



ICIS report to CWP
08-09 July 2008, Halifax Canada

The Integrated Capture Information System - ICIS

M. Taconet, S. Tsuji
FAO



- Objectives of ICIS requirements gathering meeting (June 08)
 - Community
 - Gather a small community to pilot a prototype implementation
 - Requirements
 - Build community consensus on the overall vision and a set of high-level requirements
 - This provides basis for report to wider CWP community

D4Science vision

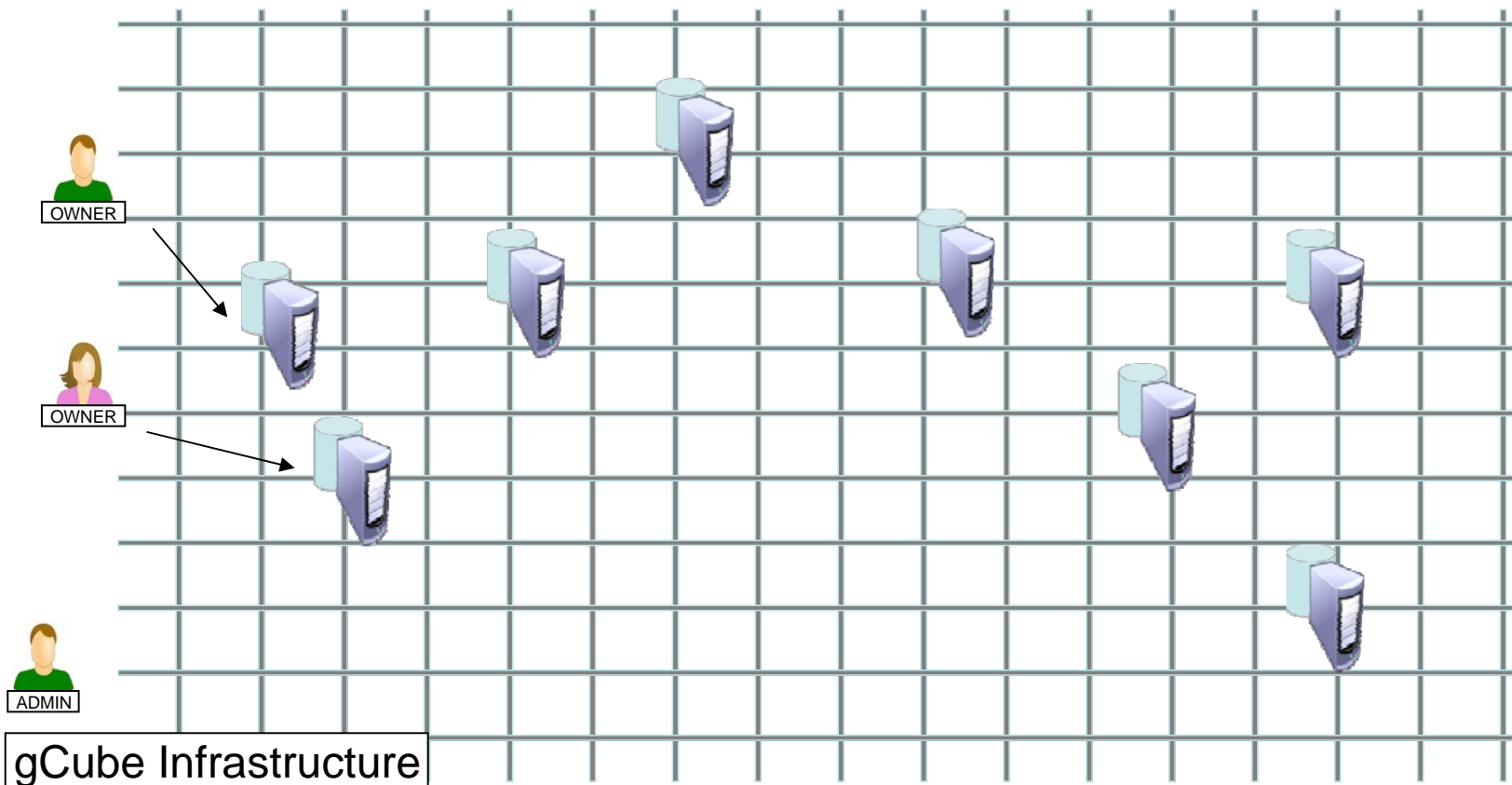
calls for the realization of scientific e-Infrastructures that will remove all heterogeneity, sustainability, scalability, and other technical concerns from the minds of scientists, *hide all related complexities from their perception*, and enable them to focus on their science and collaborate on common research challenges

