The VAMDC infrastructure

Standard procedures to publish, search and process atomic and molecular data

N.Moreau

Paris Observatory, LERMA

Gathering heterogeneous data

- Many services provide atomic and molecula data

- Content of data is different (atomic and molecular spectroscopy, collisions, solids) but we sometimes users have to compare or merge them

- Many services have defined their own data format

- Problematics :
  - How to query multiple databases
  - How to identify comparable data
  - Which data format to use

- Goal of Virtual Atomic and Molecular Data Center project : to provide an interoperable e-Infrastructure for the exchange of atomic and molecular data.

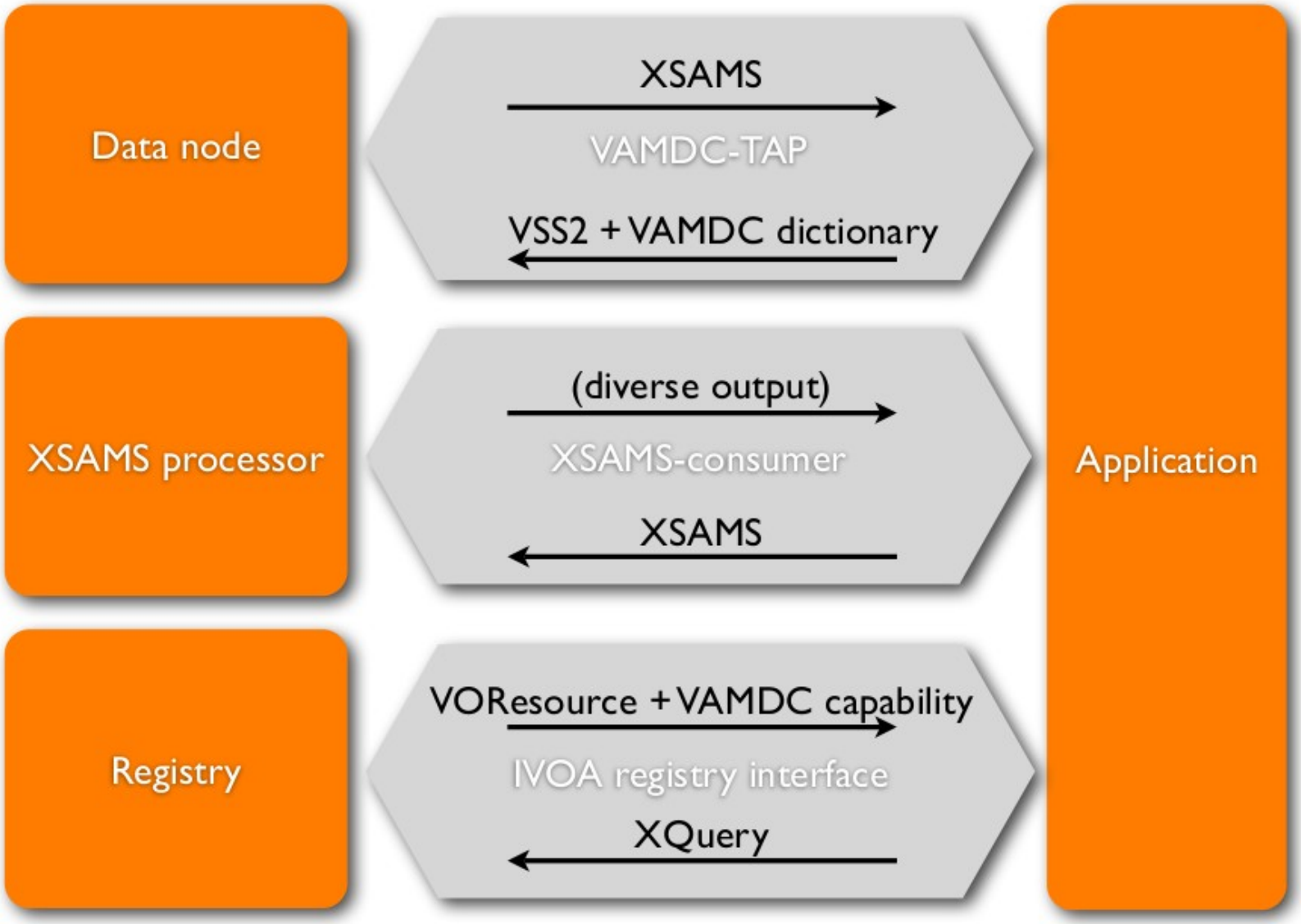
The VAMDC project

- Supported by EU in the framework of the FP7 (2009-2012)
- Involved 15 administrative partners representing 24 teams from 6 European Union member states
- Built an interoperable e-Infrastructure for the exchange of atomic and molecular data
- Now managed by the VAMDC Consortium
- 30 databases available

