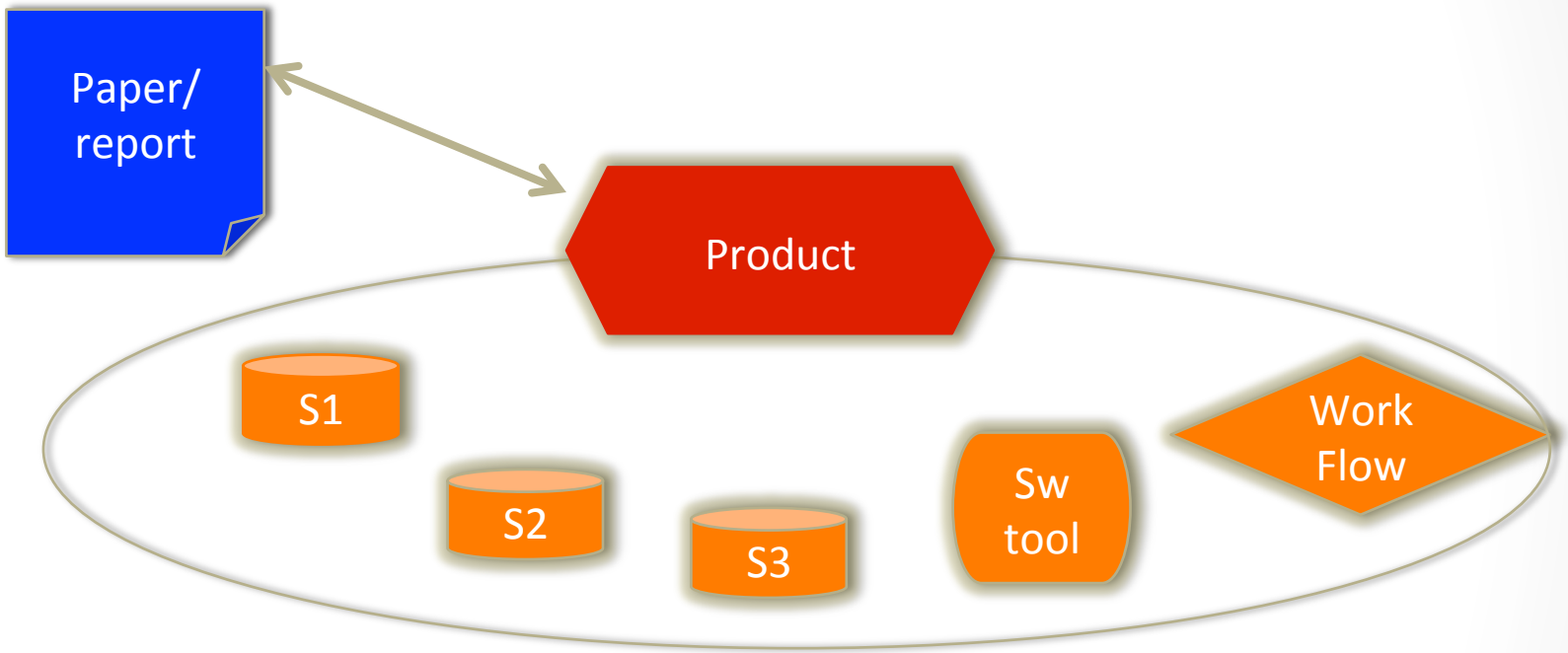
Interoperability of Research Data

Donatella Castelli
CNR-ISTI
donatella.castelli@isti.cnr.ir

Data-driven science



Research and scientific advice to policy makers are increasingly based on data

Supporting environments

- The set and type of scientific products that can be generated is largely influenced by the available **IT supporting environment**

- The power of an environment is a function of the **available tools** and of the **collections of data that can be easily exploited by these tools**

To maximize the data collections that can be exploited by the most effective tools (for each target task)

Crossing project and community boundaries

- Data are collected by projects with a **limited temporal and geographical coverage**
  - stored in different archives
  - with different metadata descriptions
  - with different granularity and quality
  - with different access and usage policies

- Addressing current scientific challenges requires working with
  - data from **multiple projects** in the same domain
  - data collected and produced in **multiple domains**

Interoperability issue

- The consumer service must be interoperable with the data provider services

- The solution implemented by the consumer service must supports interoperability across multiple data provider services

