VAMDC

Virtual Atomic and Molecular Data Centre

D7.2

–

Publishing Tools Report 1

Version 0.2

Grant agreement no: 239108

Combination of Collaborative Projects & Coordination and Support Actions

Project Information

Project acronym: VAMDC

Project full title: Virtual Atomic and Molecular Data Centre

Grant agreement no.: 239108

Funding scheme: Combination of Collaborative Projects & Coordination and Support Actions

Project start date: 01/07/2009

Project duration: 42 months

Call topic: INFRA-2008-1.2.2 Scientific Data Infrastructure

Project web sites: http://www.vamdc.eu

http://voparis-twiki.obspm.fr/twiki/bin/view/VAMDC/WebHome

Consortium:

Beneficiary Number *	Beneficiary name	Beneficiary short name	Country	Date enter project**	Date exit project**
1(coordinator)	Centre National de la Recherche Scientifique	CNRS	France	Month 1	Month 42
2	The Chancellor, Masters and Scholars of the University of Cambridge	CMSUC	UK	Month 1	Month 42
3	University College London	UCL	UK	Month 1	Month 42
4	Open University	OU	UK	Month 1	Month 42
5	Universitaet Wien	UNIVIE	Austria	Month 1	Month 42
6	Uppsala Universitet	UU	Sweden	Month 1	Month 42
7	Universitaet zu Koeln	KOLN	Germany	Month 1	Month 42
8	Istituto Nazionale di Astrofisica	INAF	Italy	Month 1	Month 42
9	Queen's University Belfast	QUB	UK	Month 1	Month 42
10	Astronomska opservatorija	AOB	Serbia	Month 1	Month 42
11	Institute for Spectroscopy RAS	ISRAN	Russian Federation	Month 1	Month 42
12	Russian Federal Nuclear Centre All-Russian Institute of Technical Physics	RFNC-VNIITF	Russian Federation	Month 1	Month 42
13	Institute of Atmospheric Optics	IAO	Russian Federation	Month 1	Month 42
14	Corporacion Parque Tecnologico de Merida	CTPM	Venezuela	Month 1	Month 42
15	Institute of Astronomy of the Russian Academy of Sciences	INASAN	Russian Federation	Month 1	Month 42

This project is funded under *"Combination of Collaborative Projects and Coordination and Support Actions"* Funding Scheme of The Seventh Framework Program of the European Union

Document

Deliverable number:	D7.2
Deliverable title:	Publishing Tools Report 1
Due date of deliverable:	June 2010
Actual submission date:	8th September 2010
Authors:	N. Piskunov, T. Marquart, P. Loboda, A. Fazliev, L. Nenadovic, M.L. Dubernet
Work Package no.:	WP7-JRA2
Work Package title:	Publishing Tools
Work Package leader:	UU
Lead beneficiary:	UU
Dissemination level:	PU
Nature:	Report
No of pages (incl. cover):	

Abstract	The objective of D7.2 is to describe VAMDC Publishing Tools Report for Period 1. This report corresponds to Activities in WP7: JRA2 "Publishing Tools". This report is included in the VAMDC Periodic Report for Period 1.

Versioning and Contribution history

Version	Date	Reason for modification	Modified by
V0.1	**20/07/2010**	**Section 5**	**N. Piskunov**
V0.2	**31/08/2010**	**Draft of D7.2**	**M.L. Dubernet & L. Nenanovic**

Final Version (v0.2) released by		Circulated to	
Name	Date	Recipient	Date
M.L. Dubernet	8th September 2010	Mrs Asero	8th September 2010

WP7 ACTIVITIES DESCRIPTION

Work package number	7		**Start date or starting event:**		3						
Work package title	JRA2: Publishing Tools										
Activity Type	RTD										
Participant id	1	6	8	12	13						
Person-months per beneficiary: **(Total = EU + Node Contributions)**	12	12	12	5	24						

Table of Content

1. WP7 Objectives

JRA2 provides generic tools for A&M data in the VAMDC infrastructure. We envisage two publishing paths: (A) inclusion of the new data into existing databases, which are already integrated into the VAMDC, and (B) publishing data in a new database instance using an existing open source relational database software instrumented with the VAMDC interface. The selection of specific path will be decided between the data producer and the VAMDC depending on the completeness of the new data, logical and structural similarity with existing databases, availability of the necessary resources at the producer site etc.
The process of publishing includes two steps: data quality assessment and the open offering of the new data to the VAMDC clients. In the first stage the new data will be technically included into the infrastructure but not made public. Instead, a group of experts working with the VAMDC will have access to the data in order to verify the quality assessment provided by the producer. After the quality assessment is completed the data will be offered in open access.
In order to implement this plan we need to formulate a number of procedures and develop or adopt a number of software tools. This work will not start from scratch: we will build on long experience accumulated by large databases handling many datasets from different data providers (e.g. VALD). In addition, several tasks in the WP7 are closely related to other work packages and thus require close and intensive collaboration.

WP7 Leader is UU(6).