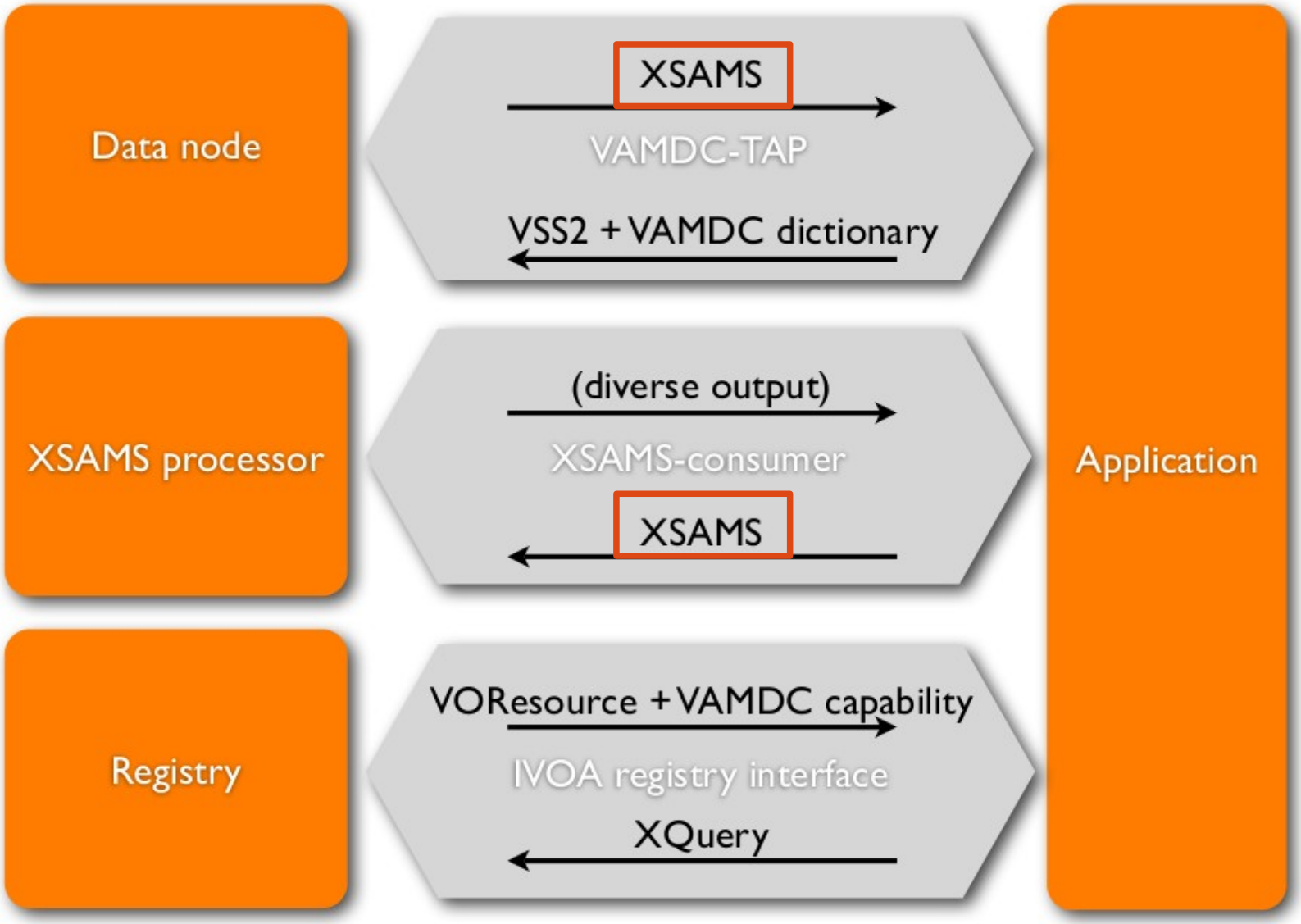
The VAMDC infrastructure

- Organization of the infrastructure capitalized on IVOA experience :
  - Registry
  - Extension of VOResource data model to describe VAMDC resources
  - Services query protocol is a simplified version of TAP (VAMDC-TAP)

