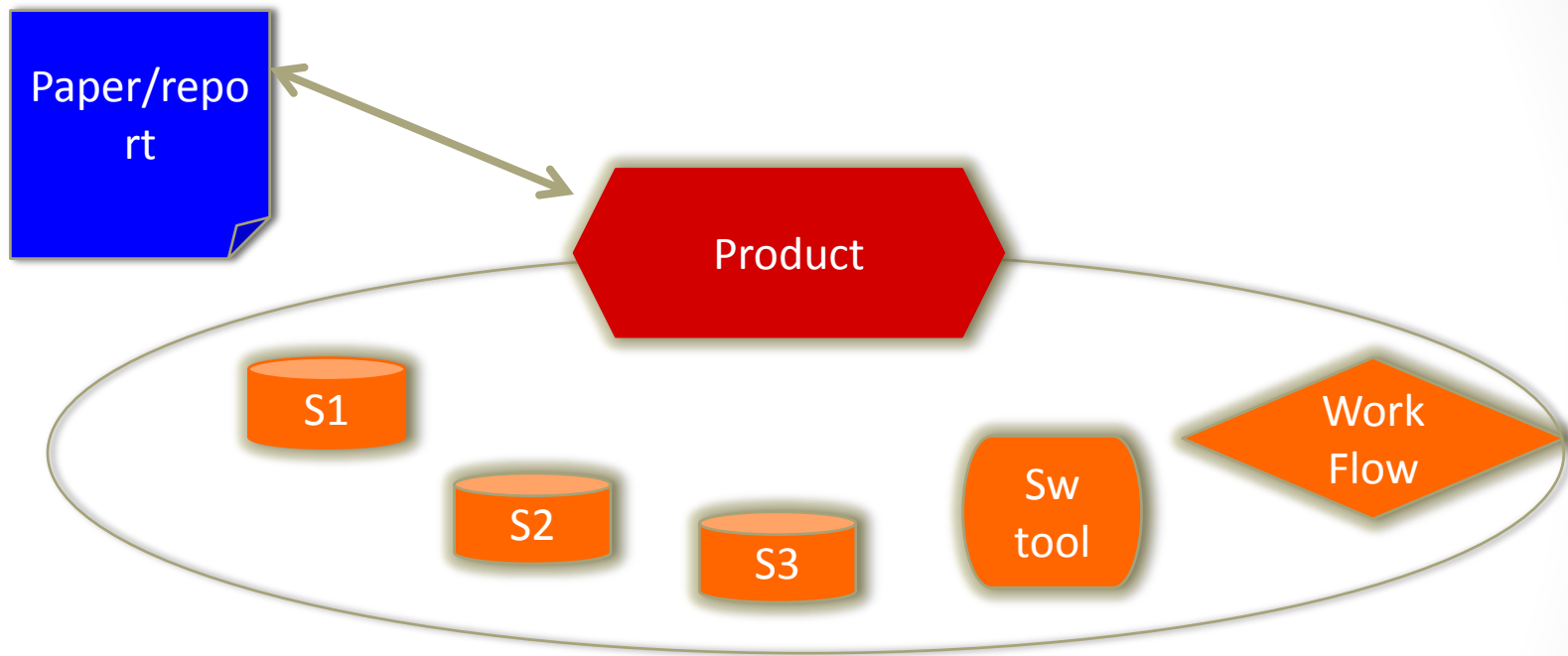# Interoperability of Research Data

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

Inter*operability*

*"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
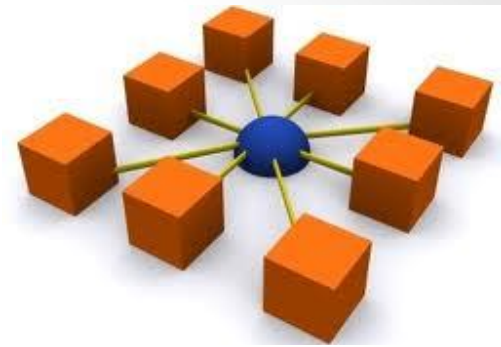
e-Infrastructure