2. WP7 Milestones and Deliverables

Milestones

M7.1	Technical Meetings	WP7	UU	Months 5,10, 16, 22, 28, 34, 40, 42	Minutes. Presentations on internal Website
M7.2	Evaluation of softwares	WP7	UU	Months 10, 22, 34	

Deliverables

D7.1 Publishing Tools Plan (PM 3)
D7.2 Publishing Tools Report to be included in report to the EU – Year 1 (PM 10)
D7.3 Publishing Tools Report to be included in report to the EU – Year 2 (PM 22)
D7.4 Publishing Tools Report to be included in report to the EU – Year 3 (PM 34)
D7.5 Final Report of Publishing Tools to be included in final report to the commission (PM41)

Annual Publishing Tools Plan revisions included in Revised Annual VAMDC Project Plans – Year 1,2,3

3. WP7 Tasks Description

WP7 Leader (co)		
Task Number	Leader	Other Partners
1	M.-L. Dubernet/M. Doronin (CNRS)	All others
2	M.-L. Dubernet/M. Doronin (CNRS)	All others
3	N. Piskunov (UU)	All others
4	P. Loboda (RFNC-VNIIT)	All others
5	A. Fazlief (IAO)	All others

Description of work (possibly broken down into tasks)

This WP will develop software that will be deployed within the VAMDC infrastructure. Significant part of this software will be based on the standards developed in JRA1. The general software available to the VAMDC community will be accessible via the VAMDC technical website. Two alternative solutions are being developed. The first one offers a possibility to import new data to the existing informational resources, capable of importing data in standardized forms (tasks 2-4) developed within JRA1. The second option implies the design of a generic information system accessible via the Internet (task 5) instrumented with the VAMDC integration tools developed in tasks 1-3. In this variant an automatic generation of semantic metadata for uploaded information resources is realized, taking into account the restrictions imposed by formal models of molecules and atoms. All software will be documented.

Task 1: Create/adapt tools to go from a DM/XML schema to a full database deployment with generation of automatic administrative interface. ***(lead by CNRS(1) with (6))***
Task 2: Create/adapt tools to build registries from the content of databases ***(lead by CNRS(1) with (6))***
Task 3: Create/adapt interfaces to easily update dictionaries ***(lead by UU(6) with (1))***

Task 4: Develop software libraries using various languages allowing to easily generate output of already existing resources in standardized format ***(lead by RFNC-VNIIT(12) with (1), (6), (8))***
Task 5: Create tools to upload, modify, retrieve, compare, visualize and publish information in molecular spectroscopy ***(lead by IAO(13))***

4. WP7 Tasks Description for Period 1

The WP7 Work Plan includes final products of two types: formulation of a number of procedures and development/adoption of a number of software tools. Several tasks in the WP7 are closely related to other work packages and require close and intensive collaboration.

The procedures to be formulated for the Publishing Tools WP7 are:

- *Preparation rules for the original data.* We intend to offer reasonable flexibility for the format of data sets to be imported to a VAMDC database but the format must be restricted by small number of clear and easy-to-follow rules. These rules will ensure full description of the data (physical meaning, units, bibliography etc.) and the cross-referencing if the data comes in several parts. In the end we may support several formats but we start by developing a tool for importing data in form of ASCII tables. This document will also describe a unified DB interface and the methods of tuning it to the existing DBs.
- *Guidelines for the expert quality assessment.* This document will contain a summary of data quality assessment procedures adopted in major existing databases (e.g. VALD, NIST etc.) and a description of the data access protocol, available tools and data usage policy during the assessment period. This document will be revised in the future incorporating the VAMDC own experience, expert comments and suggestions and therefore it must be updated on a regular basis.
- *Selection of an open-source relational database to be used in Path B.* This document will contain a report explaining the motivation behind the selection of specific software.
- *Detailed deployment/verification/maintenance instructions for Path B.* Although the software and the VAMDC interface for Path B will be provided by the WP7, the actual deployment and data publishing must be thoroughly tested before adding a new resource to VAMDC. Both procedures must be described very clearly under the assumption that the data producer does not have special training in installing and running relational databases. A critical requirement for Path B is service availability and sustainability on a long run. These issues will also be addressed by this document.
- *Statistics and data bibliography.* We have identified the lack of proper referencing to the original data producers as the major concern for publishing in large databases. Solving this issue will be a major incentive for the data producer to publish their data through VAMDC. This document should specify the rules for storing and cross-referencing the bibliographic and A&M data as well as the logs of processed requests so any request can be traced back to the source of the extracted data.

Software tools to be developed in WP7:

- VAMDC interface for publishing new A&M data
- Data import tool
- Automatic data verification tool

- VAMDC interface for a selected open source database with the support of VAMDC data exchange protocols
- VAMDC interface for data quality experts (Quality assessment report)

Coordination and Cooperation:
Many of the items to be developed in WP7 are closely related or dependent on other work packages and thus require very close coordination and often collaboration to avoid effort duplication and interface incompatibility. While those questions will be in competence of WP2 we feel that it is our responsibility to identify them. The following specific topics of the WP7 involve interaction with other work packages:

- The format of internal data representation/transport in the VAMDC is critically important for the Data Preparation Rules.
- Guidelines for the quality assessment will be prepared by the WP7 but must be discussed and approved by the VAMDC VPB (WP1 and WP2). See e.g. the point on "quality assurance systems" in Task 1 of WP2 plan for period 1.
- The selection of the open source database and its instrumentation is dependent on both the internal data representation and the registry query protocol adopted in VAMDC. This requires coordination with WP5 and WP6.
- Path B deployment document aimed at data providers must coordinated with WP4 but also with WP3.
- The cross-referencing system and the request logging mechanism must be distributed between individual databases and the VAMDC server. This is as much policy question as it is a technical issue. Thus it must be a cooperation topic with WP2 and WP4.
- The data Import Tool interface to the VAMDC database(s) must comply with the VAMDC internal data representation. This calls for cooperation with the WP6.
- The software package for the Path B must be coordinated with WP4 and WP5.
- Automatic data verification tool overlaps with SA2:T5 of the WP5.
- VAMDC interface will be mostly developed within WP6 but parts of it will be included in WP7 while testing/verification will involve WP4 and WP5.