The VAMDC infrastructure



VAMDC standards : XSAMS



VAMDC standards : XSAMS

- XSAMS stands for XML Schema for Atomic, Molecular and Solids (http://www.vamdc.org/documents/vamdc-xsams-guide_v12.07.pdf)

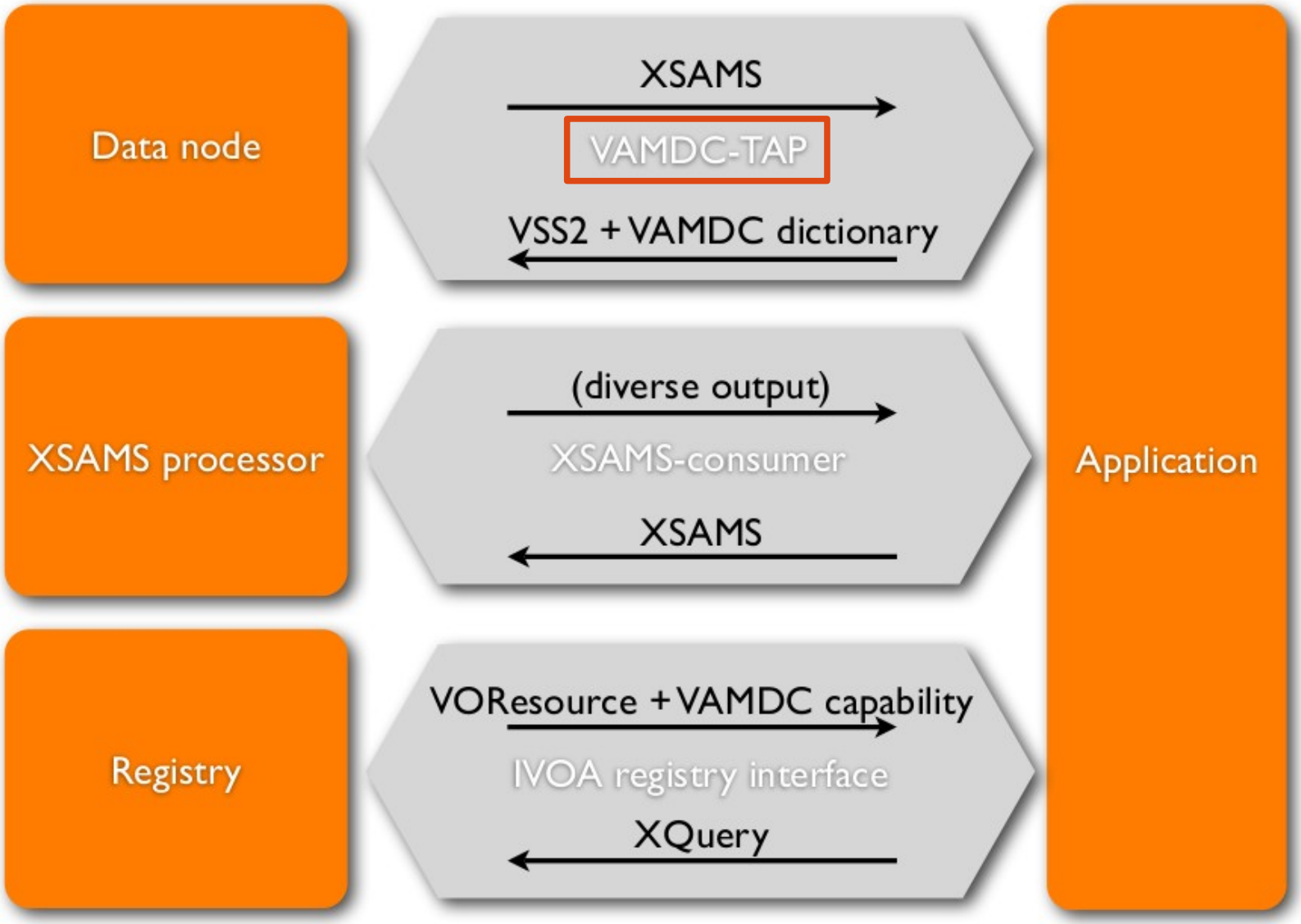
- A common format was necessary because VAMDC includes databases providers from very different fields (atomic, molecular and solid spectroscopy)

