



4th EGEE User Forum/OGF 25 and
OGF Europe's
2nd International Event

Le Ciminiere, Catania, Sicily, Italy
2-6 March 2009

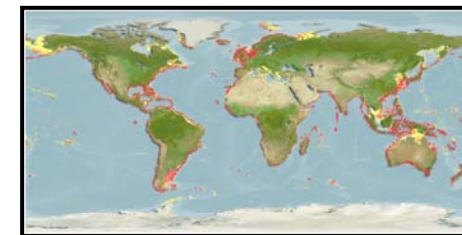
On the D4Science Approach Toward AquaMaps Richness Maps Generation

Pasquale Pagano - CNR-ISTI
pasquale.pagano@isti.cnr.it

Pedro Andrade – CERN
pedro.andrade@cern.ch

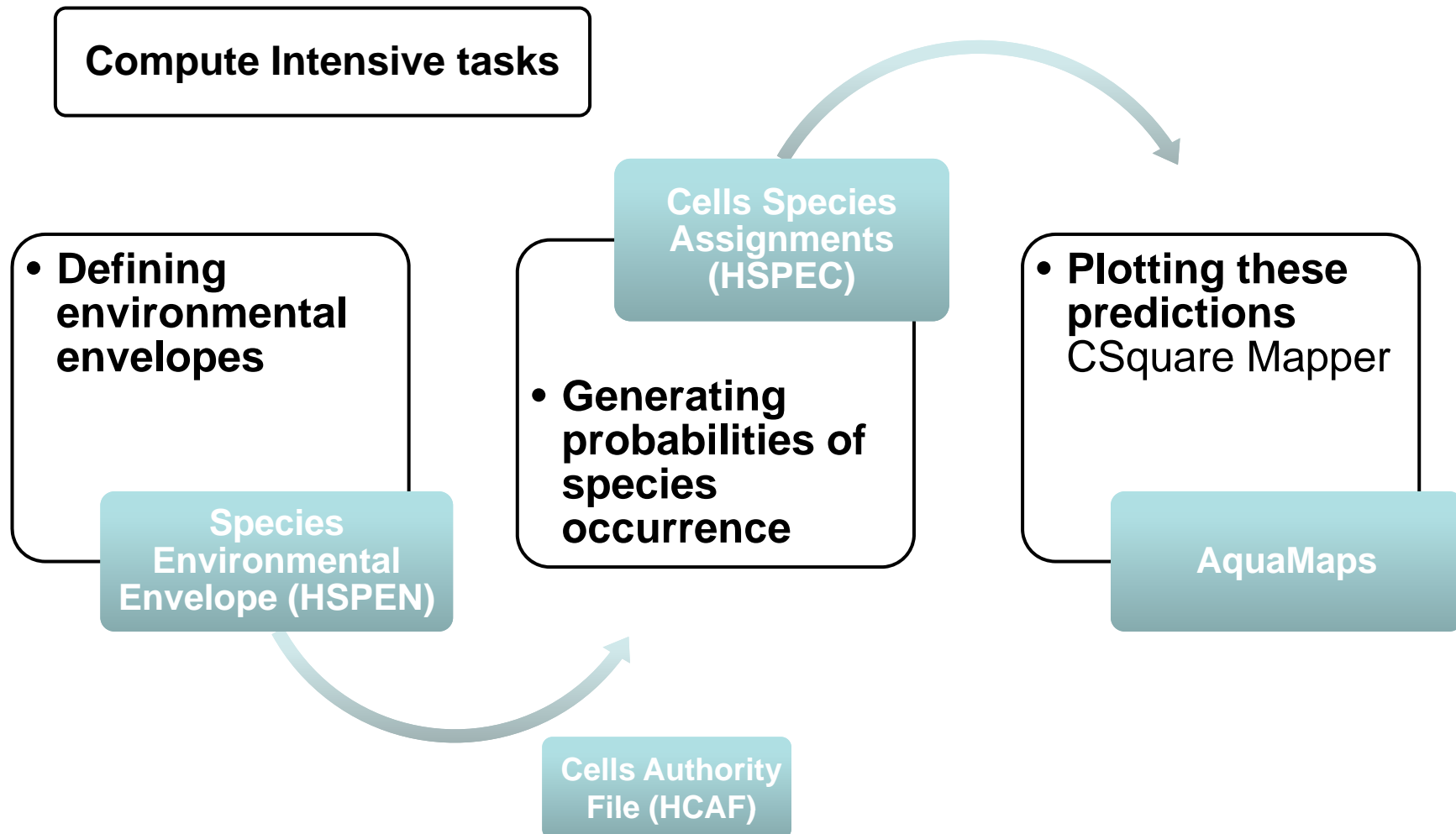


- Model-based, large-scale predictions of known natural **occurrence of marine species**
- Originally developed by Kashner *et al.* (2006) to predict global distributions of marine mammals
- Predictions made by matching species tolerances (environmental envelope) against local environmental conditions
- **Color-coded species range map**, using a half-degree latitude and longitude dimensions
- Supplements existing occurrence data