Coordinating activities will be carried out through systematic updating of the wiki pages, extensive email exchange, use of version control software for the toolkit development and regular working meetings. The following working meetings have been planned for Period 1:

1. Coordination with WP6, Köln, February 2010
2. XSAMS and registry toolkits, Paris, March 2010
3. Automatic registry updates and referencing system within VAMDC, Vienna or Florence, March 2010
4. Test deployment of the VAMDC interface prototype for VALD, Moscow (TBC), May 2010

Here is a description of specific steps within each task planned for Period 1:

Task 1: The current version of the XSAMS will be used is a basis for creating a toolkit for various VAMDC applications. The toolkit will be primarily aimed at data conversion to and from XSAM to ASCII tables and other formats (April 29th, 2010, Doronin, Nenadovic, Dubernet).

Task 2: A prototype toolkit will be created for collecting exhaustive information about the content of and A&M database, available extraction tools and procedures as well the usage statistics (June 4th, 2010, Doronin, Nenadovic, Dubernet).

Task 3:

- Formulating the requirements for the "standard" VAMDC database software based on the analysis of the structure, content and selection tools of the existing databases (March 1st, 2010, Marquart, Heiter) Note: WP2 Task 1 includes a "census of all VAMDC resources" - the results of this task will be used in WP7/Task 3.
- Finalizing the WP7 plan, internal deliverables and time lines for each partner and writing Publishing Tools Report 1 (April 12th, 2010, Piskunov, Heiter).
- Preparing first version of specification for the format of the data to be imported into VAMDC (April 29th, 2010, Marquart, Piskunov).
- Preparing the prototype of the import tool (together with the RFNC-VNIIT, May 3rd, 2010, Marquart).
- Selection of the "short list" candidates for the VAMDC "standard" database (May 28th, 2010, Marquart, Stempels).
- Testing the import tool with VALD and Spectr W^3 databases (June 11th, 2010, Marquart, Heiter, Stempels).

Task 4: Prototype development and deployment of the VAMDC XSAMS interface for the VALD and Spectr W^3 databases (June 4th, 2010, Marquart, Laboda).

Task 5: A prototype automatic data verification tool developed for a “test” atom and CO-molecule will be created and tested for primary data sources. (May 28th, Fazliev).

- Formulating the requirements for the "standard" middleware based on the analysis of the atomic and molecular data and metadata structures and constructing first version of middleware (June 30, 2010, Akhlyostin)
- Preparing specification for the quantum numbers of CO-molecule, a “test” atom and the corresponding selection rules (March 1st, 2010, Perevalov, Ryabchikova)
- Preparing the prototype of the upload data system for the molecular data (CO-molecule transitions and line profiles data structures) from primary data sources (March 1, 2010, Lavrentiev, Privezentsev).
- Preparing the prototype of the upload data system for the atomic data (the atom transitions and line profiles data structures) from primary data sources (May 1, 2010, Lavrentiev, Privezentsev).
- Preparing the prototype of the tabular presentation system of molecular data (CO-molecule transitions and line profiles data structures) from primary data sources (April 1, 2010, Akhlyostin).
- Preparing the prototype of the tabular presentation system of atomic data (transitions and line profiles data structures) from primary data sources (April 1, 2010, Akhlyostin).
- Formulating the requirements for publication tools (analysis of the relations between the data owner and person reviewing the data). Preparing specification of these relations and prototype of the software implementing the publishing procedure. (May 28, 2010, Kozodoev)

Release of the software, documentation, test results, performance assessment and procedure documents for Tasks 1-5 will be available through links from the VAMDC web-site.

5. WP7 Tasks Reports for Period 1

VAMDC

Periodic Report Template (per Workpackage)

Period: 01/07/2009 – 30/06/2010
WorkPackage: WP7 Publishing tools
WorkPackage Leader and co-Leader: N. Piskunov (UU)
Participants in the WorkPackage: All Partners

Part 1

A summary of progress towards objectives and details for each tasks

Individual tasks of this work package are closely related to other work packages through the selected data model, VAMDC - individual database interface, registry content, query language and extraction tools, specific properties of existing databases etc. Thus, lots of activities described below is also relevant to various tasks in other WPs. The goals of the efforts within WP7 was during Cycle 1 was to (1) come up with a stable model for data description (together with WP6 Task 1), (2) to define a general, portable and reliable VAMDC-DB interface, (3) to propose a database-independent method for populating and maintaining VAMDC registry reflecting the data content of the whole infrastructure but also useful for interpreting user requests, (4) to demonstrated the functionality of the two proposed paths for publishing new data in VAMDC and (5) to explore various options for data producer interface, formal and expert data verification etc.
A number of tasks required parallel development in order to compare the alternative solutions: for example, we have implemented several XSAMS generators to compare the performance and data traffic for various database organizations, used frameworks etc.

One of the key solution which was proposed and implemented in Cycle 1 was the construction of VAMDC global dictionaries based on the XSAMS data model. The goal of a dictionary is to translate the names of data fields used in individual DBs and the common names used across VAMDC. These dictionaries are specific for each DB and they are the basis for the query construction, the conversion of the extracted data to a XSAMS structure and to the interaction with a registry. For the publishing tools the dictionaries are used for creating and registering new, VAMDC-compatible DB or for importing new data to an existing DB.

The work, the comparison of results and reporting were organized in a form of workshop series across all JRAs. During Cycle 1 we had three such workshops with the fourth planned for the early fall.