Interoperability

"The ability of two or more systems or components to ***exchange information*** *and to* ***use the information*** *that has been exchanged"*

IEEE Computer Glossary

Interoperability facets

Exchangeability

The consumer is able to access "meaningful" data

- Syntactic, structural and semantic aspects

Compatibility

The consumer policies are compatible with those regulating the data source

- Consumer and Provider Organizations are mutually trusted
- Terms of Usage and Usage policies are compatible

Usability

The consumer can effectively use the accessed data in order to perform target tasks

- Quality mismatching
- Right temporal coverage
- Data-incomplete mismatching

Interoperability across multiple heterogeneous data sources for supporting multiple usages is a too complex problem to be addressed with ad hoc-solutions

Data e-infrastructure

e-Infrastructure

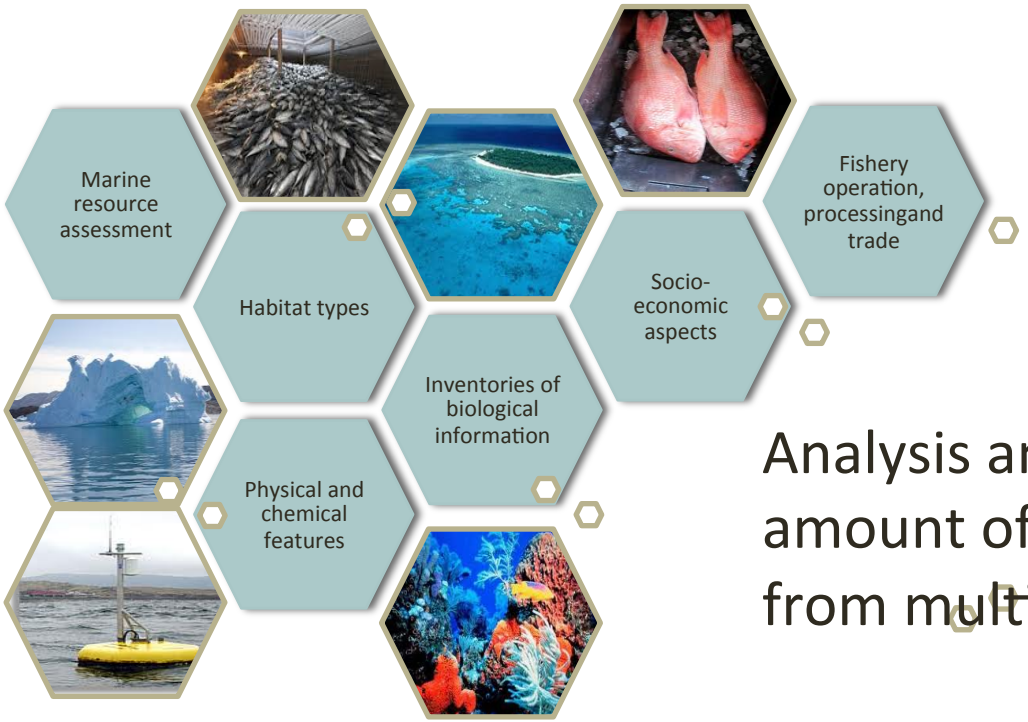
Electronic platform operated by a **responsible entity** offering an open set of **basic enabling services** (including access to resources) to a distributed **Community of Practice**. By exploiting these shared services the members of the Community of Practice realise **economies of scale**.

A ***data infrastructure*** offers services for dealing with data (e.g. access, management, curation, ….)

iMarine data infrastructure

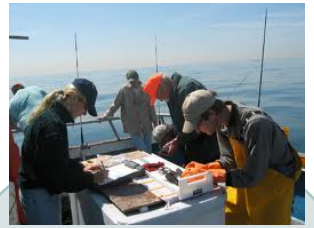
www.i-marine.eu

Supports a Community of Practice aiming at implementing the principles of the Ecosystem Approach to Fisheries Management and Conservation of Marine Living Resources