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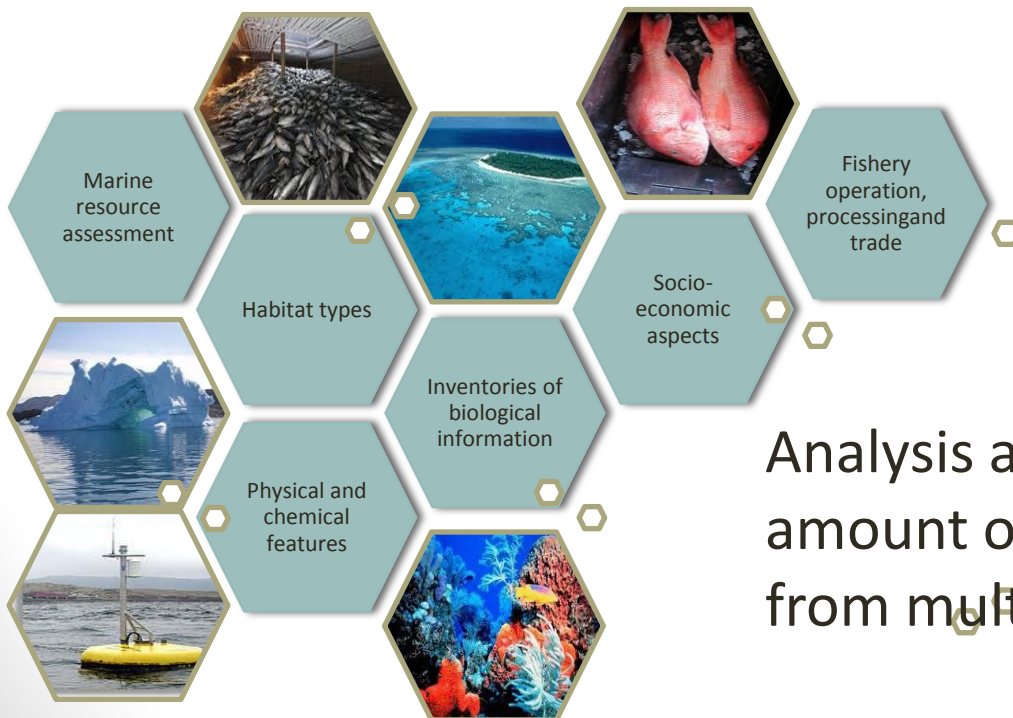
A **_data infrastructure_** offers services for dealing with data (e.g. access, management, curation, ...)