- Standard for exchange of atomic, molecular and particle-surface-interaction (AMPSI) data

- Informations concerning sources and generation of the data must be provided

- Correctness or applicability of the data is left to the producer responsibility

VAMDC standards : VAMDC-TAP



VAMDC standards : VAMDC-TAP

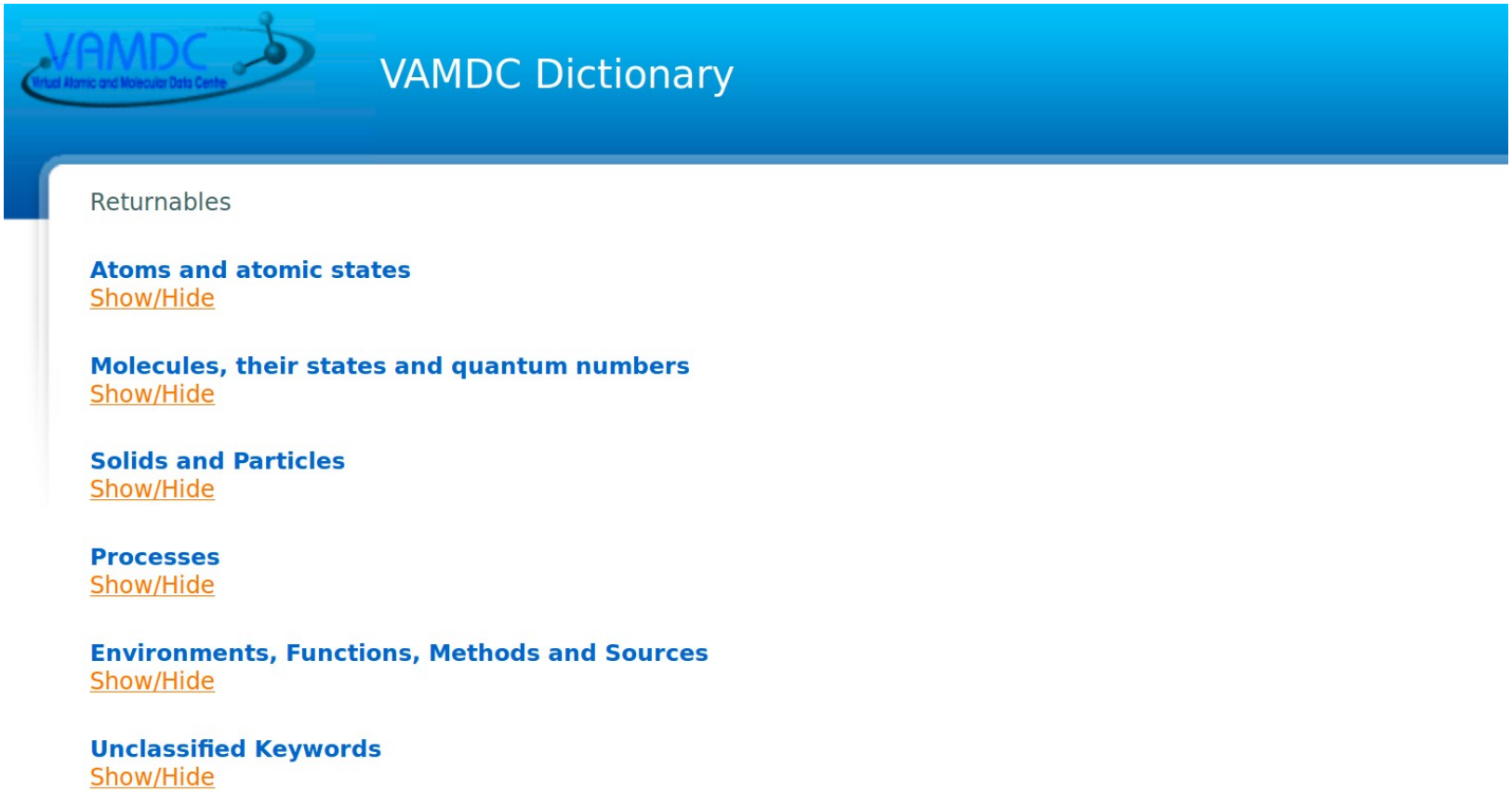
- Based on IVOA TAP (sync, async requests, all services have capabilities / availability)