Significant results (Activities and Deliverables)

Task 1: Toolkit for conversion of data to and from XSAMS

Many of the items in WP7 are closely related or dependent on the tasks in other work packages and have required close collaboration between partners. WP6-T1 has made progress in the definition of VAMDC data format. In this cycle we have seen an evolution of XSAMS from the original version 0.1 (see schema and documentation on the IAEA XSAMS site), to a proposed modified schema with less nested structure for the representation of molecular states,

and a different " case-by-case" approach to the molecular state part of the schema proposed by C. Hill (UCL).
Tools have been developed for conversion of data from various databases to XSAMS. M. Doronin (CNRS) has implemented version 0.1, modified version 0.1, and is in the process of implementing the case-by-case schema on BASECOL; C. Hill (UCL) has implemented both version 0.1 and case-by-case on HITRAN; T. Marquart (UU) has implemented version 0.1 on VALD; C. Endres (KOLN) has implemented the case-by-case on CDMS.

Prototype of conversion form XSAMS to ASCII has been developed as a result of work on WP8-T2. It includes conversion of the energy values and quantum numbers associated with individual energy levels from XSAMS files to ASCII tables.

A meeting is planned in Paris on the 22nd of July 2010 in order to finalise the standard VAMDC data format. Pending the outcome of this meeting additional tools will be developed to convert data from XSAMS to VOTable and other formats.

Task 2: Registry population/updating from database content

This task is postponed for cycle 2. For cycle one registry is manually filled with resource records for databases, legacy access services, DSA/Catalog TAP service installations.

Task 3: Reference database structure and data import tools

The process of publication of the new data in VAMDC includes two possibilities: addition of the new data to the existing DB already equipped with a VAMDC interfaces and the creation of a new DB. For the 2nd option a selection of a recommended open-source database software was required. In order to make a choice we used a comprehensive study of existing DBs carried out by Theresa Lüftinger and performed a systematic analysis of the existing options. There results compiled by Thomas Marquart and Ulrike Heiter. The conclusions are based on the progress of the VAMDC-DB interface, in particular, the interface based on the Django framework. This promising technology simplifies several of the tasks: e.g. configuring a new DB according to the data content, configuring the VAMDC interface for a relational DB capable of accepting an SQL query and converting the results into an XSAMS structure with good performance. In order to test the ease of implementation and specific properties of this framework in connection with the publishing tools we have created a new database from the existing VALD data. The procedure included 3 steps:

1. Exporting a subset of the VALD3 data (about 5% of the total data content) in form of four tables (transition data, state description, references to the original data sources and the VALD3 list of species).
2. Creating a dictionary for translating the VALD field naming to the VAMDC XSAMS-compatible names.
3. Using Django framework for generating a new database with preconfigured VAMDC TAP-XSAMS interface, importing the data, generating the keys connecting species, states, transitions and the bibliography.
4. Testing data integrity and correctness of the DB structure by comparing results the extractions with the original VALD. This stage was also used to assess the performance and the requirements on the data traffic.

Note that the newly created DB was instantly compatible with the VAMDC in accepting SQL request using XSAMS-based dictionary and returning the extracted data in form of XSAMS structure or a VO table. This test was repeated with MySQL and SQLite open source databases. The test bench server can be accessed here and a

snapshot of the test site GUI is shown in Figure 1.

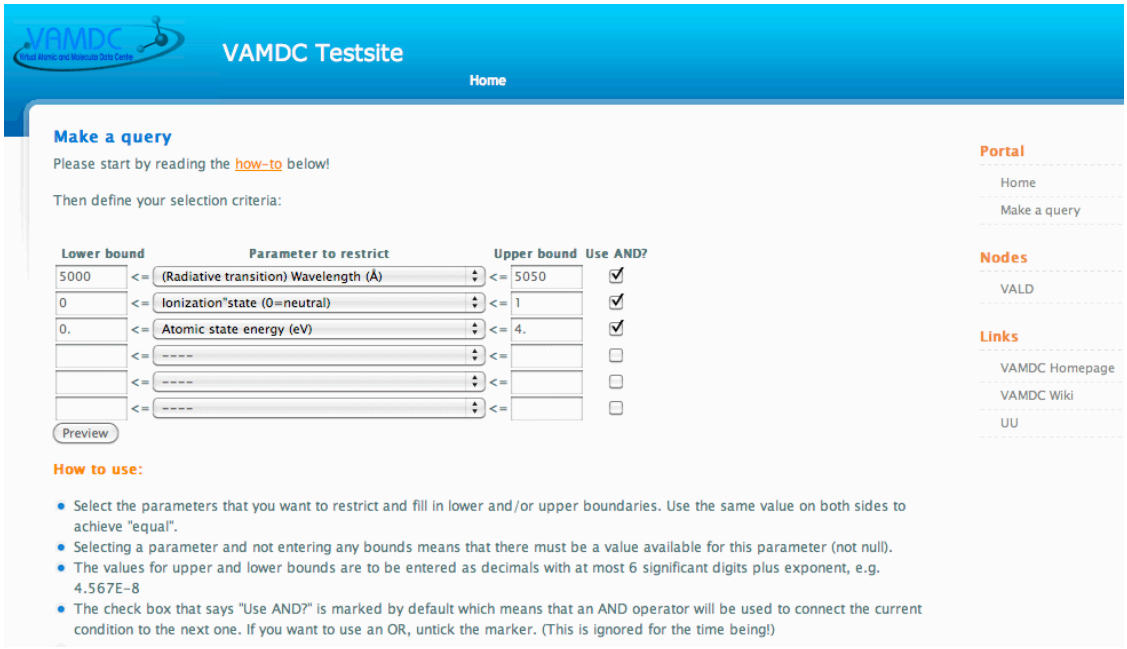


Figure 1. Snapshot of the VAMDC test bench at UU.