gCube is

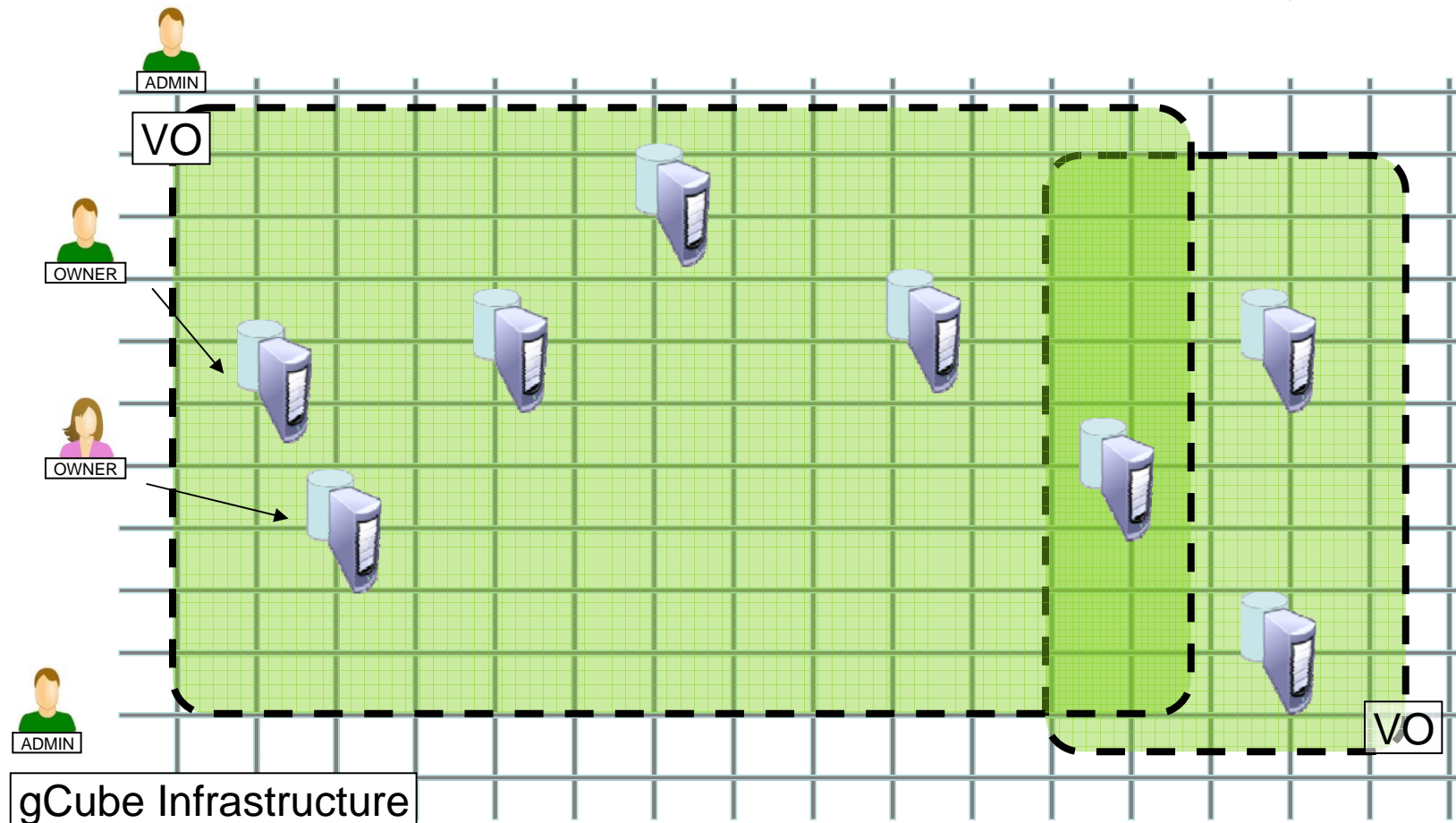
a framework to manage distributed **e-infrastructures** where it is possible to define, host, and maintain dynamic **Virtual Research Environments** (VREs) capable to satisfy the collaboration needs of distributed **Virtual Organizations** (VOs)

- An **infrastructure** is the basic physical and organizational structures and facilities (roads, power supplies, ..) needed for the operation of a society or enterprise
- An **e-Infrastructure** provides support for *effective consumption of shared resources*:
 - hardware-bound resources (i.e. networks, storage, instruments, and computational resources),
 - system-level software resources (i.e. basic middleware services),
 - and application-level software resources (i.e. data sources and services).