- SQL-like requests

- Simplified to avoid join :
  - The data model is seen as one big table
  - All quantities are well defined into a dictionary
  - http://dictionary.vamdc.eu

- Example : select * where ((AtomSymbol = 'he') OR (AtomSymbol = 'li'))

VAMDC standards : VAMDC-TAP

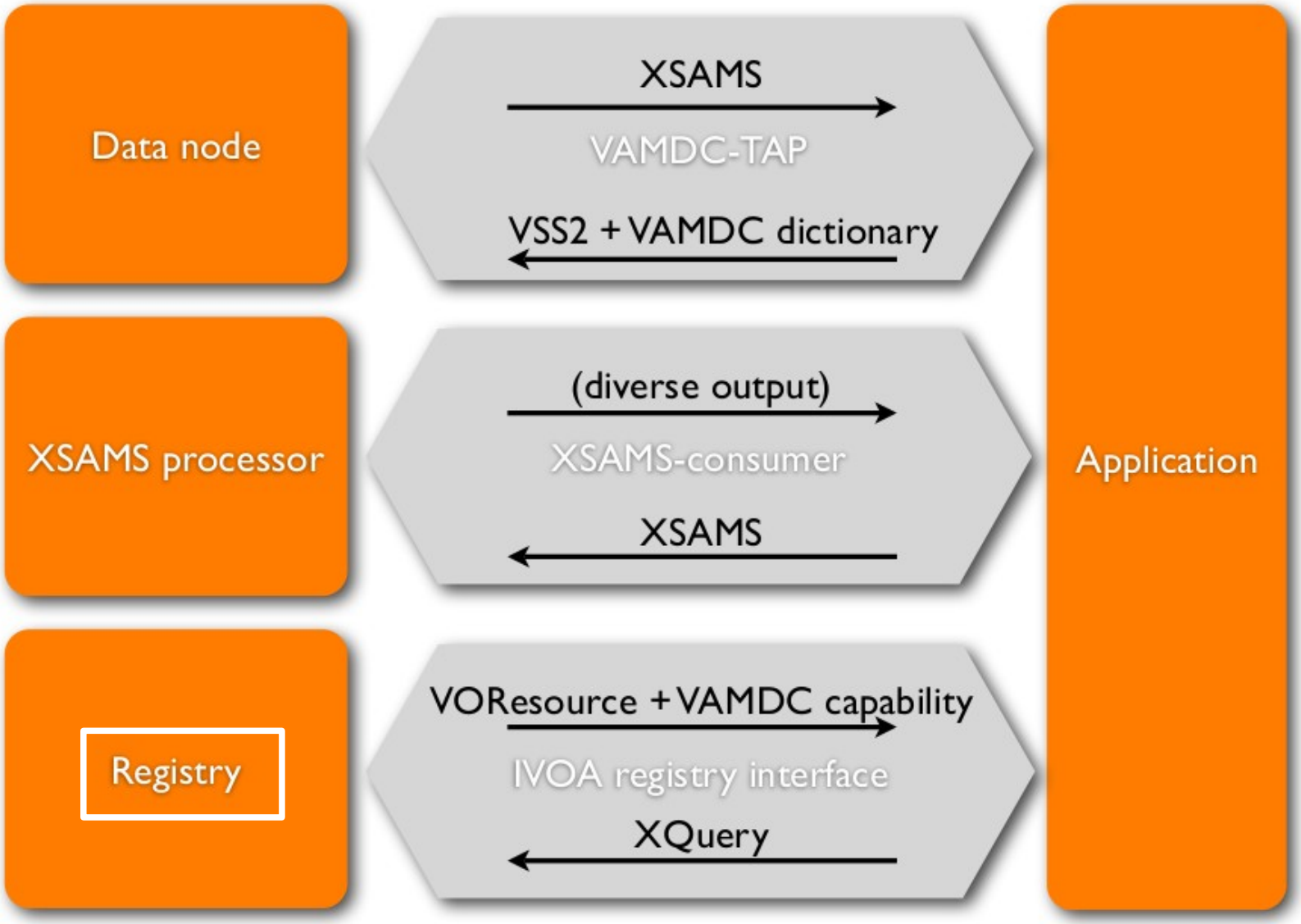


VAMDC standards : VAMDC-TAP

Molecules, their states and quantum numbers
Show/Hide

Keyword	Short Description	Long Description	Type	*
MoleculeBasisStates	The basis states for a set of molecular states expressed as a linear combination on some basis	The basis states for a set of molecular states expressed as a linear combination on some basis		
MoleculeCASRegistryNumber	CAS registry number of molecule	CAS registry number of molecule	String	
MoleculeCNPIGroup	CNPI group of molecule	CNPI group of molecule	String	
MoleculeChemicalName	Molecule name	Conventional molecule name, e.g. CO2, NH3, Feh (may not be unique)	String	
MoleculeComment	Comment on molecule	Comment on molecule	String	
MoleculeIUPACName	IUPAC name of molecule	IUPAC name of molecule	String	
MoleculeInchi	Inchi of molecule	Inchi of molecule	String	
MoleculeInchiKey	InchiKey if molecule	InchiKey if molecule	String	
MoleculeIonCharge	Molecule ion charge	Molecule ion charge	Integer	
MoleculeMolecularWeight	molecular weight	molecular weight	Float	*
MoleculeNormalModeDisplacementVectorComment	Comments concerning this normal mode's displacement vectors	Comments concerning this normal mode's displacement vectors	String	
MoleculeNormalModeDisplacementVectorMethod	Method for displacement vector in molecule normal mode	Method for displacement vector in molecule normal mode	String	
MoleculeNormalModeDisplacementVectorRef	A reference to the atom in the molecule's structure to which this displacement vector applies	A reference to the atom in the molecule's structure to which this displacement vector applies	String	