We have also tested the ways of adding new data to the created database and the same preconfigured Django frontend provided all the necessary functionality and flexibility. Finally, together with our partners we were able to configure Django frontend for three other DBs (CDMS, BASECOL and xstarDB). Porting the Django frontend to a new dataset requires the creation of two dictionaries and description of the data model. The current version of the full list of global keywords to be used in the dictionaries (the data dictionary and the query dictionary) can be found here while the description of the data model is to be coded as a Python function according to the guidelines and the example found here.

The conclusions are that the proposed prototype for the publishing tools (paths A and B) is:

- easy to use for creating a new database from data tables
- compatible with most of the relational databases
- easy to configure (mostly one needs to edit the two dictionaries, one function describing data model and one function for data import)
- compatible with the VAMDC data exchange standards
- compatible with the proposed query language and the VAMDC registry content.

Task 4: XML schema compatibility with Publishing Tools, data publishing GUI, development and testing the data import tool

During the 1st year of the VAMDC project implementation, an operating prototype of a special-purpose tool to export selected data from the Spectr-W3 database to XML format as well as to import XML data into Spectr-W^3 was improved and extended. Specifically, using current version 0.1 of XML Schema for Atoms, Molecules and Solids (XSAMS) being developed under the IAEA auspices, the tool has been

implemented for the ionization-potential, energy-level, and spectral-lines data sections of the off-line local analog of the Spectr-W^3 database (Spectr-CD) and for the ionization-potential and energy-levels sections of the Spectr-W^{3}database (in trial mode).
Initially, Spectr-CD was developed for the off-line use on PCs under Windows. In the process of the implementation of the ISTC project # 3504, previous interface of the Spectr-CD package was completely rewritten in C# that enabled to considerably improve its flexibility and functionality. At the moment, the Spectr-CD package is mostly employed for the development and testing of the export/ import options to XML & plain text formats. To generate XML output, all the fields to display in the requested-data tables should be selected. Figures 2 and 3 demonstrate the setting and the results of the Spectr-W^3-to-xSAMS data export (respectively) done for the energy levels of Co-like W by clicking the "Export to XSAMS" option of the "Search results" window.

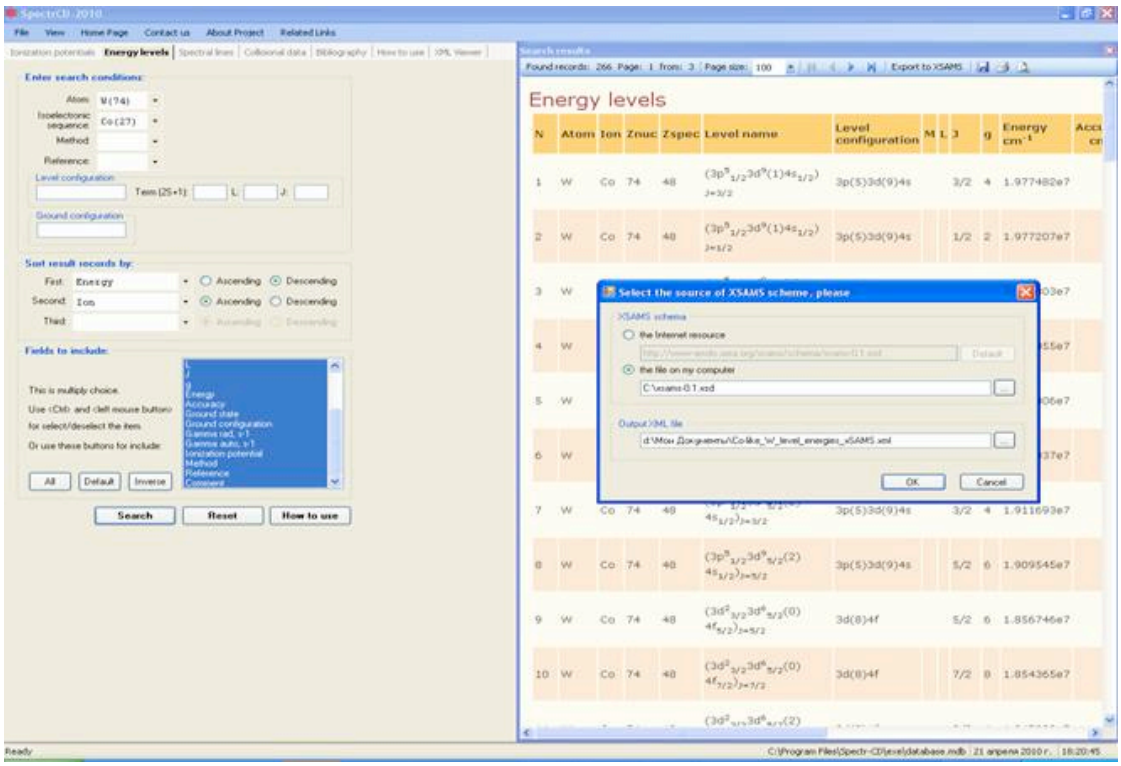


Figure 2. Setting the export procedure of the selected Spectr-W^3 energy-level data for Co-like W to the xSAMS-gauged XML file.

