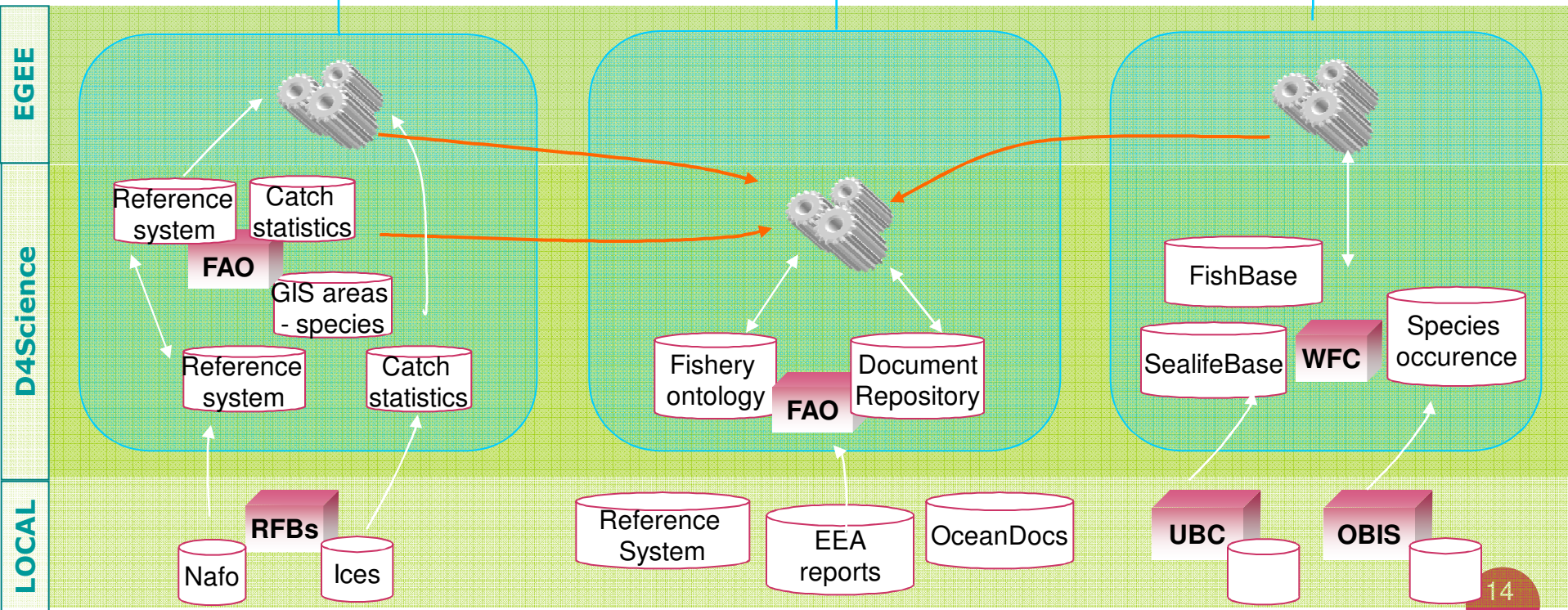
ICIS VRE
Fisheries



FCPPS VRE
Fisheries



AquaMaps VRE
Biodiversity



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

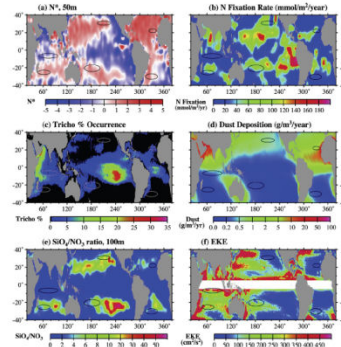
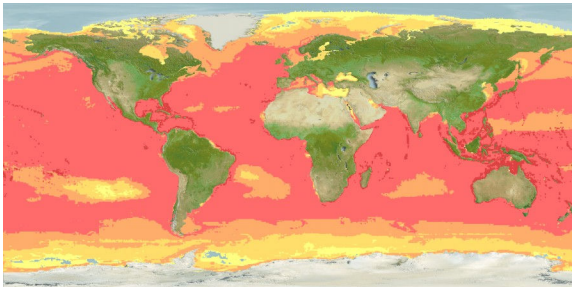
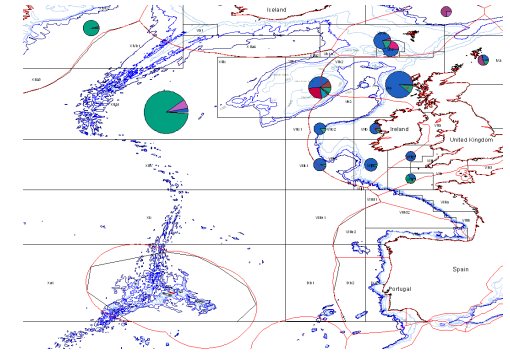


Fig. 11. Global distribution of (a) N^* at 50m, calculated using nutrient data from the NCAR/NOCC World Ocean Atlas 2005. (b) Nitrogen fixation rates from Donath et al. (2007). (c) Trichodesmium presence (the percentage of time biomass is present), scaled by the frequency of the daily occurrence from Strickland and Partridge (1990). (d) Dust deposition from Mahowald et al. (2005). (e) SiO_2/NO_3 ratio calculated using nutrient data from the NCAR/NOCC World Ocean Atlas 2005 and (f) eddy kinetic energy (EKE) calculated using geostrophic velocity fields from AVISO. The scale on the maps represents schematically the area where chlorophyll biomass doubling...

and diazotrophic zooplankton are of the same magnitude (and as variable as those from Trichodesmium (Mahaffey et al., 2005)). Diatoms with diazotrophic symbioses are capable of rapid growth and forming localized blooms (Villalón, 1992), and they commonly develop near the NE Pacific chlorophyll blooms (Wilson et al., 2008). These other types of diazotrophs they have not been studied as much as Trichodesmium, but their distribution appears fairly ubiquitous. Diatoms containing *Richelia* have been observed in the NE Pacific (Chinowski, 1988; Maguer et al., 1974; Yonck, 1974; Villalón, 1992), the western Pacific off of Japan (Gómez et al., 2005), the Gulf of California (Wilce et al., 2007a), the Indian Ocean (Srinivasan, 2001) and the western tropical Atlantic (Carpenter et al., 1999; Foster et al., 2007; Villalón, 1994). Unicellular diazotrophs have been observed in the Atlantic Ocean (Falcón et al., 2002; Langlois et al., 2005), the NE Pacific Ocean (Church et al., 2005; Montoya et al., 2004; Zehr et al., 2007) and the SW Pacific (Campbell et al., 2005; Garcia et al., 2007).

It is more likely that the N fixed from these different sources of still remain about the fate of N fixed by Trichodesmium, the most well understood marine diazotroph, and even less is known about the fate of N fixed by other diazotrophs (Mahaffey et al., 2005).

5.1.1.4. Vertical migrators. Vertically migrating phytoplankton could provide a new source of N into the euphotic zone to fuel chlorophyll blooms. Many of Rhizosolenia diatoms change buoyancy to descend below the nutricline to acquire nitrate, and return to the surface for photosynthesis (Richardson et al., 1996; Villalón et al., 1993, 1996, 1999). In the North Pacific the estimated N flux into the euphotic zone from such is on the order of $40-50 \mu mol N m^{-2} day^{-1}$ (Pridmore et al., 2005; Richardson et al., 1996; Villalón et al., 1996, 1999), which is comparable to the low end of estimated rates of nitrogen fixation, $2-500 \mu mol N m^{-2} day^{-1}$ for the same area (Dore et al., 2002; Montoya et al., 2004). Most studies of Rhizosolenia mats (Gildebrand and Silver, 1982; Martínez et al., 1983; Villalón et al., 1993, 1996; Villalón and Cavanaugh, 1998) have been in the NE Pacific.



www.d4science.eu

www.gcube-system.org