



Discussions for compatibilities and integration between the VAMDC standards and SLAP/SSLDM

N. Moreau, C.M. Zwölf, M.L. Dubernet
Trieste Interop, October 2016



Introduction

This presentation contains two parts:

1. Technical discussion on three possible ways for integrating SLAP access into the VAMDC infrastructure (Moreau).
2. Convergence between VAMDC-XSAMS and SLAP-SSLDM: Elements for preparing the discussion (Zwölf).

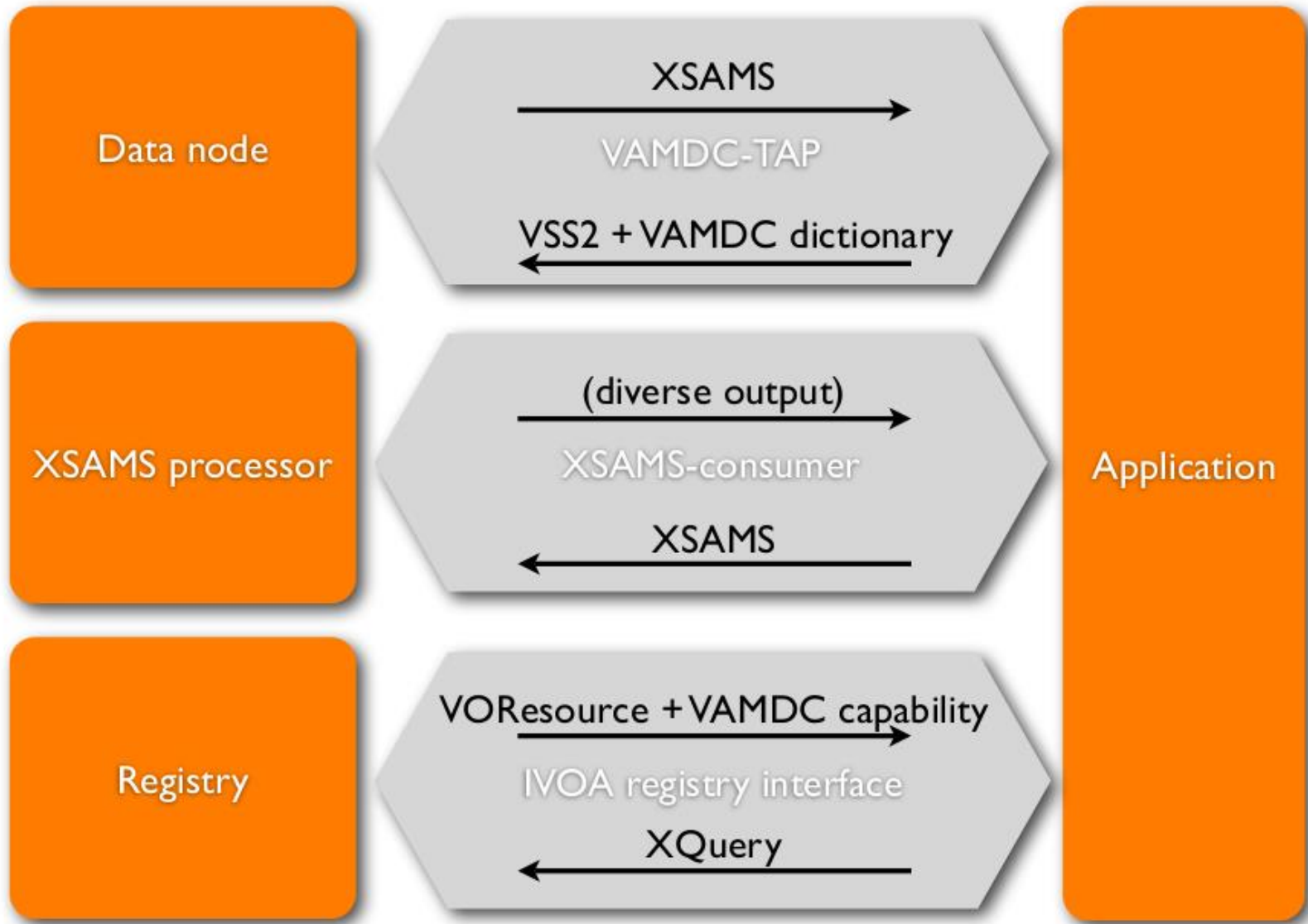
Part 1.