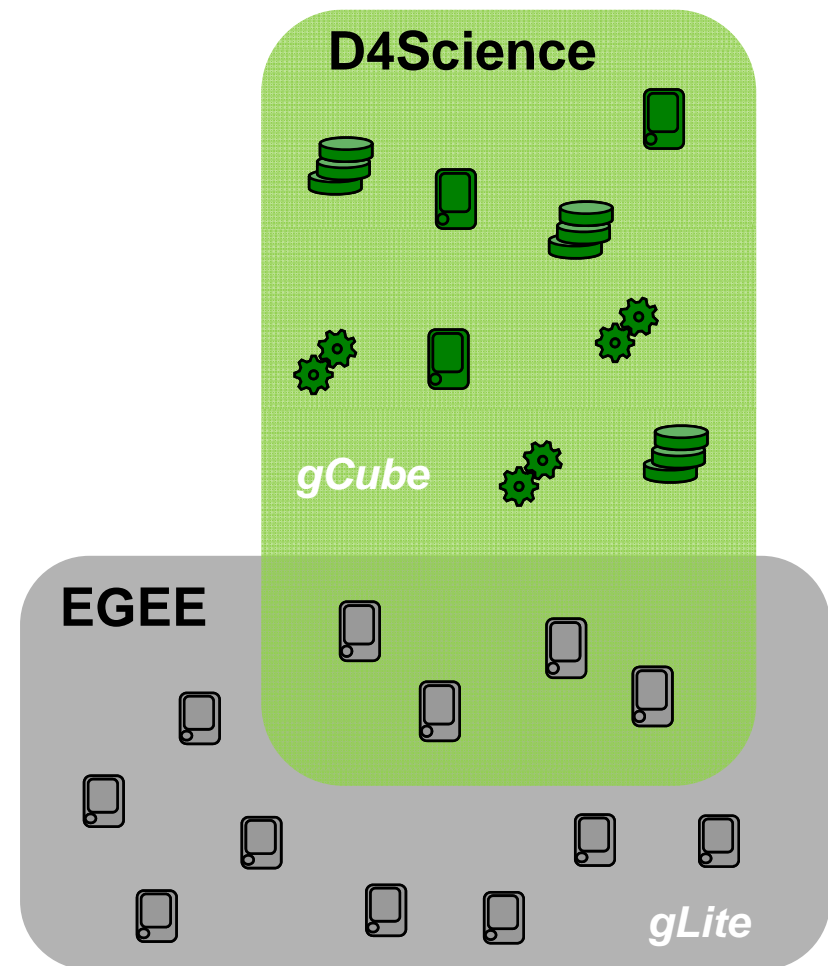
- Web based interface @ www.aquamaps.org
- On-demand generation of maps by implementing an environmental envelope type modeling approach
 1. **Defining environmental envelopes** describing the tolerance of a species w.r.t. each *environmental parameter* (e.g. sea temperature, salinity)
 2. **Generating probabilities of species occurrence** by matching the species environmental envelope against local environmental conditions
 3. **Plotting these predictions** to document large-scale and long-term presence of a species

- Relational DB containing tables for:
 - Species Environmental Envelope (HSPEN)
 - Range of environmental tolerance and preference of a species
 - Cells Authority File (HCAF)
 - Metadata about half degree cells
 - E.g. name, membership (FAO, EEZ), physical attributes (salinity)
 - Cells Species Assignments (HSPEC)
 - Probability of occurrence of a species in a given cell