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
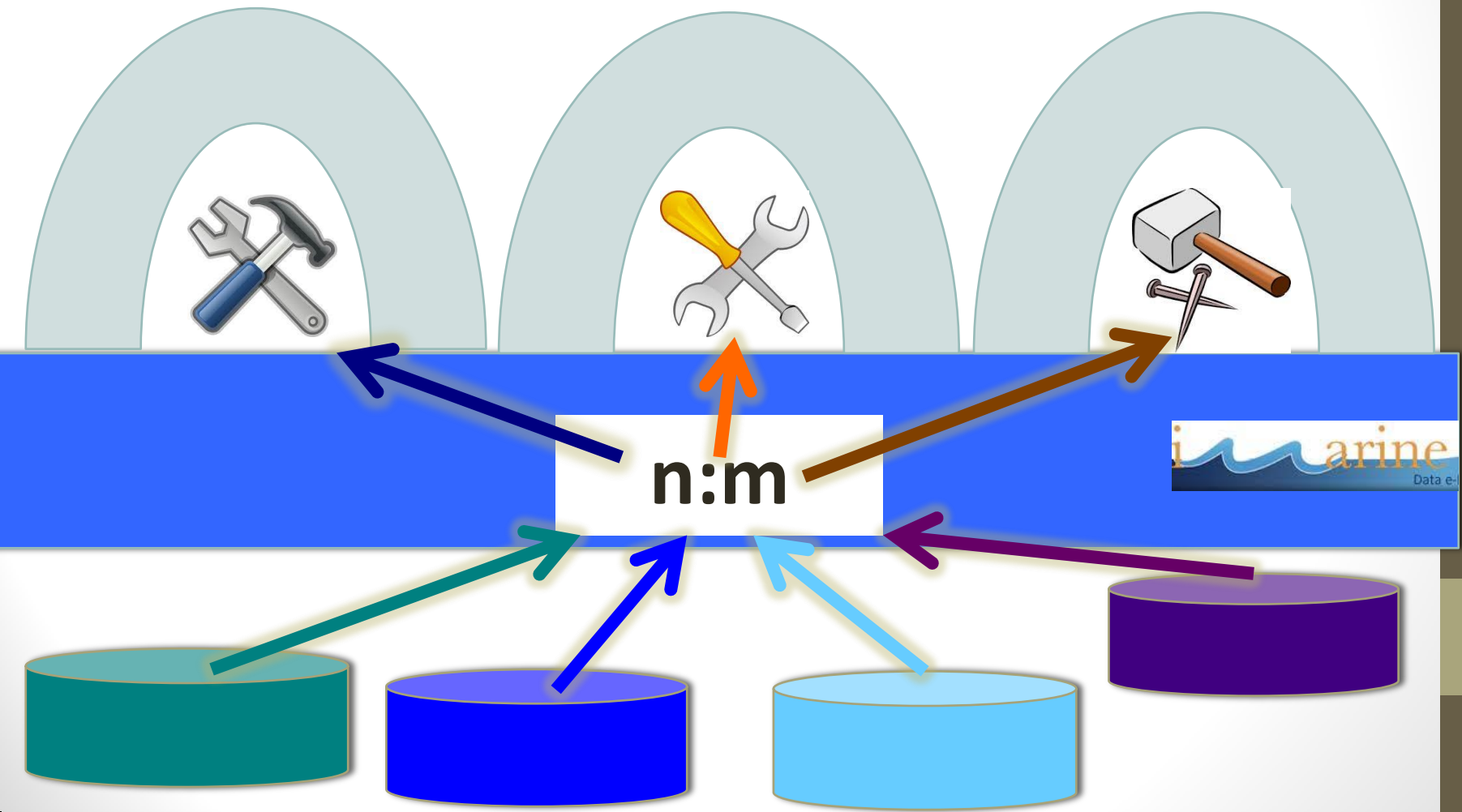
Marine resource assessment

Habitat types

Fishery operation, processing and trade

Socio-economic aspects

Inventories of biological information

Physical and chemical features

Analysis and processing of a large amount of heterogeneous data from multiple domains
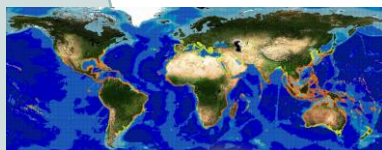
Virtual Research Environments
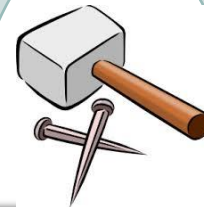
Multiple data sources
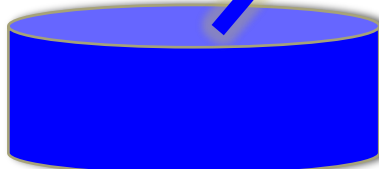# Multiple sw tools

n:m

Example 1

ISO 19115/19139

Numerical Table

4	5200:215:3	−21.5	−5	0	1
5	5500:209:1	−50	−9	0	1
6	71301:370:4	137.5	−10	0	1
7	7800:248:2	84	−8.5	0	1
8	71001:216:1	101	−16	0	1
9	1203:143:1	24	33	0	1
10	5101:499:4	−19.5	−19	0	1
11	31303:110:1	−131	30	0	1
12	1400:134:4	43.5	4.5	0	1
13	1401:246:4	44.5	16.5	0	1
14	5701:485:2		−15	0	1
15	3703:14	4	30	0	1