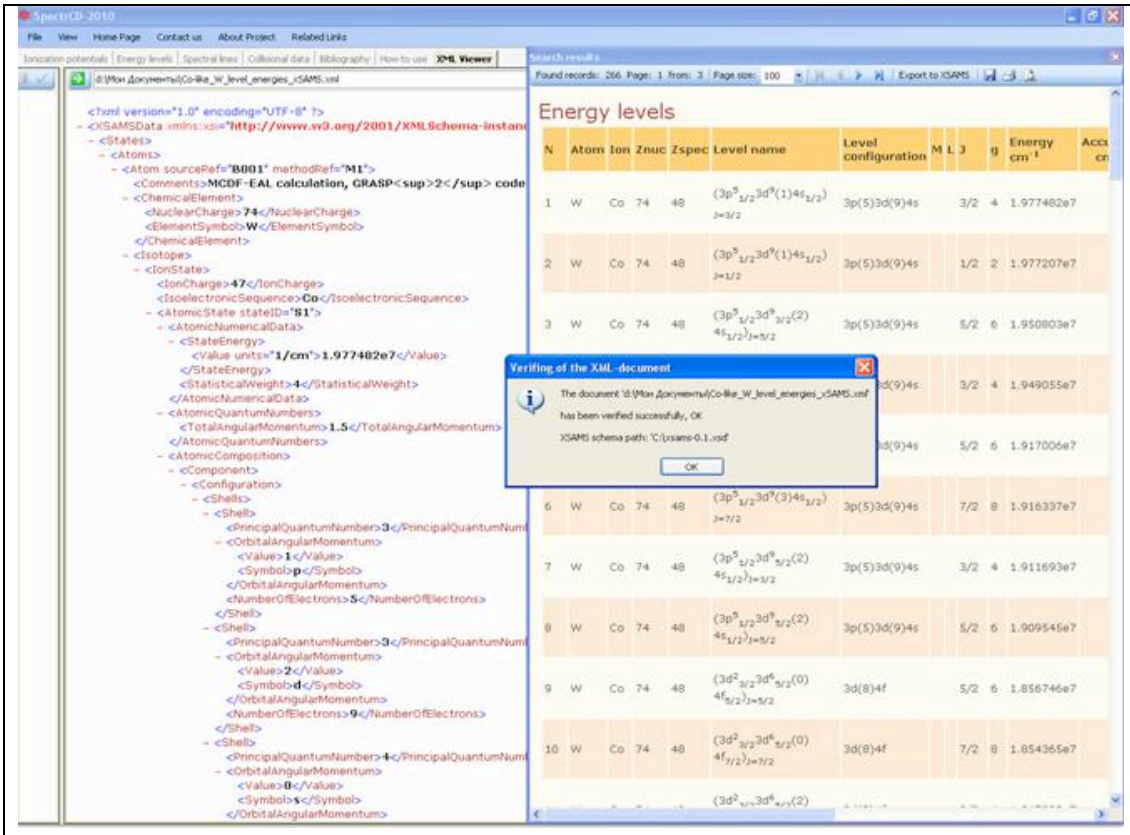


Figure 3. Result of the export procedure of the selected Spectr-W^3 energy-level data for Co-like W to the xSAMS-gauged XML file (left window).

Using the classes generated with the Altova XMLSpy package (Enterprise edition), an import capability of the prototype tool was created to convert atomic data from the xSAMS-gauged XML files to the internal format of the Spectr-W^3 database. In the process of preparing this prototype, a Content viewer enabling to easily inspect the contents of the input XML file was developed. Figure 4 shows an example of the contents viewing of the xSAMS-gauged file generated by Dr. Yuri Ralchenko from the ASD database at NIST for energy levels of Ne-like Si.