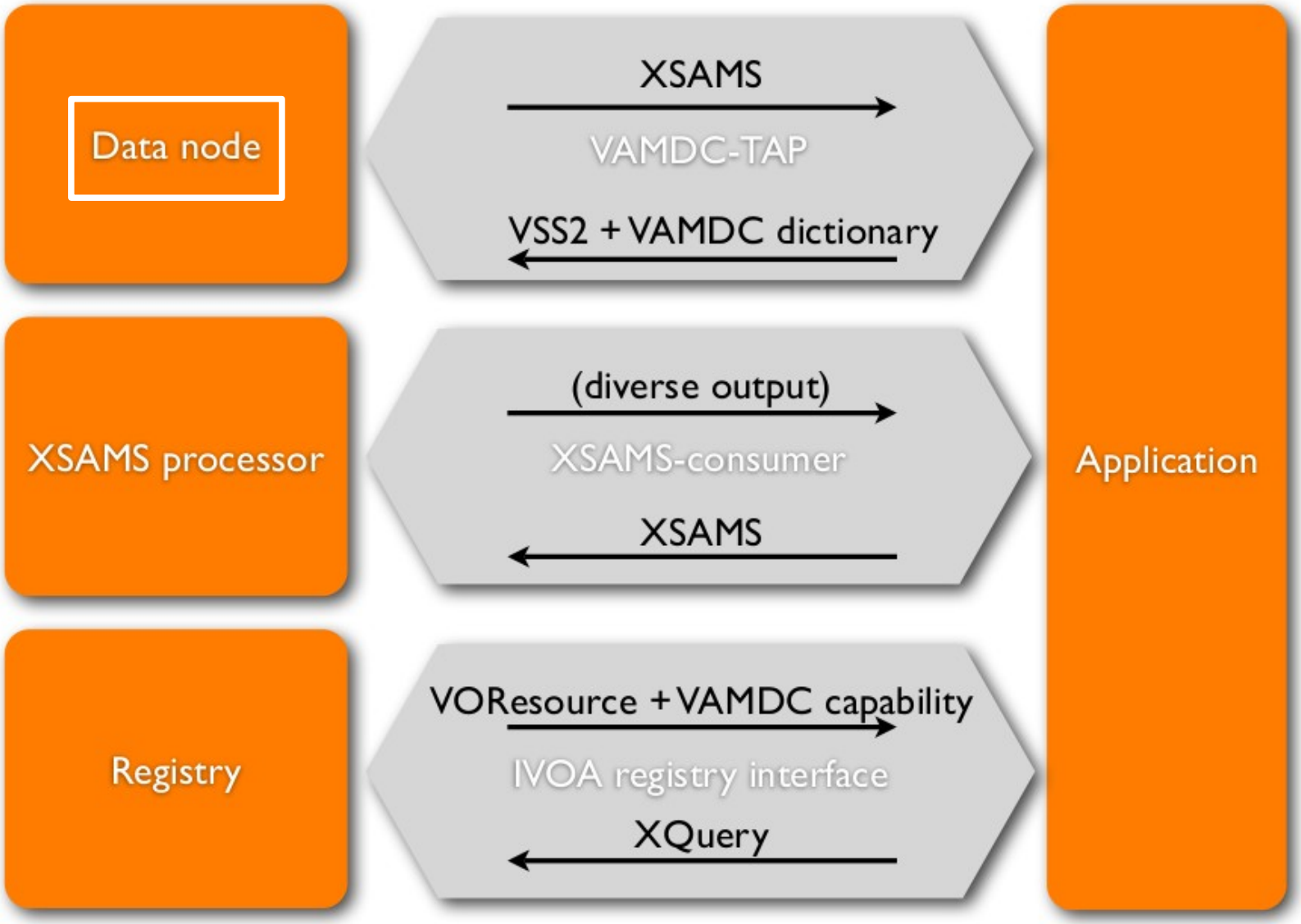
VAMDC registry



VAMDC registry

- Astrogrid registry (UK's virtual observatory development project)
- http://registry.vamdc.eu/
- Contains several types of resources
  - VAMDC nodes (databases)
  - Processors (data conversion tools)
  - VAMDC species database
- All services registered their capabilities
  - Services endpoints
  - Queryable and returnables quantities (using dictionary)

VAMDC node


Data node
XSAMS
VAMDC-TAP
VSS2 + VAMDC dictionary
XSAMS processor
(diverse output)
XSAMS-consumer
XSAMS
Registry
VOResource + VAMDC capability
IVOA registry interface
XQuery
Application

Data node

- A node is a database registered in the VAMDC infrastructure
- It can understand VAMDC-TAP query and returns data in XSAMS
- We provide a middleware to implement the VAMDC layer in two versions :
  - Java
  - Python (used by the majority of nodes)
- Source code available on github