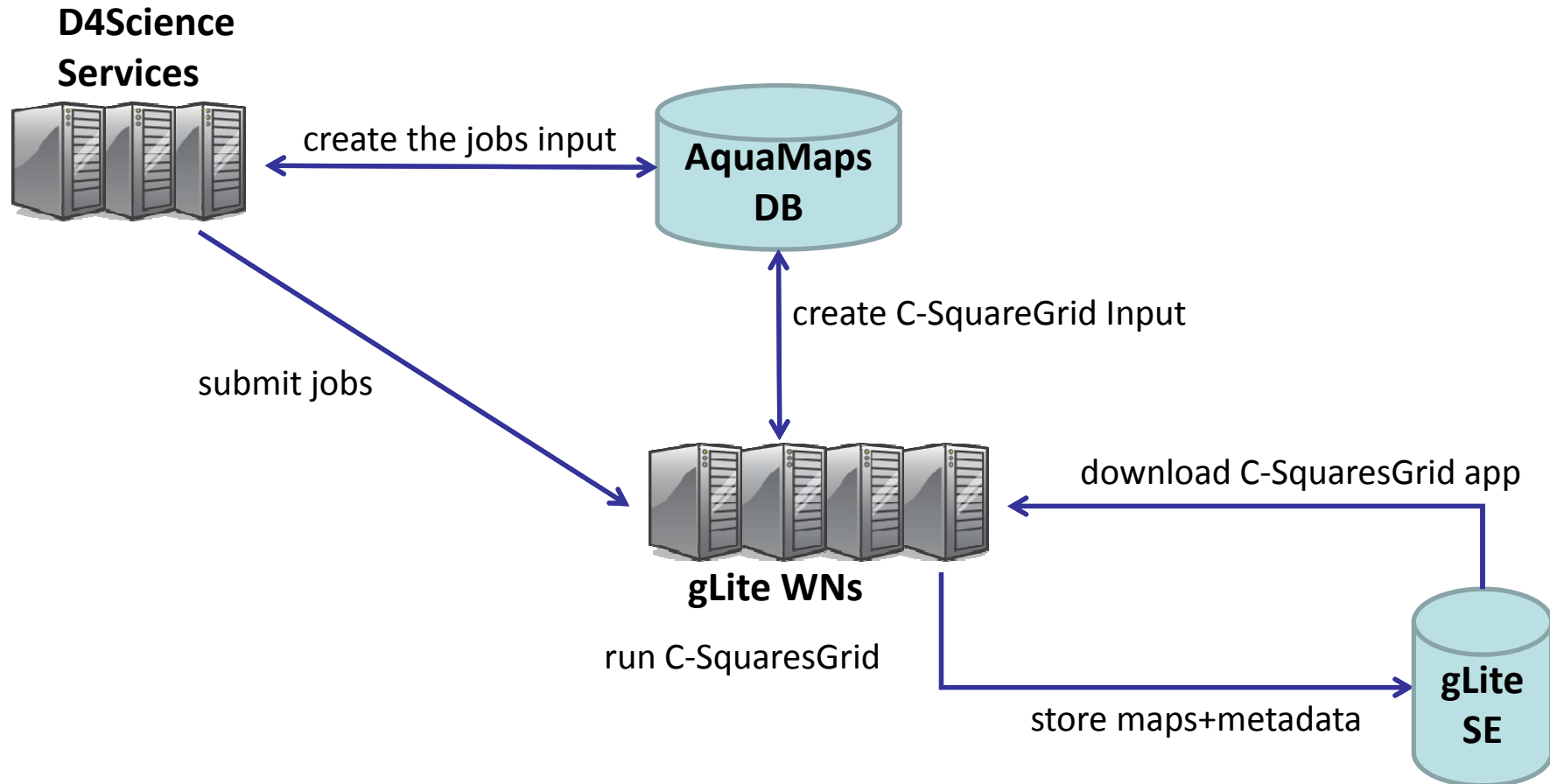


- Difficult Periodic Maintenance
 - Size of the map data, mainly HSPEC
 - Re-generate all data files:
 - every time a species is included or excluded
 - Re-generate maps every:
 - every time the algorithm for “probability of occurrence” is updated
- Limited Performance
 - Huge number of combinations of different richness maps
 - species, family, country, ecosystem, bounding box, etc.

- D4Science provides an e-Infrastructure based on the concept of Virtual Research Environments (VREs).
- These VREs bring together different resources (data collections, services, computing and storage) to serve particular needs of distributed VOs.