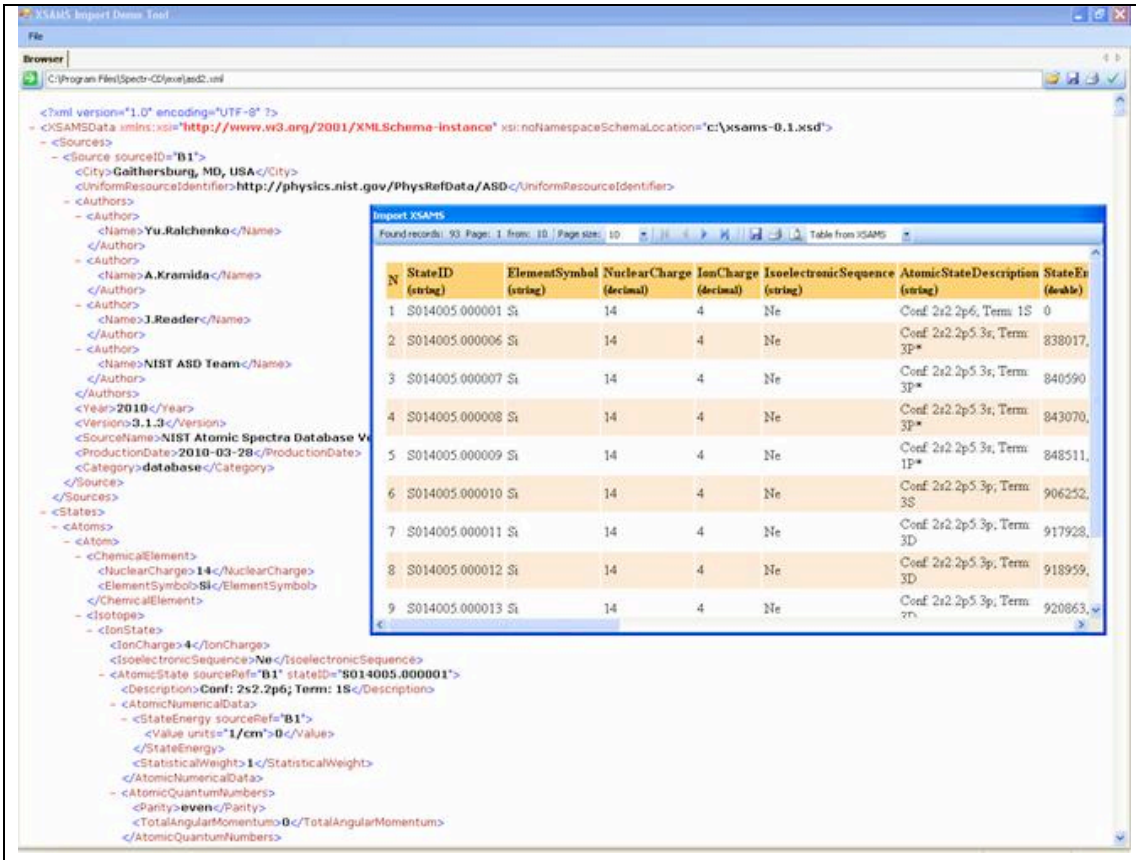


Figure 4. Content viewing of the xSAMS-gauged file on the energy-level data for Ne-like Si selected from the NIST ASD database.

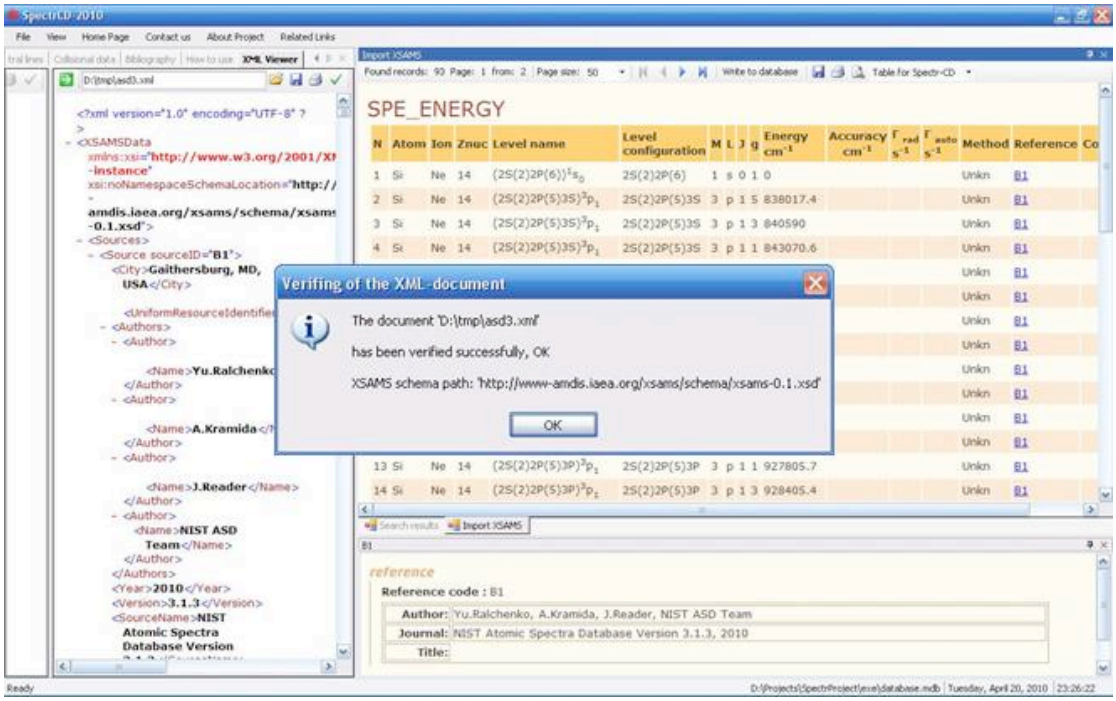


Figure 5. The result of import of the ASD-retrieved energy-level data for Ne-like Si to to the energy-level section of the Spectr-W^3 database.

With this file, the operation of the import capability is demonstrated in Figure 5. As a result, 93 new records and 1 bibliography reference source were successfully read

from the ASD xSAMS-gauged XML-file. From those, however, only 16 new records were tentatively added to the energy-level section of the Spectr-W^3 database due to duplications found by Spectr-W^3 in the key fields of the input-data file. The reason was evidently due to insufficiency of the "as is" ground for data import because of different formats for atomic-state descriptions inherent to the ASD and Spectr-W^3 databases. This data-import experiment highlighted the fact that detailed parsing of the atomic-state descriptions — especially atomic-configuration specifications including shell (pair) terms — should make a high-priority effort to ensure appropriate operation of the export/import data tool during the Cycle 2 activities under the VAMDC project.

Task 5: Automatic data verification tool

The publishing tools are designed for the solution of the problem formulated in [1]: The challenge of *knowledge publishing* or disseminating can be described as getting the right knowledge, in the right form, to the right person or system, at the right time … .
In Task 5 the efforts were focused on the solution of the task on acquiring the proper data which are interpreted as a part of a knowledge domain as well as the task on representing the acquired data to the users of the system.

The task on proper data acquiring consists of a series of subtasks. The key subtask is the one on the definition of primary data validity according to the criteria defined by the mathematical models of atoms and molecules, in particular, according to the selection rules and relations of reflexivity and transitivity.