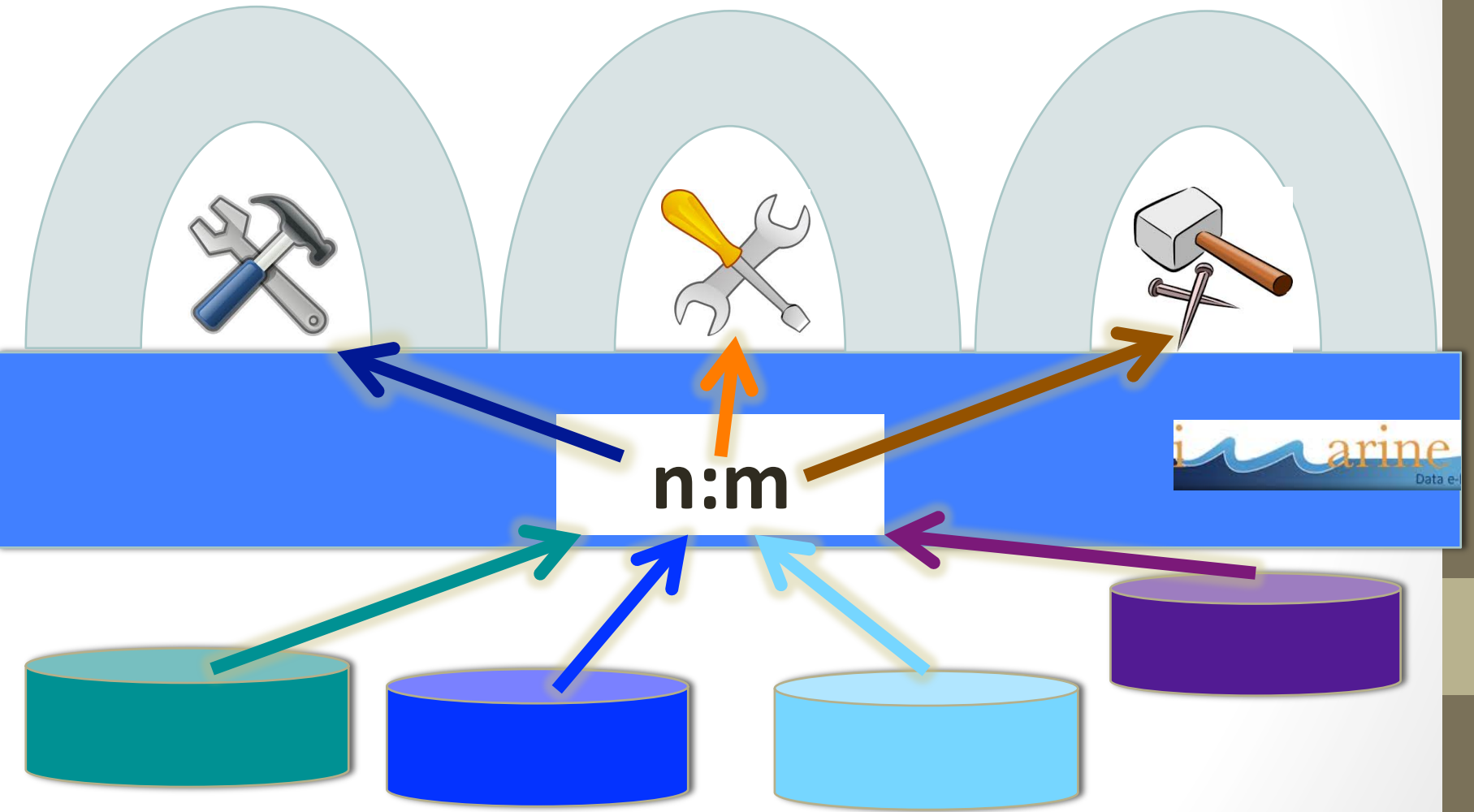


Analysis and processing of a large amount of heterogeneous data from multiple domains