These e-Infrastructures offer mechanisms that concurrently exploit networks, grids and data in a seamless fashion, and will thus enable scientific communities to operate within a coherent model, regardless of the location of their research facilities.

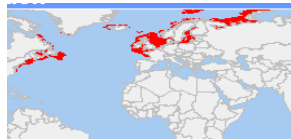


- A **Virtual Organization** (VO) models sets of users and resources (information and services) defining clearly and carefully
 - what is shared,
 - who is allowed to share,
 - and the conditions under which sharing occurs, usually based on an authentication and authorization policies.





Ontologies
"mapping rules"



Maps

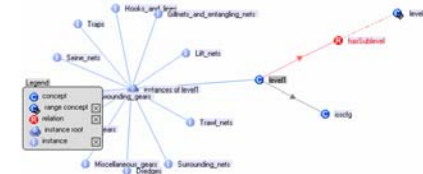
Year	Area	Species	Quantity	Value	Value	Value	Value
2001	Area 1	Species 1	100	100	100	100	100
2002	Area 2	Species 2	200	200	200	200	200
2003	Area 3	Species 3	300	300	300	300	300
2004	Area 4	Species 4	400	400	400	400	400
2005	Area 5	Species 5	500	500	500	500	500

fishery time series:
now catch stats
then VMS, ...

← Data sources

Services

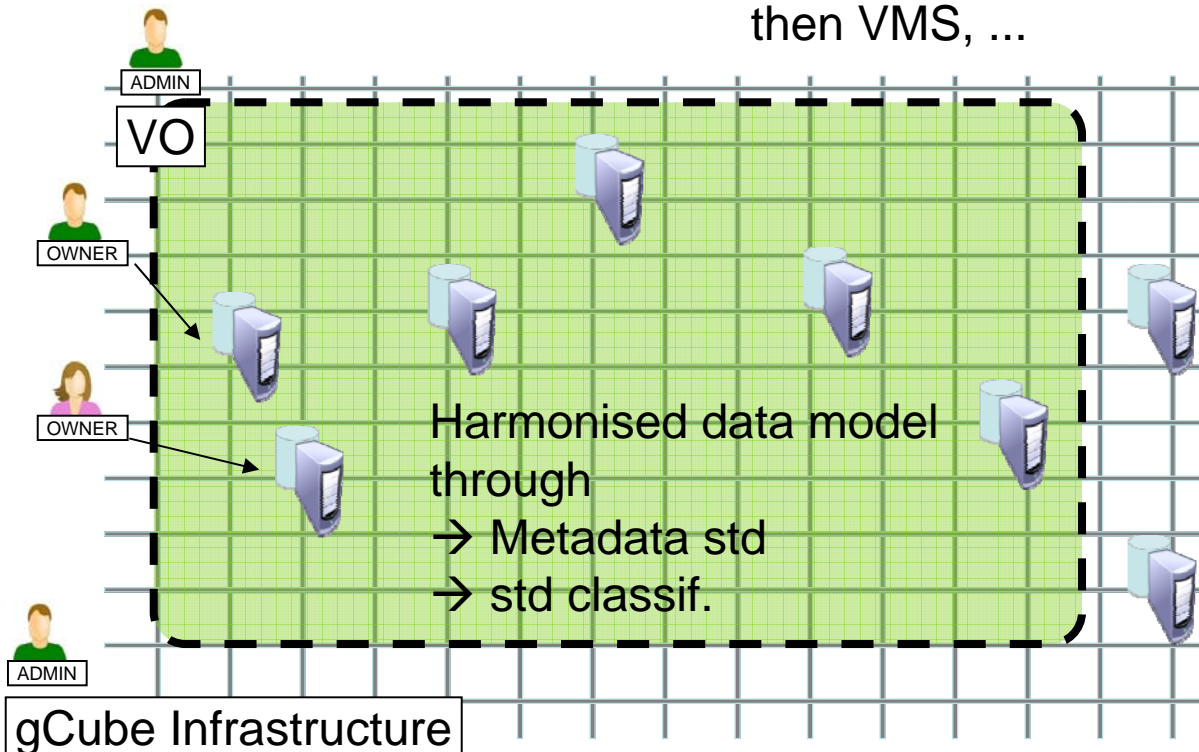
↓
Ontology services



FishStat

GeoNetwork

Aquamaps



- Data integration (the repository)