n:m

Example 2

Darwin Core

LOD

n:m

Example 3

Darwin Core

ISO 19115/19139
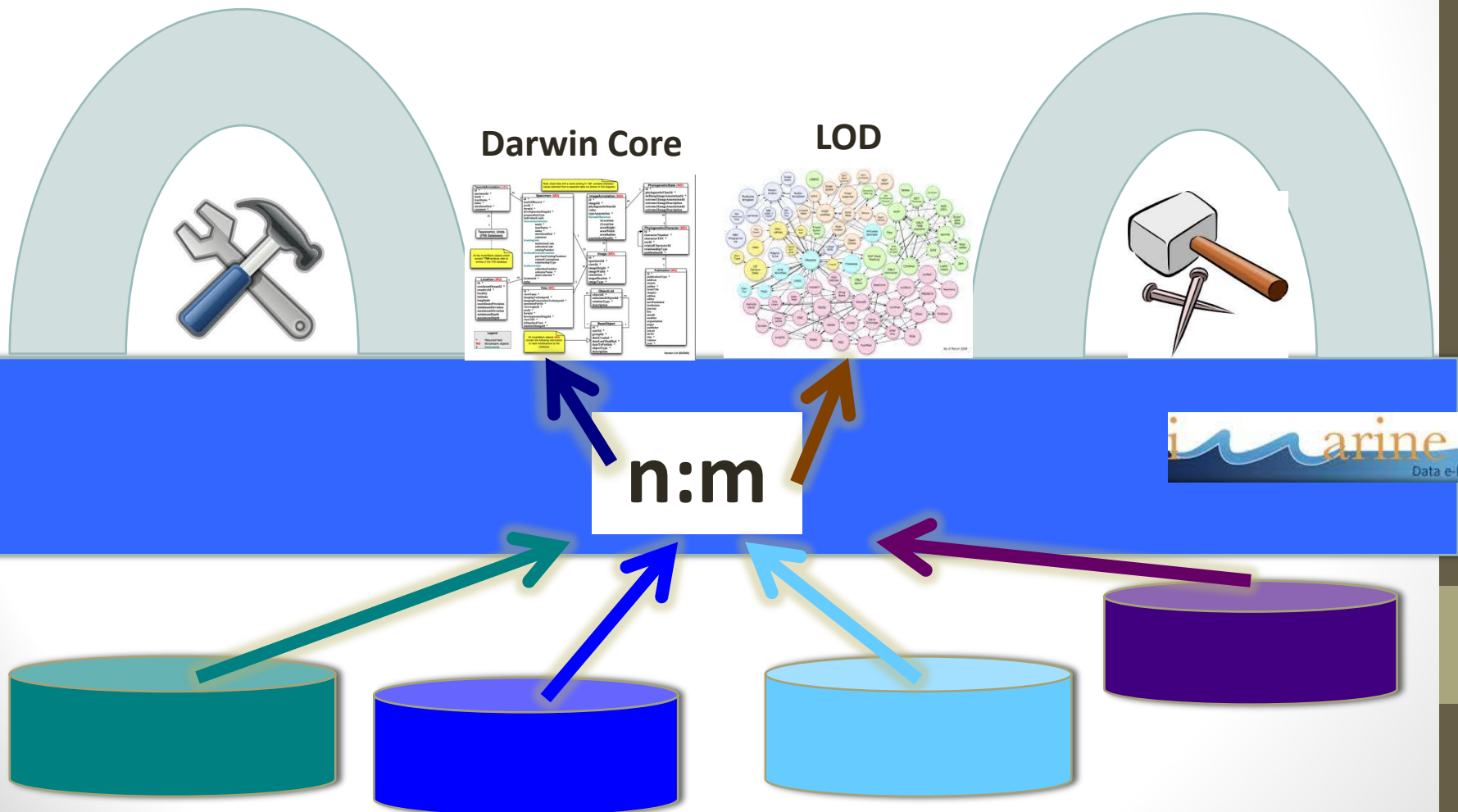