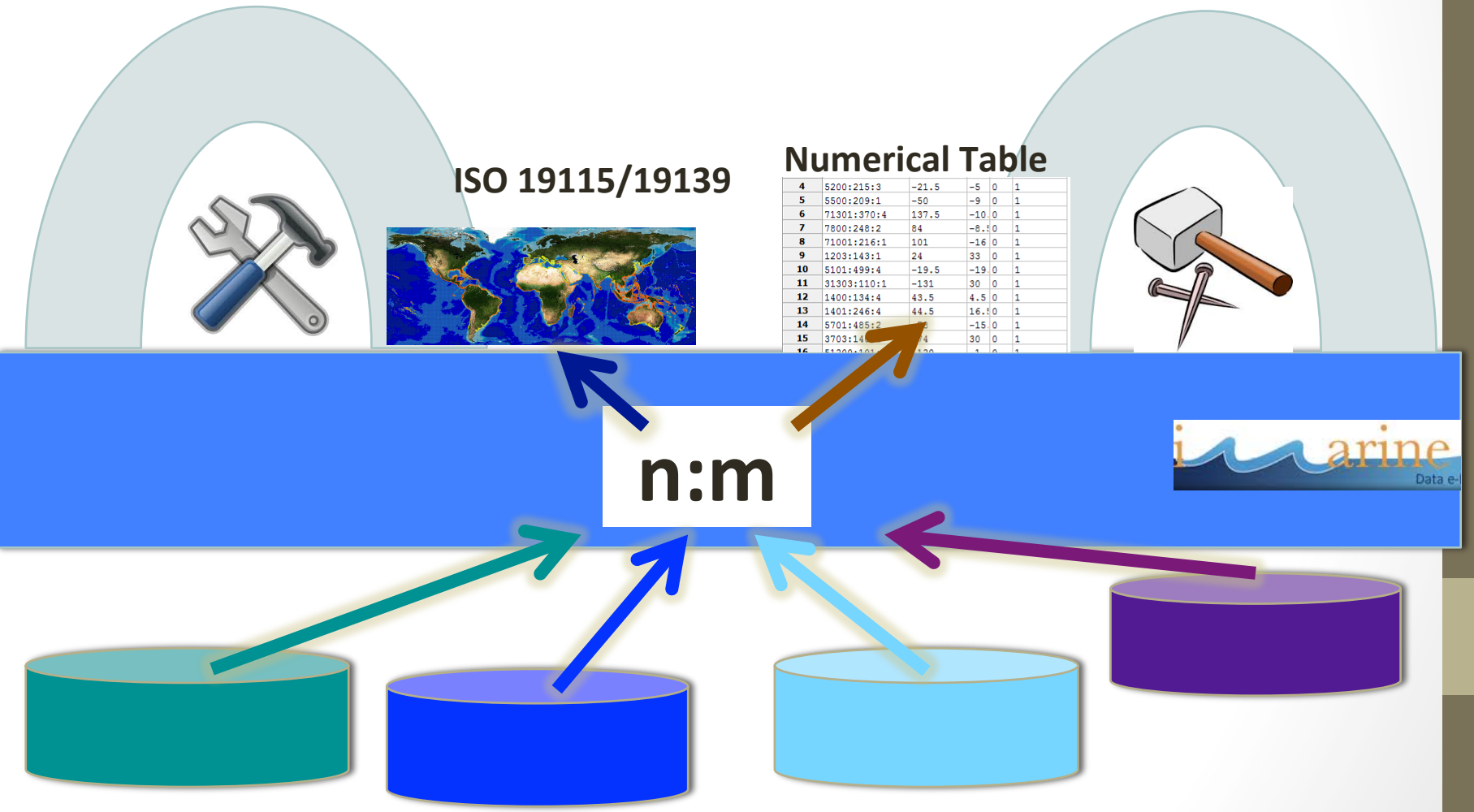
Virtual Research Environments



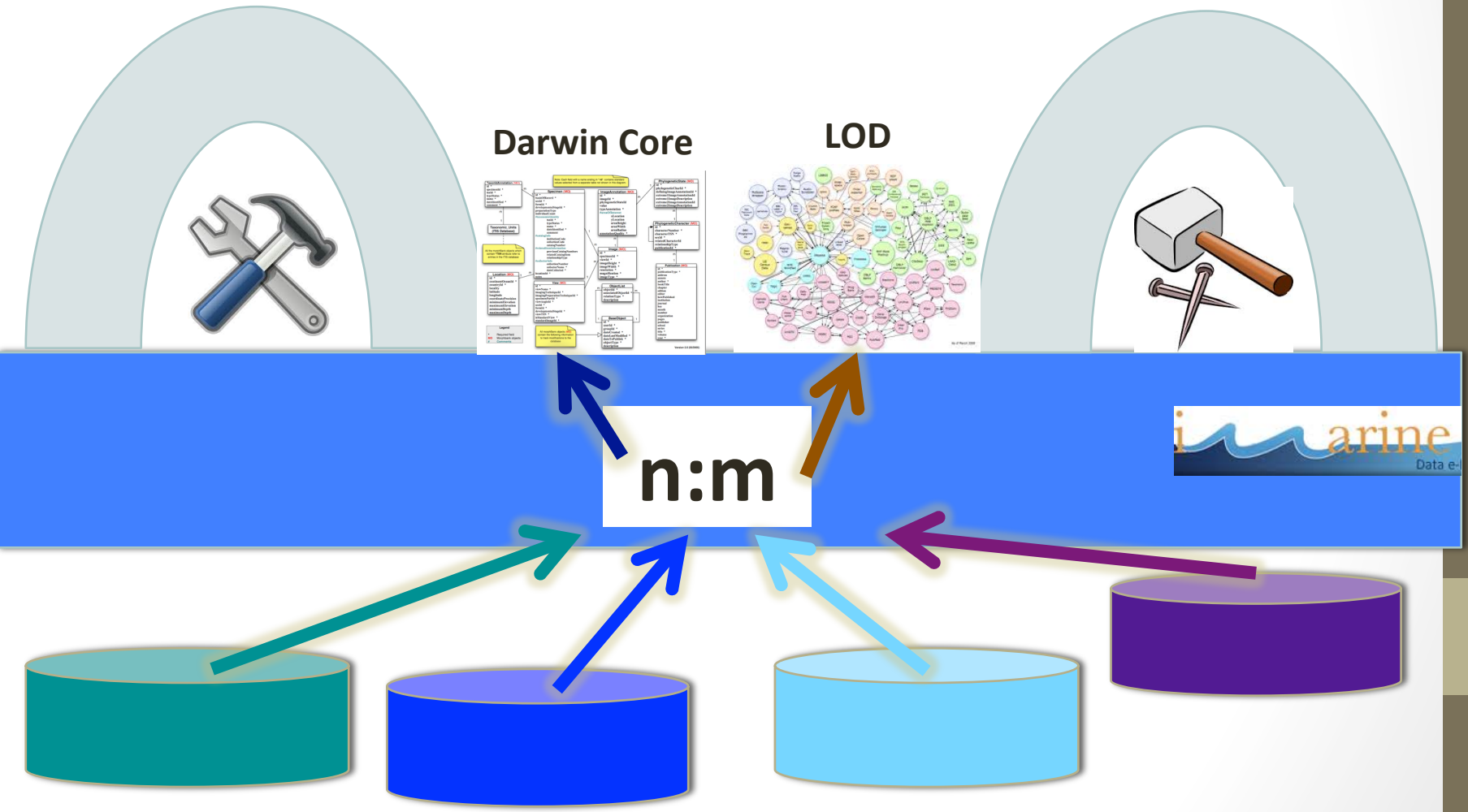
Multiple data sources
Multiple sw tools



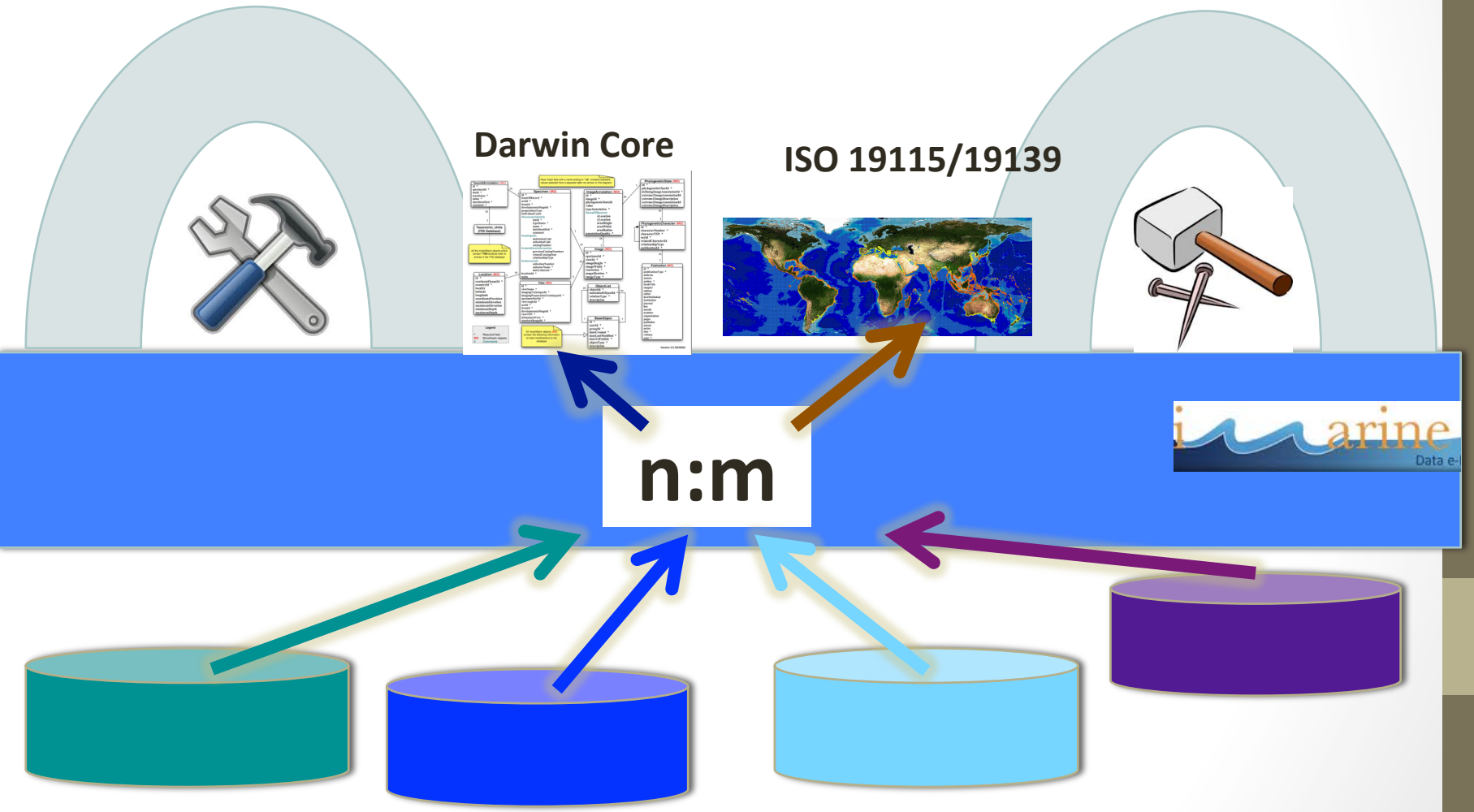
Example 1



Example 2



Example 3



iMarine solutions

- Standards
- Controlled vocabularies
- Generic frameworks and plug-ins