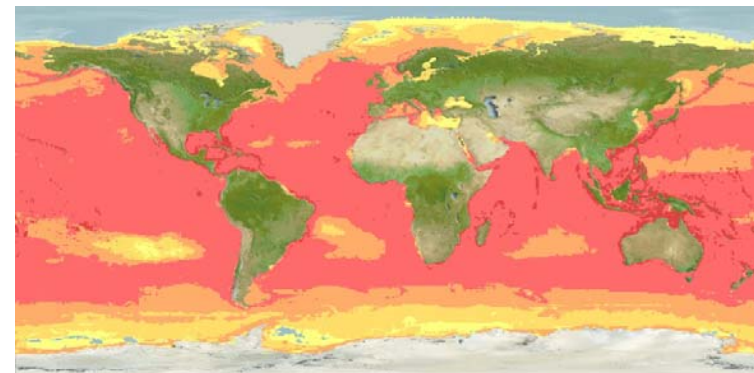
- Direct Grid Execution (DONE)
 - Exploit the grid to generate and store AquaMaps
 - i.e. replace “on-demand generation” with “search for” maps
- VRE Grid Execution (ONGOING)
 - Support the whole workflow including maintenance
 - Generate data from different predictive algorithms
 - Alternative predictive maps will coexist, support for comparison and validation;



- C-SquaresGrid Mapper is the application resulting of the porting of C-Square Mapper to the gLite environment
 - C-Square Mapper
 - a web-oriented perl-based utility
 - plotting dataset on a 2D base maps starting from a csquares string
 - C-SquaresGrid Mapper
 - standalone perl application executed as gLite job
 - accept csquare input string as files
 - generate global 3D maps also by using Xplanet
 - generate AquaMaps metadata and provenance data
 - store the products to gLite SE

- AquaMaps are generated by:
 - Species, Family, Order, Class, Phylum
- Each AquaMaps grid job elaborates 50 Species distribution maps (or Class, Family)
 - Data are obtained by aggregating Species data and distributing these probability data in 5 clusters
 - C-SquareGridMapper is provided with files containing 5 csquare string for each AquaMaps
- Each AquaMaps grid job produces:
 - one 2D map and 13 Global 3D views (for each Species)
 - provenance data and metadata

- Data Processed (1st round)
 - 6086 Species Maps
 - ~ 80000 products (5.3 GB of images+metadata)
 - 592 Family Maps
 - ~11000 products (673 MB)
 - 132 Order Maps
 - ~2500 products (127 MB)
 - 39 Class Maps
 - ~700 products (50 MB)
 - 13 Phylum Maps
 - ~250 products (25 MB)
















- The total number of job submitted was 136
 - success rate of 99%
 - 510 min for 6000 species

- The maps creation time heavily depended on the generation of the distribution maps itself:
 - Species: ~15 sec for each species
 - Family: ~25 sec
 - Order: ~1 minute
 - Class: ~3 minutes
 - Phylum : ~5 minutes

- A gCube VRE handling AquaMaps generation
 - Exploiting D4Science and gLite services
 - Equipped with a AquaMaps DB containing data on fishery species and their distribution
 - AquaMaps generation orchestrated via a VRE workflow
 - Job submission and management over the EGEE infrastructure
 - Interacting with the Archive Import Service to store generated AquaMaps compound objects in Content Management Service
 - Execution of Time Series Services to support the comparison and validation of multiple distribution maps
 - Discovery, access, and retrieval of distribution maps using different criteria (environmental, biological, geographical)

- Enhanced performance in the creation of AquaMaps
 - Distribution maps covering larger number of species
 - More often updates of distribution maps
 - Easier to tune “probability of occurrence” algorithm
- VRE exploitation brings advanced functionality
 - Support for alternative predictive maps
 - Execution of comparison and validation of predictive maps
- Scientists can focus on their core activities !

Thanks! Questions?

<p>Technology</p> <ul style="list-style-type: none"> gCube System gCube Issue Tracker  gCube Software Repository  gCube Software Documentation (wiki) gCore Issue Tracker  gCore Software Repository  gCore Software Documentation (wiki) gCore DEV Software Documentation (wiki) Software Distribution Site Build and Integration System  Software Integration Report  Software Release Procedures and Tools (wiki) Standards 	<p>Infrastructure</p> <ul style="list-style-type: none"> Resources Inventory (wiki) Real time Map Support Virtual Organization Management System  Infrastructure Management (wiki)
<p>Communities</p> <ul style="list-style-type: none"> Learn about gCube Environmental Monitoring  Fisheries and Aquaculture Resources Management  Requirements and Practices  Training Environment  Learn about e-Infrastructures 	<p>Project</p> <ul style="list-style-type: none"> Collaborative Working Space (BSCW)  User Profile Management  Quality Assurance (wiki) Support D4Science (e-mail) Photo gallery on Flickr D4Science on Delicious