INTEGRATION OF SLAP ACCESS INTO VAMDC

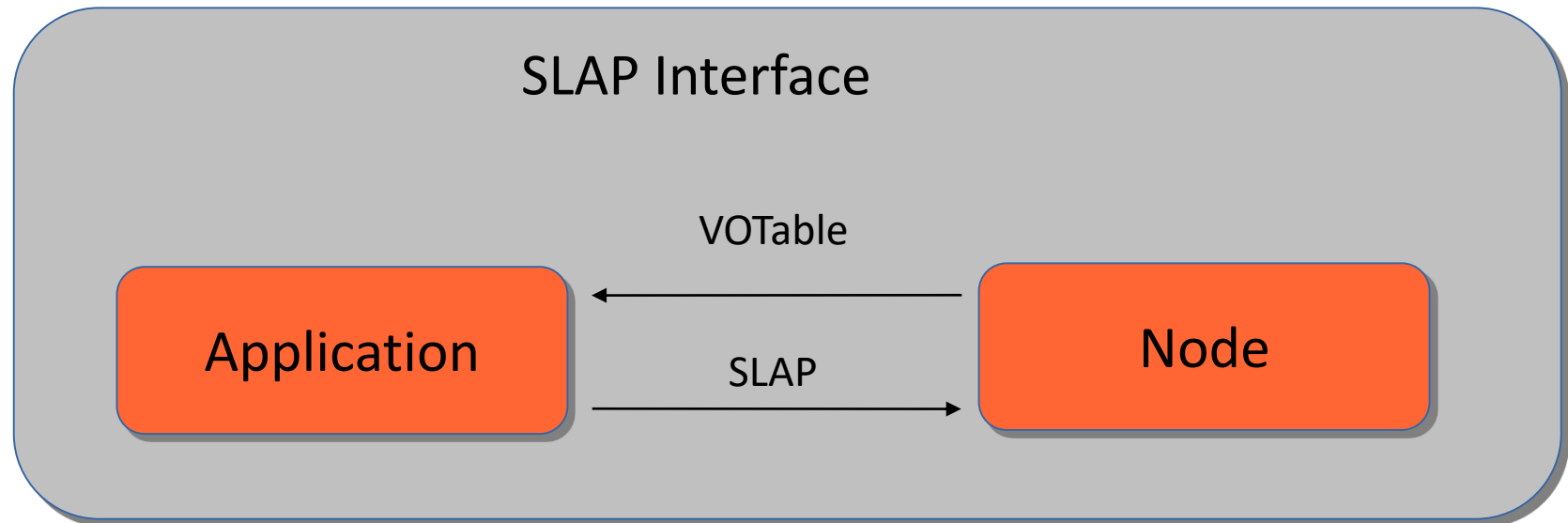
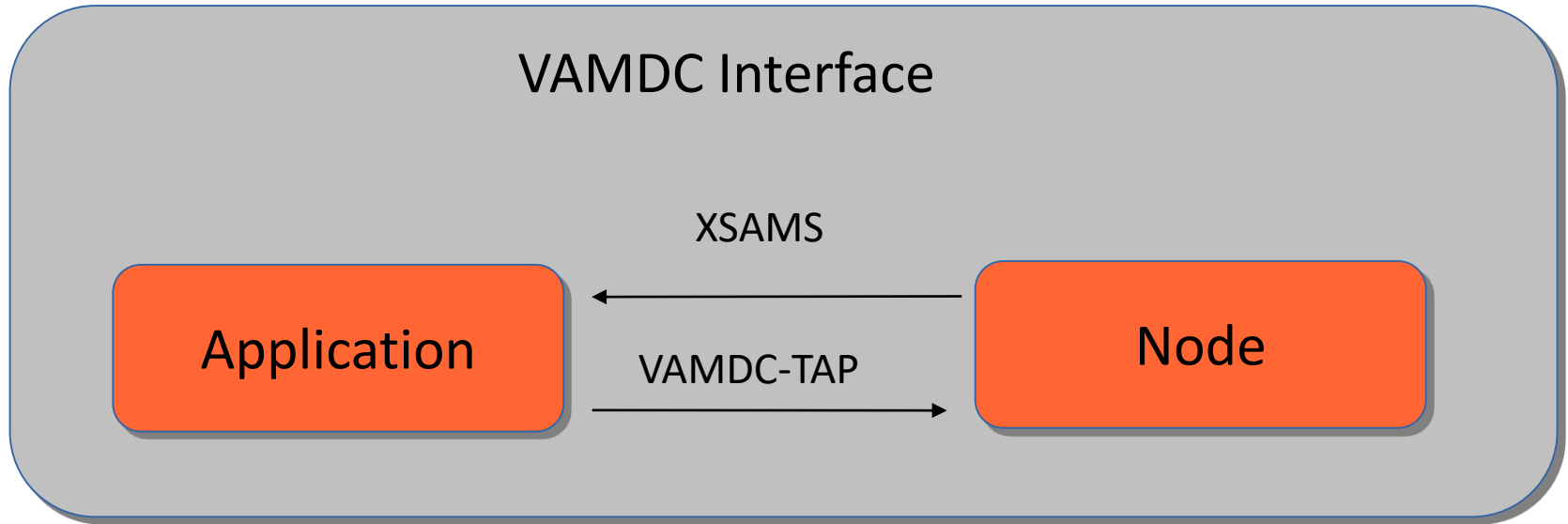
SLAP / SSLDM

- Simple Line Access Protocol is a data access protocol to retrieve spectral lines
- SSLDM is the corresponding data model
- Parameterized queries with at least a wavelength interval
- Returns a VOTable with corresponding quantities