- Mechanisms for improving data quality and completeness

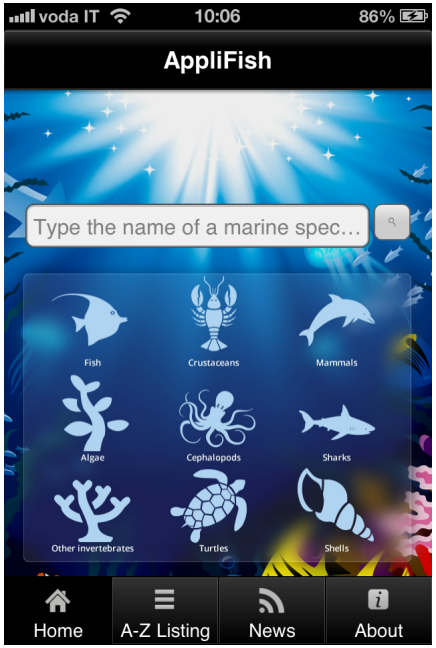
Still many open issues

- Many different standards
- Many plug-ins
- Evolving controlled vocabularies
- Lack of explicit representation of contextual information (e.g. policies)
- Context dependent interpretations (e.g. quality)

Concluding remarks

- Supporting interoperability across heterogeneous data sources for different tasks is an extremely complex problem
- Data infrastructures can help by providing shared solutions
- Complexity can only managed with a more in-depth understanding of the problems and a more systematic approach to them

Thank you



AppliFish

www.i-marine.eu