The first attempt of data repository and development of mapping rules will be based on:

- Statistics: publicly available data (e.g. StatLant, FAO statistics). This may include an input tool (like FiStat WS).
- The pilot will try to integrate some other data set(s) such as VMS catch, observer/port/inspection reports or stock/effort data (for authorized users only).
- Species distribution: likeliness of occurrence data. Output formats from FAO sp. maps and Aquamaps will be harmonized, each with reliability indices

- Data analysis
 - Comparison of multiple data sets should be possible, through summarization, aggregation

- Outputs

- Provenance metadata important, including authorship.
- Public has restricted interface compared to internal user.
- Should be possible to get out exactly what was put in

- Sources of species distribution (for ICIS)
 - FAO species maps (900): based on expert knowledge
 - Aquamaps (7000+): modeling predicting species distribution based on ecological parameters
 - output formats from 2 sources will be harmonized, each with reliability indices
 - Evaluation and refinement can be made by comparing them / integrating FAO maps in the Aquamap algorithm

- Aquamaps modelling

- Aquamap algorithm to be integrated into VRE system as a minimum.
- Should be easy to add new parameters.
- Peer review important: Weight in reallocation rules influenced by whether or not map has been validated.
- Can use catch stats to improve accuracy of species distribution

- Annotation
 - Need annotation, at cell level and/or above
 - Keep institutional memory of what people do with the system
 - Annotate both data and process used to retrieve the data

- Quality control
 - How do we check that some errors have not been introduced through the process?

- Security concerns (servers, contents)
 - Need to be sure access control is rigid.

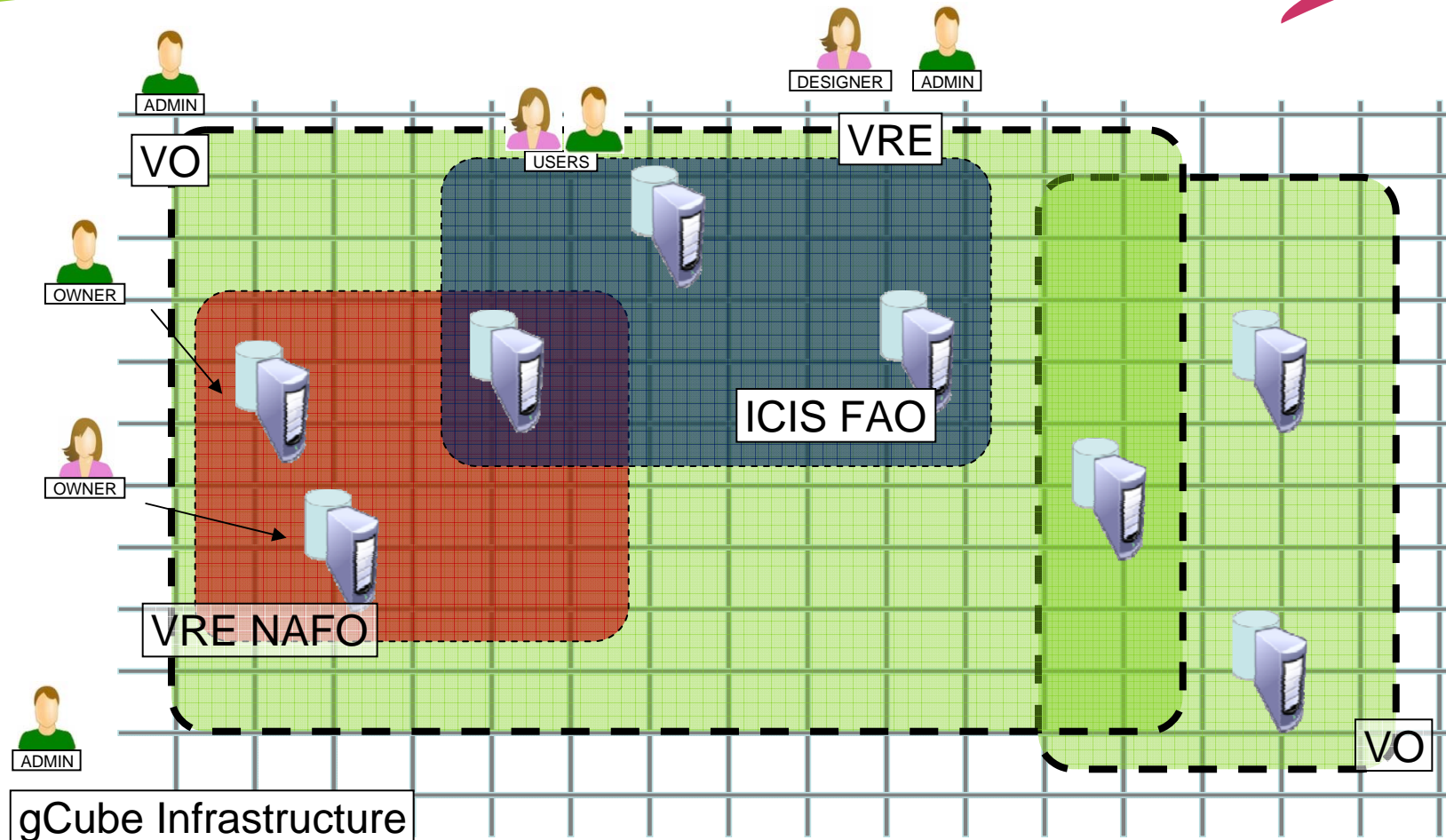
- Communication
 - The VREs should function also as a forum that encourages community discussion of datasets and analysis.