VAMDC Infrastructure



Solution 1 : Direct integration in node



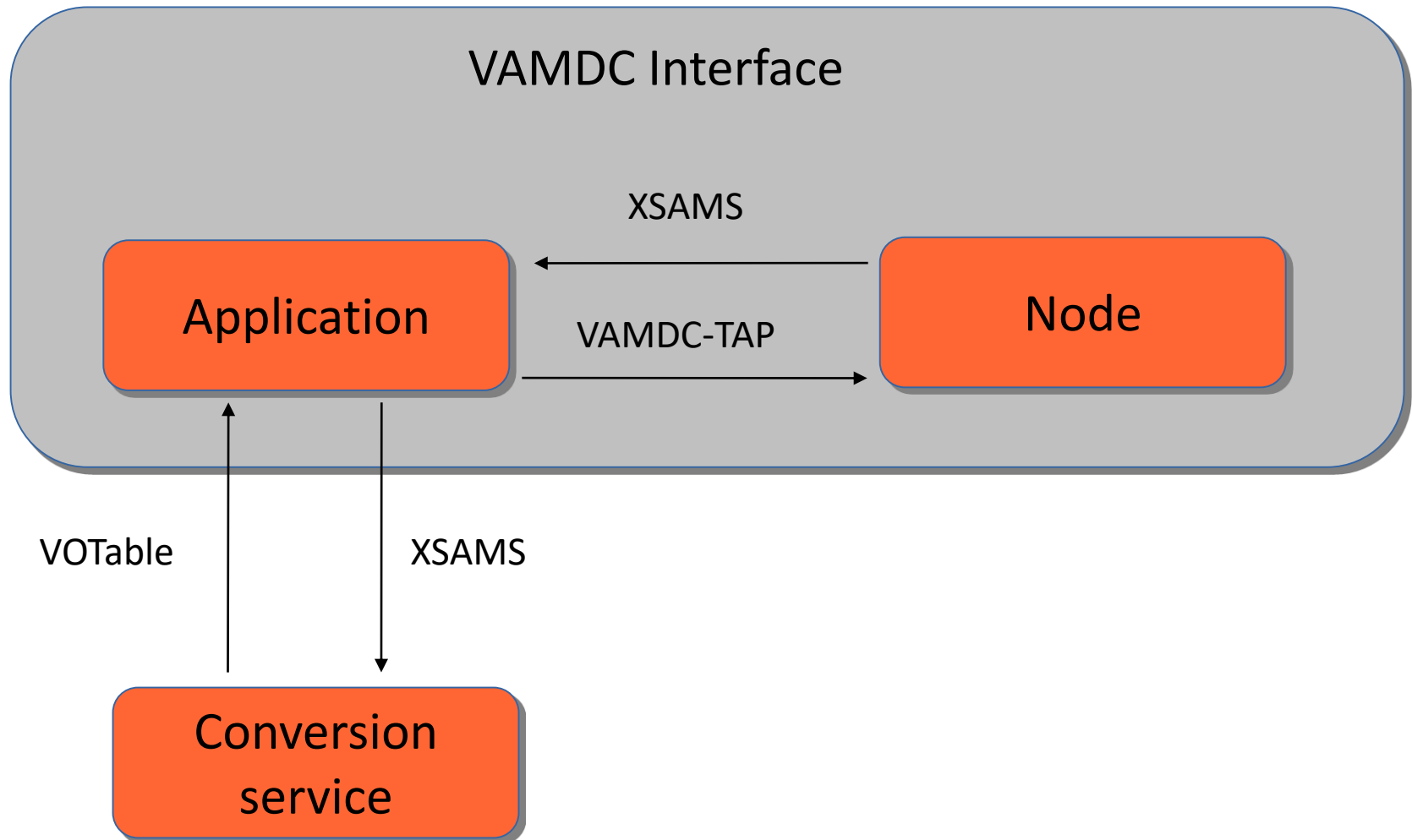
Solution 1 : Advantages

- Quite easy to implement in the node software
- No need to transform SLAP request into VAMDC-TAP or XSAMS into VOTable
 - Natively understood and generated
- All nodes are independant SLAP services and can be registered as such
- Could be a solution for big requests
- Data producer can choose by themselves what to return in VOTable and how to format it

Solution 1 : Drawbacks

- Must be manually deployed on each node each time standards are updated
- Update must be made when XSAMS or SSLDM changes
- Formatting of identical quantities could differ (quantum numbers) between nodes