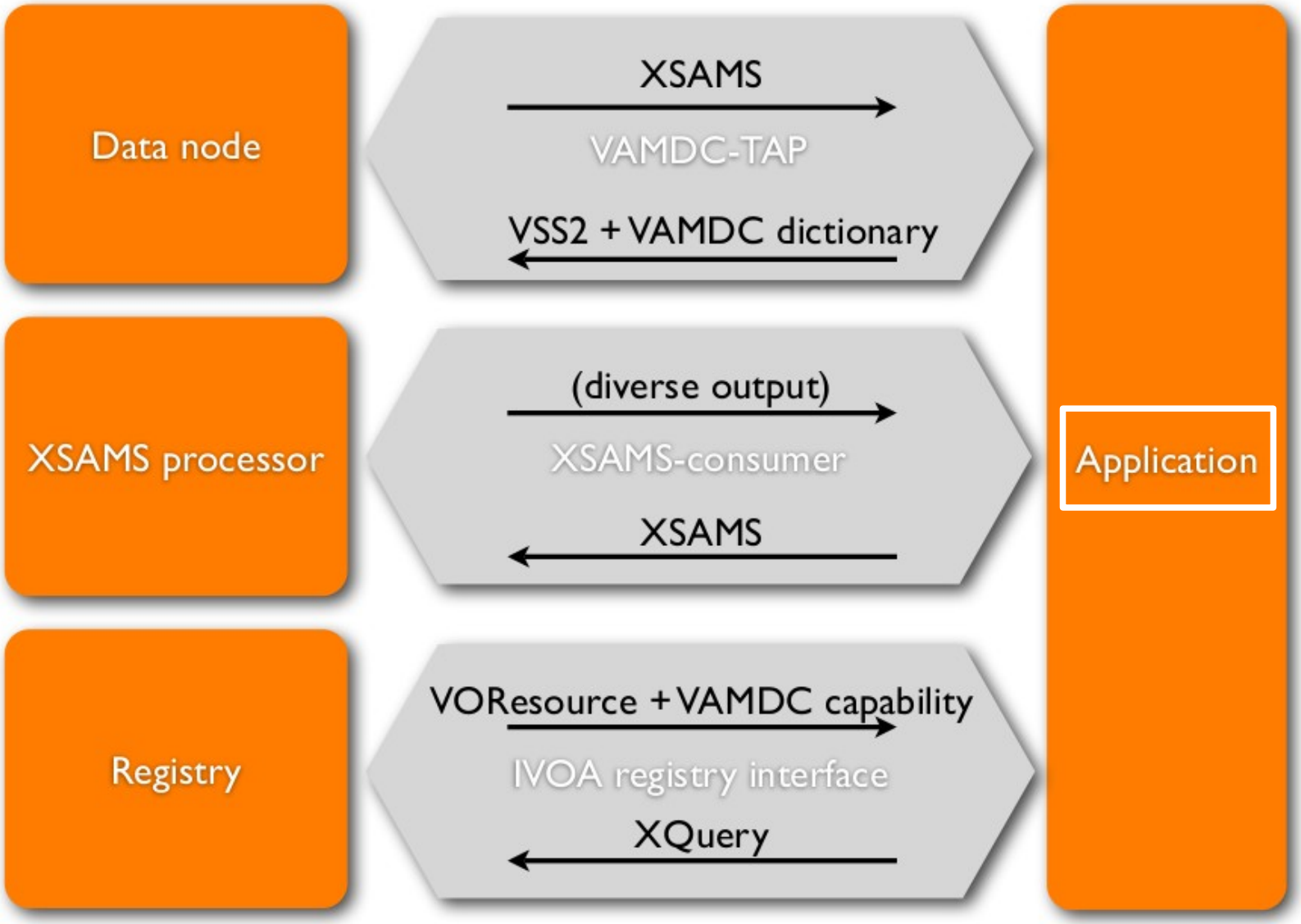
Data node

- If the standards are updated, the node simply updates its middleware version.

- Requirements :
  - Data have to be stored in a relational database
  - Data provider writes a mapping between the column of its database and VAMDC dictionary columns
  - Node can respond to « select species » query that returns all the species contained in the database

- Before inclusion in the registry, node conformity is checked :
  - Is returned XSAMS valid ?
  - Is « Select species » implemented ?

VAMDC portal



Species database

- Repository of all species contained in the infrastructure, sorted by database
- http://species.vamdc.eu
- Browsable through a web site to find quickly where a species can be found,
- Data can be exported in a xls file, easy to sort or to convert to csv
- Queryable through an API :
  - http://species.vamdc.eu/api/v12.07/nodes
  - http://species.vamdc.eu/api/v12.07/species
- Returns JSON structured data

VAMDC Portal

- http://portal.vamdc.eu
- Main entry point to look for data
- Provides two interfaces to build VAMDC-TAP requests and query all nodes
- Uses all elements of the infrastructure :
  - Registry to get a list of nodes that are queried
  - Species database to autocomplete species names
  - XSAMS transformation tools (xml is converted to html)

Further evolutions

- Updates in the XSAMS schema
- Including new data nodes (NIST is almost done)
- Providing a «simplified » database registration procedure into the infrastructure :
  - Answer only to «select species »
  - Get visibility in the species database
  - Not included in the VAMDC portal