- End-user interfaces:
 - Tightly related to VRE instances (e.g. ICIS, NAFO).
 - Should be possible to customise for various user types
 - Queries
 - Simple to complex depending on VRE/user type.
 - Search terms definition may be an issue. When issue raised, resolve case by case basis in a collaborative way.
 - Another meeting focusing on end-user requirements is planned together with evaluation/recommendations on the first early version of the VRE (towards the end of 2008).

- A **Virtual Research Environment** (VRE) provides a framework of applications, services and data sources dynamically identified to support the underlying processes of research/collaboration/cooperation.

The purpose of a VRE is to help researchers* in all disciplines by managing the increasingly complex range of tasks involved in carrying out their activities.

*Researcher has to be considered in the large, i.e. end-user, decision-makers, resource and data providers, etc.



- Data exploration – maps, visualize trends, comparison of diff data sets,
- Management (e.g. quota) .vs. Catch monitoring
- See the data from others
- easy queries, easy extraction to importing other applications, no sophistication in visualization
- NAFO will provide access to its resources (i.e. data, tools, and information system) useful for the ICIS project

- Data analysis
 - Aggregation / Reallocation rules (only as ICIS FAO VRE)
 - FAO will develop the rules, for feedback and advice from CWP level and by expert groups when possible
 - When data processed/reallocated, it must be clearly indicated under the name of FAO (see “Output”)
 - Any output products will carry the FAO badge and be clearly identified as analytic products and not raw source data.

- Aggregation
 - System will provide mechanisms of aggregation to public only in the way defined by ICIS developers.

- Reallocation (including disaggregation)
 - Spatial reallocation will be considered within the context of the need to separate between High Seas and within EEZ catches

 - Reallocation scenarios must be carefully reviewed before releasing as analytic products to the general public.

 - Other reallocation types might be considered later (time, ...)

 - Process used must be transparent in final product.

- Reliability rules
 - Highest for those maintained by RFB who “is in competence of” management of species in a certain area (e.g. NAFO is responsible for management of the 23 stocks of 11 species).
 - How to define? some lists available in FIRMS. Key will be based on linkage between RFB-species-type of mandate-dataset
 - This rule is only applicable for consolidating FAO statistics, Statlants, and formal catch statistics.
 - Needs further refinement for individual cases, starting to identify the substantially conflicting cases.
- Are all of these questions a matter for the CWP?

- Outputs

- Any output products will carry the FAO badge and be clearly identified as analytic products and not raw source data.
- Process used for any reallocation very important.
- System does not allow arbitrary reallocation.



Thank you.