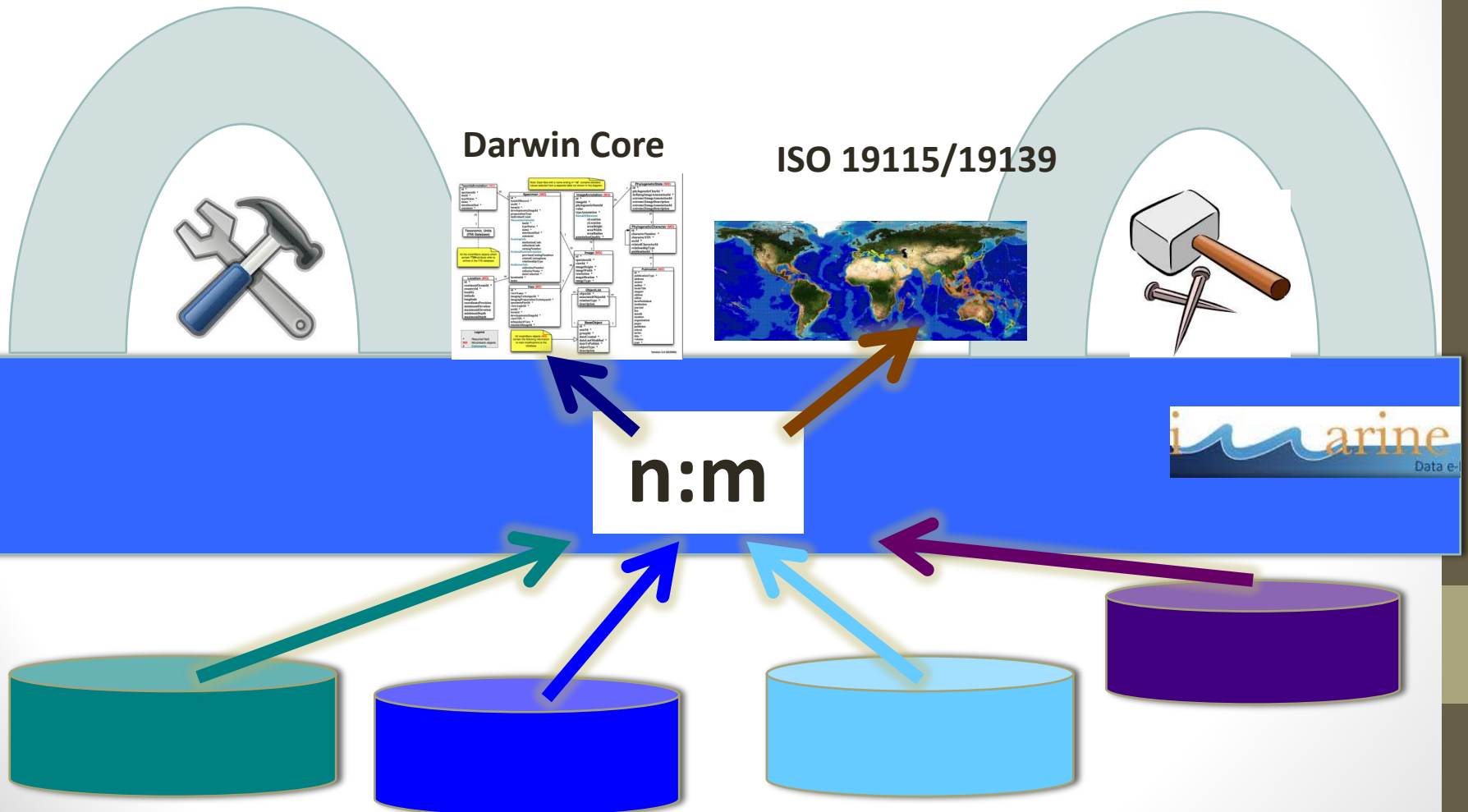
n:m

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