Solution 2 : Result conversion



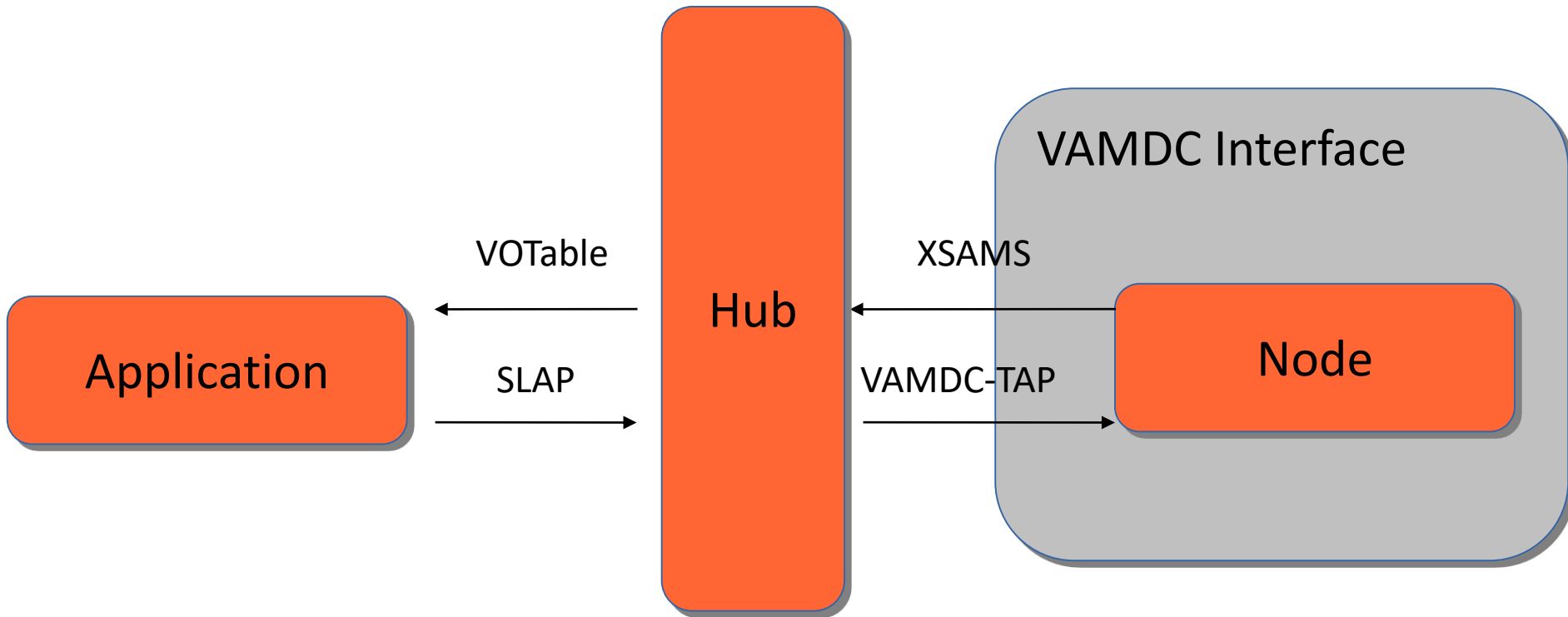
Solution 2 : Advantages

- Very loose coupling with infrastructure
- No impact on the nodes
- Only this service must be updated when standards changes
- Possibility to convert files generated by a code or manually

Solution 2 : Drawbacks

- Not a SLAP implementation, only SSLDM
- Necessary to convert XSAMS to VOTable
- The service will have to know all the quantities that can be exported from a XSAMS file to a VOTable
- It will have to test their existence in the XSAMS document

Solution 3 : Intermediary layer



Solution 3 : Advantages

- Complete SLAP integration without impact on the nodes
- Only this service must be updated if SLAP standards are changed
- Formatting of output will be identical for all the nodes

Solution 3 : Drawbacks

- Necessary to convert SLAP to VAMDC-TAP and XSAMS to VOTable
- This service must be aware of all the quantities that can be converted from XSAMS to SLAP VOTable
- It will not be node specific
- It will probably return a common subset of fields

Part 2.

PREPARING THE DISCUSSION ON THE DATA MODELS CONVERGENCE

Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

We extract from **CDMS** and from **Splatalogue** all the radiative data having a wavelength between 0.0165066m and 0.0165067m.

- On this restriction, the data from **CDMS** and **Splatalogue** are supposed to be the same, since this last takes the data from **CDMS**.
- The extracted data concern the molecule Cyanoacetylene (Propynenitrile)

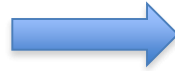
Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

We extract from **CDMS** and from **Splatalogue** all the radiative data having a wavelength between 0.0165066m and 0.0165067m.

- On this restriction, the data from **CDMS** and **Splatalogue** are supposed to be the same, since this last takes the data from **CDMS**.
- The extracted data concern the molecule Cyanoacetylene (Propynenitrile)

The species Cyanoacetylene (Propynenitrile):

- is called HCC13CN on **CDMS**
- is called HC13CCN on **Splatalogue**



We discovered it was the same specie by

- using the **CDMS** internal ID (non standard and not in SLAP output, only on **Splatalogue** site) and
- manually comparing the energies states and transitions

Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

We extract from **CDMS** and from **Splatalogue** all the radiative data having a wavelength between 0.0165066m and 0.0165067m.

- On this restriction, the data from **CDMS** and **Splatalogue** are supposed to be the same, since this last takes the data from **CDMS**.
- The extracted data concern the molecule Cyanoacetylene (Propynenitrile)

The species Cyanoacetylene (Propynenitrile):

- is called HCC13CN on **CDMS**
- is called HC13CCN on **Splatalogue**



We discovered it was the same specie by

- using the **CDMS** internal ID (non standard and not in SLAP output, only on **Splatalogue** site) and
- manually comparing the energies states and transitions

Let us take a look to the output file from **CDMS-XSAMS** and from **Splatalogue-SLAP** (mind the colour convention!)

Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

Definition of the chemical species:

```
<MolecularChemicalSpecies>
  <OrdinaryStructuralFormula>
    <Value>HCC-13-CN</Value>
  </OrdinaryStructuralFormula>
  <StoichiometricFormula>C3HN</StoichiometricFormula>
  <ChemicalName>
    <Value>Cyanoacetylene, Propynenitrile</Value>
  </ChemicalName>
  <InChI>1S/C3HN/c1-2-3-4/h1H/i2+1</InChI>
  <InChIKey>LNDJVIYUJ0JFS0-VQEHIDDOSA-N</InChIKey>
  <VAMDCSpeciesID>LNDJVIYUJ0JFS0-VQEHIDDOSA-N</VAMDCSpeciesID>
  <StableMolecularProperties>
    <MolecularWeight>
      <Value units="unitless">52</Value>
    </MolecularWeight>
  </StableMolecularProperties>
  <Comment> 52515- v1*:HCC-13-CN; $v7=1$</Comment>
</MolecularChemicalSpecies>
```

In XSAMS schema the species are uniquely identified using their standard InChiKey

(we maintain a catalogue of all the available species at species.vamdc.org)

Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

Definition of the energy states involved into a radiative transition:

```
<FIELD ID="lowerstateenergy" name="lowerstateenergy" datatype="double">
  <DESCRIPTION>Lower state energy</DESCRIPTION>
</FIELD>
<FIELD ID="upperstateenergy" name="upperstateenergy" datatype="double"
  utype="ssldm:Line.initialLevel.energy">
  <DESCRIPTION>Upper state energy</DESCRIPTION>
</FIELD>
<FIELD ID="lowerstateenergyK" name="lowerstateenergyK" datatype="double">
  <DESCRIPTION>Lower state energy in Kelvin</DESCRIPTION>
</FIELD>
<FIELD ID="upperstateenergyK" name="upperstateenergyK" datatype="double">
  <DESCRIPTION>Upper state energy in Kelvin</DESCRIPTION>
</FIELD>
<TD>219.8545</TD>
<TD>220.460317852162</TD>
<TD>316.319764433268</TD>
<TD>317.19139617282906</TD>
```

What is the unit used for energy?
According to SSLDM it should be "J".
This does not fit with the conversion in
Kelvin!
Let us look on CDMS...

Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

Definition of the energy states involved into a radiative transition:

```
<FIELD ID="lowerstateenergy" name="lowerstateenergy" datatype="double">
  <DESCRIPTION>Lower state energy</DESCRIPTION>
</FIELD>
<FIELD ID="upperstateenergy" name="upperstateenergy" datatype="double"
  utype="ssldm:Line.initialLevel.energy">
  <DESCRIPTION>Upper state energy</DESCRIPTION>
</FIELD>
<FIELD ID="lowerstateenergyK" name="lowerstateenergyK" datatype="double">
  <DESCRIPTION>Lower state energy in Kelvin</DESCRIPTION>
</FIELD>
<FIELD ID="upperstateenergyK" name="upperstateenergyK" datatype="double">
  <DESCRIPTION>Upper state energy in Kelvin</DESCRIPTION>
</FIELD>
<TD>219.8545</TD>
<TD>220.460317852162</TD>
<TD>316.319764433268</TD>
<TD>317.19139617282906</TD>
```

What is the unit used for energy?
According to SSLDM it should be "J".
This does not fit with the conversion in
Kelvin!
Let us look on CDMS...

```
<MolecularState stateID="SCDMS-3293669">
  <MolecularStateCharacterisation>
    <StateEnergy methodRef="MCDMS-3588" energyOrigin="SCDMS-None-origin-418">
      <Value units="1/cm">219.85447</Value>
    ...
  </MolecularStateCharacterisation>
</MolecularState>
<MolecularState stateID="SCDMS-3293670">
  <MolecularStateCharacterisation>
    <StateEnergy methodRef="MCDMS-3588" energyOrigin="SCDMS-None-origin-418">
      <Value units="1/cm">220.460288</Value>
    ...
  </MolecularStateCharacterisation>
</MolecularState>
```

XSAMS explicitly defines the units. By cross-matching the data we find that the unit used in SLAP is actually (1/cm).

Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

About the accuracy of values:

```
<FIELD ID="lowerstateenergy" name="lowerstateenergy" datatype="double">
  <DESCRIPTION>Lower state energy</DESCRIPTION>
</FIELD>
<FIELD ID="upperstateenergy" name="upperstateenergy" datatype="double"
  utype="ssldm:Line.initialLevel.energy">
  <DESCRIPTION>Upper state energy</DESCRIPTION>
</FIELD>
<FIELD ID="lowerstateenergyK" name="lowerstateenergyK" datatype="double">
  <DESCRIPTION>Lower state energy in Kelvin</DESCRIPTION>
</FIELD>
<FIELD ID="upperstateenergyK" name="upperstateenergyK" datatype="double">
  <DESCRIPTION>Upper state energy in Kelvin</DESCRIPTION>
</FIELD>
<TD>219.8545</TD>
<TD>220.460317852162</TD>
<TD>316.319764433268</TD>
<TD>317.19139617282906</TD>
```

These values are the same.

```
<MolecularState stateID="SCDMS-3293669">
  <MolecularStateCharacterisation>
    <StateEnergy methodRef="MCDMS-3588" energyOrigin="SCDMS-None-origin-418">
      <Value units="1/cm">219.85447</Value>
    ...
  </MolecularStateCharacterisation>
</MolecularState>
<MolecularState stateID="SCDMS-3293670">
  <MolecularStateCharacterisation>
    <StateEnergy methodRef="MCDMS-3588" energyOrigin="SCDMS-None-origin-418">
      <Value units="1/cm">220.460288</Value>
    ...
  </MolecularStateCharacterisation>
</MolecularState>
```

Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

About the accuracy of values:

```
<FIELD ID="lowerstateenergy" name="lowerstateenergy" datatype="double">
  <DESCRIPTION>Lower state energy</DESCRIPTION>
</FIELD>
<FIELD ID="upperstateenergy" name="upperstateenergy" datatype="double"
  utype="ssldm:Line.initialLevel.energy">
  <DESCRIPTION>Upper state energy</DESCRIPTION>
</FIELD>
<FIELD ID="lowerstateenergyK" name="lowerstateenergyK" datatype="double">
  <DESCRIPTION>Lower state energy in Kelvin</DESCRIPTION>
</FIELD>
<FIELD ID="upperstateenergyK" name="upperstateenergyK" datatype="double">
  <DESCRIPTION>Upper state energy in Kelvin</DESCRIPTION>
</FIELD>
<TD>219.8545</TD>
<TD>220.460317852162</TD>
<TD>316.319764433268</TD>
<TD>317.19139617282906</TD>
```

These values are the same.

These values are different:

- Splatalogue has more significant digit
- Which value is the “original” to trust?
- Where the difference come from?

```
<MolecularState stateID="SCDMS-3293669">
  <MolecularStateCharacterisation>
    <StateEnergy methodRef="MCDMS-3588" energyOrigin="SCDMS-None-origin-418">
      <Value units="1/cm">219.85447</Value>
    ...
  </MolecularStateCharacterisation>
</MolecularState>
<MolecularState stateID="SCDMS-3293670">
  <MolecularStateCharacterisation>
    <StateEnergy methodRef="MCDMS-3588" energyOrigin="SCDMS-None-origin-418">
      <Value units="1/cm">220.460288</Value>
    ...
  </MolecularStateCharacterisation>
</MolecularState>
```


Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

About the accuracy of values:

```
<FIELD ID="lowerstateenergy" name="lowerstateenergy" datatype="double">
  <DESCRIPTION>Lower state energy</DESCRIPTION>
</FIELD>
<FIELD ID="upperstateenergy" name="upperstateenergy" datatype="double"
  utype="ssldm:Line.initialLevel.energy">
  <DESCRIPTION>Upper state energy</DESCRIPTION>
</FIELD>
<FIELD ID="lowerstateenergyK" name="lowerstateenergyK" datatype="double">
  <DESCRIPTION>Lower state energy in Kelvin</DESCRIPTION>
</FIELD>
<FIELD ID="upperstateenergyK" name="upperstateenergyK" datatype="double">
  <DESCRIPTION>Upper state energy in Kelvin</DESCRIPTION>
</FIELD>
<TD>219.8545</TD>
<TD>220.460317852162</TD>
<TD>316.319764433268</TD>
<TD>317.19139617282906</TD>
```

These values are the same.

These values are different:

- Splatalogue has more significant digit
- Which value is the “original” to trust?
- Where the difference come from?

```
<MolecularState stateID="SCDMS-3293669">
  <MolecularStateCharacterisation>
    <StateEnergy methodRef="MCDMS-3585" energyOrigin="SCDMS-None-origin-418">
      <Value units="1/cm">219.85447</Value>
    ...
  </MolecularStateCharacterisation>
</MolecularState>
<MolecularState stateID="SCDMS-3293670">
  <MolecularStateCharacterisation>
    <StateEnergy methodRef="MCDMS-3585" energyOrigin="SCDMS-None-origin-418">
      <Value units="1/cm">220.460288</Value>
    ...
  </MolecularStateCharacterisation>
</MolecularState>
```

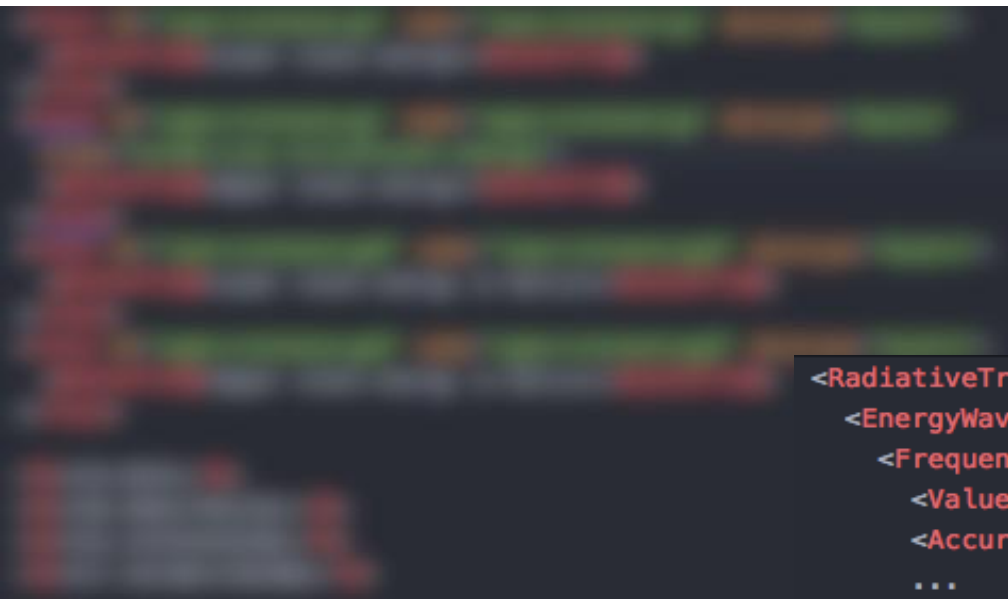
Splatalogue compute the upper energy as:

$$E_{up} = E_{low} + h\nu$$

The number of digits comes from the Planck constant.

Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

About the accuracy of values:



The wavelength accuracy is explicitly defined in XSAMS.

```
<RadiativeTransition id="PCDMS-R7672603" process="excitation">
  <EnergyWavelength>
    <Frequency methodRef="MCDMS-3971">
      <Value units="MHz">18161.9623</Value>
      <Accuracy>0.0009</Accuracy>
    ...
  </EnergyWavelength>
</RadiativeTransition>
```

```
<MolecularState stateID="SCDMS-3293669">
  <MolecularStateCharacterisation>
    <StateEnergy methodRef="MCDMS-3588" energyOrigin="SCDMS-None-origin-418">
      <Value units="1/cm">219.85447</Value>
    ...
  </MolecularStateCharacterisation>
</MolecularState>

<MolecularState stateID="SCDMS-3293670">
  <MolecularStateCharacterisation>
    <StateEnergy methodRef="MCDMS-3588" energyOrigin="SCDMS-None-origin-418">
      <Value units="1/cm">220.460288</Value>
    ...
  </MolecularStateCharacterisation>
</MolecularState>
```

Splatalogue compute the upper energy as:

$$E_{up} = E_{low} + h\nu$$

The number of digits comes from the Planck constant.

Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

Tracing the origin and scientific sources for data

```
<Sources>
  <Source sourceID="BCDMS-2016-09-27-13-54-53">
    <Comments> This Source is a self-reference.
      It represents the database and the query that produced the xml document.
      The sourceID contains a timestamp.
      The full URL is given in the tag UniformResourceIdentifier but you need
      to unescape ampersands and angle brackets to re-use it.
      Query was: select * where (RadTransWavelength &gt;= 1.650657735932166E8 AND RadTransWavelength &lt;= 1.6507486261769727E8)
    </Comments>
    <Year>2016</Year>
    <Category>database</Category>
    <UniformResourceIdentifier>http://cdms.ph1.uni-koeln.de/cdms/tap/sync?LANG=VSS2&REQUEST=doQuery&FORMAT=XSAMS&QUERY=select \* where \(RadTransWavelength >= 1.650657735932166E8 AND RadTransWavelength <= 1.6507486261769727E8\)
    <ProductionDate>2016-09-27</ProductionDate>
    ...
  </Source>
</Sources>
```

- The database producing the data is documented (CDMS VAMDC node)
- The file is timestamped. We may know exactly what the version of CDMS was at the data production.
- The query used for extracting data is embedded into the data

Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

Tracing the origin and scientific sources for data

```
<Sources>
  <Source sourceID="BCDMS-2016-09-27-13-54-53">
    <Comments> This Source is a self-reference.
      It represents the database and the query that produced the xml document.
      The sourceID contains a timestamp.
      The full URL is given in the tag UniformResourceIdentifier but you need
      to unescape ampersands and angle brackets to re-use it.
      Query was: select * where (RadTransWavelength &gt;= 1.650657735932166E8 AND RadTransWavelength &lt;= 1.6507486261769727E8)
    </Comments>
    <Year>2016</Year>
    <Category>database</Category>
    <UniformResourceIdentifier>http://cdms.ph1.uni-koeln.de/cdms/tap/sync?LANG=VSS2&REQUEST=doQuery&FORMAT=XSAMS&QUERY=select
    <ProductionDate>2016-09-27</ProductionDate>
    ...
```

- The database producing the data is documented (CDMS VAMDC node)
- The file is timestamped. We may know exactly what the version of CDMS was at the data production.
- The query used for extracting data is embedded into the data

```
<DESCRIPTION>Splatalogue SLAP Service</DESCRIPTION>
<INFO name="QUERY_STATUS" value="OK"/>
<INFO name="dbName" value="splat"/>
<INFO name="tableName"/>
<INFO name="REQUEST" value="queryData"/>
<INFO name="WAVELENGTH" value="0.0165066/0.0165067"/>
<INFO name="ServiceEngine" value="slap: SLAP version 1.0 DALServer version 0.3"/>
<INFO name="VERB" value="3"/>
<INFO name="TableRows" value="1"/>
...
```

- The database producing the data is documented
- The query generating the data file is documented
- There is no timestamping or versioning of data (just the version of standards).

Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

Tracing the origin and scientific sources for data

Tracing data version and sources is fundamental for the reproducibility of the scientific process and trustworthiness of analysis

Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

Tracing the origin and scientific sources for data

```
...  
<Source sourceID="BCDMS-141">  
  <Authors>  
    <Author>  
      <Name>Thorwirth, S.</Name>  
    </Author>  
    <Author>  
      <Name>Müller, H. S. P.</Name>  
    </Author>  
    <Author>  
      <Name>Winnewisser, G.</Name>  
    </Author>  
  </Authors>  
  <Title>  
</Title>  
  <Category>journal</Category>  
  <Year>2001</Year>  
  <SourceName>Phys. Chem. Chem. Phys.</SourceName>  
  <Volume>3</Volume>  
  <PageBegin>1236</PageBegin>  
  <DigitalObjectIdentifier>10.1039/b009743h</DigitalObjectIdentifier>  
</Source>  
...
```

Tracing data version and sources is fundamental for the reproducibility of the scientific process and trustworthiness of analysis

- The result file contains the references to the scientific sources (i.e. papers from journals) used for compiling the data file.

Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

Tracing the origin and scientific sources for data

```
...
<Source sourceID="BCDMS-141">
  <Authors>
    <Author>
      <Name>Thorwirth, S.</Name>
    </Author>
    <Author>
      <Name>Müller, H. S. P.</Name>
    </Author>
    <Author>
      <Name>Winnewisser, G.</Name>
    </Author>
  </Authors>
  <Title>
  </Title>
  <Category>journal</Category>
  <Year>2001</Year>
  <SourceName>Phys. Chem. Chem. Phys.</SourceName>
  <Volume>3</Volume>
  <PageBegin>1236</PageBegin>
  <DigitalObjectIdentifier>10.1039/b009743h</DigitalObjectIdentifier>
</Source>
...
```

Tracing data version and sources is fundamental for the reproducibility of the scientific process and trustworthiness of analysis

- The result file contains the references to the scientific sources (i.e. papers from journals) used for compiling the data file.

- No references to the articles used for compiling the data file.



Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

Focusing on quantum numbers

- The ways for defining quantum numbers (QNs) in SLAP/SSLDM are relatively free.
- As a consequence, QNs may be expressed in different ways by data providers.
- Only a human may understand them (usually an expert is required)
- No automatic processing of QNs is possible

```
<FIELD ID="QNs" name="quantum numbers" datatype="char"
  ucd="phys.atmol.transition;spect.line" arraysize="*">
  <DESCRIPTION>The initial and final quantum number states that
    produces this line.</DESCRIPTION>
</FIELD>
...
<DATA>
  <TABLEDATA>
    <TR>
      ...
      <TD>J= 2- 1<i>, l=1e</i></TD>
      ...
    </TR>
  </TABLEDATA>
```


Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

Focusing on quantum numbers

- The ways for defining quantum numbers (QNs) in SLAP/SSLDM are relatively free.
- As a consequence, QNs may be expressed in different ways by data providers.
- Only a human may understand them (usually an expert is required)
- No automatic processing of QNs is possible

```
<FIELD ID="QNs" name="quantum numbers" datatype="char"
  ucd="phys.atmol.transition;spect.line" arraysizes="*">
  <DESCRIPTION>The initial and final quantum number states that
    produces this line.</DESCRIPTION>
</FIELD>
...
<DATA>
  <TABLEDATA>
    <TR>
      ...
      <TD>J= 2- 1<i>, l=1e</i></TD>
      ...
    </TR>
  </TABLEDATA>
```

- In XSAMS the definition of QNs is more constrained.
- The definitions are contained into a separated ontology (the case by case:
<http://www.vamdc.eu/documents/cbc-1.0/>)
- The xsams document refer to a point of this ontology (a particular case) for defining the QNs frame.
- Using this ontology, automatic cross-matching of state is possible (implemented in SpectCol).

```
<MolecularState stateID="SCDMS-3293669">
  ...
  <Case xsi:type="lpcs:Case" caseID="lpcs"
    xmlns:lpcs="http://vamdc.org/xml/xsams/1.0/cases/lpcs"
    xsi:schemaLocation="http://vamdc.org/xml/xsams/1.0/cases/lpcs ../../cases/lpcs.xsd">
    <lpcs:QNs>
      <lpcs:ElecStateLabel>X</lpcs:ElecStateLabel>
      <lpcs:vi mode="7">1</lpcs:vi>
      <lpcs:li mode="7">1</lpcs:li>
      <lpcs:J>1</lpcs:J>
      <lpcs:kronigParity>e</lpcs:kronigParity>
    </lpcs:QNs>
  </Case>
</MolecularState>
```

Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

Summing up.

- It takes us approx. 3 hours for manually cross-matching the two files as described into this presentation.
- It should be easier than that!! How to deal with hundreds of lines otherwise?

Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

Summing up.

- It takes us approx. 3 hours for manually cross-matching the two files as described into this presentation.
 - It should be easier than that!! How to deal with hundreds of lines otherwise?
-
- We are hoping for a rapprochement between SLAP/SSLDM and XSAMS for
 - Unequivocally defining species
 - Defining explicitly the units
 - Introducing references to the scientific sources
 - Introducing versions of data
 - Converging to a common way for defining Quantum Numbers.

Feedbacks from spectroscopic data extraction and matching from VAMDC and SLAP resources

Summing up.

- It takes us approx. 3 hours for manually cross-matching the two files as described into this presentation.
- It should be easier than that!! How to deal with hundreds of lines otherwise?

- We are hoping for a rapprochement between SLAP/SSLDM and XSAMS for
 - Unequivocally defining species
 - Defining explicitly the units
 - Introducing references to the scientific sources
 - Introducing versions of data
 - Converging to a common way for defining Quantum Numbers.

Let the discussion start...