The publication criterion is important for composite data sources check as well. According to this criterion all the data of a composite data source should be published. In order to carry out the set task during the first year of the project we made the following list of subtasks related to the task of published data acquiring and representation in tabular form.

- Preparing specification for the quantum numbers of CO-molecule, a "test" atom and the corresponding selection rules.
- Preparing the prototype of the upload data system for the molecular data (CO molecule transitions and line profiles data structures) from primary data sources.
- **Preparing the prototype of the upload data system for the atomic data (the atom transitions and line profiles data structures) from primary data sources.**
- Preparing the prototype of the tabular presentation system of molecular data (CO molecule transitions and line profiles data structures) from primary data sources.
- **Preparing the prototype of the tabular presentation system of atomic data (transitions and line profiles data structures) from primary data sources.**
- Formulating the requirements for publication tools (analysis of the relations between the data owner and person reviewing the data). Preparing specification of these relations and implementing them in a prototype of software for the publishing procedure.
- Formulating the requirements for the "standard" middleware based on the analysis of the atomic and molecular data and metadata structures and constructing first version of middleware.

The spectroscopy data domain restricted by four tasks was chosen for the solution of the task on the creation of data acquiring procedure. Among these tasks are direct and inverse tasks on atom or molecule spectral line profile parameters and transitions determination.

In the first year we narrowed our efforts to the implementation of selection rules check

during data upload and to the calculation of standard deviations on the example of a single molecule and atom. In order to complete this task we solved the following problems:

- Preparing specification for the quantum numbers of CO-molecule, a “test” atom and the corresponding selection rules. In the framework of this task we specified the quantum numbers for carbon oxide molecule and a “test” atom as well as the selection rules.
- Preparing the prototype of the upload data system for the molecular data (CO molecule transitions and line profiles data structures) from primary data sources.

The software implementation of an upload system for CO molecule composing of three stages was realized (see http://wadis.saga.iao.ru/co/). The interfaces for upload procedure are shown on Fig. 6 and 7.

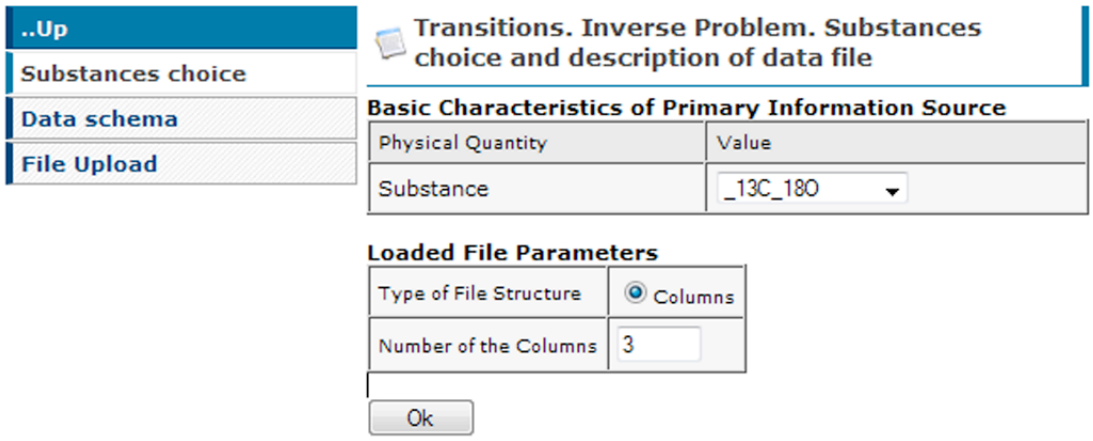


Figure 6. First step: Choice of molecule and typing the number of physical quantities.

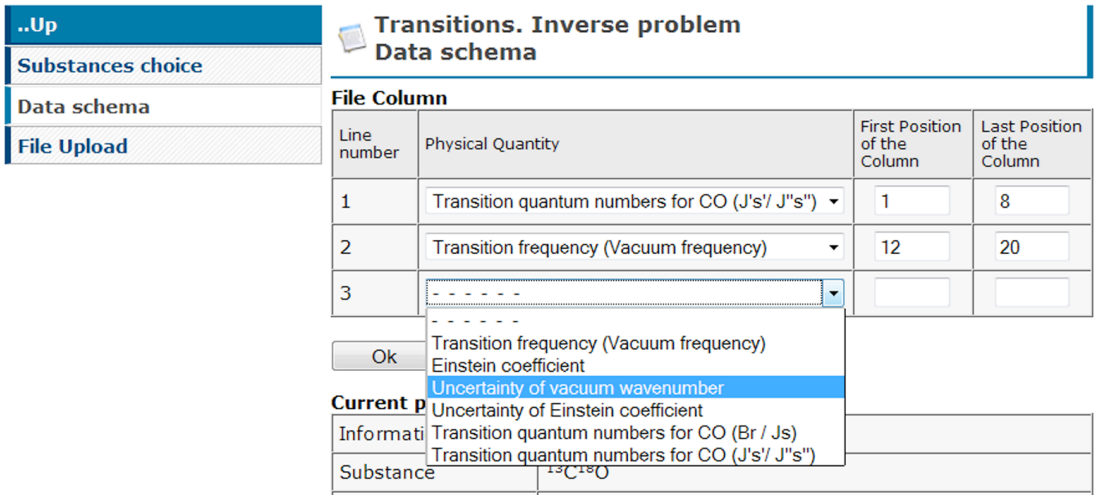
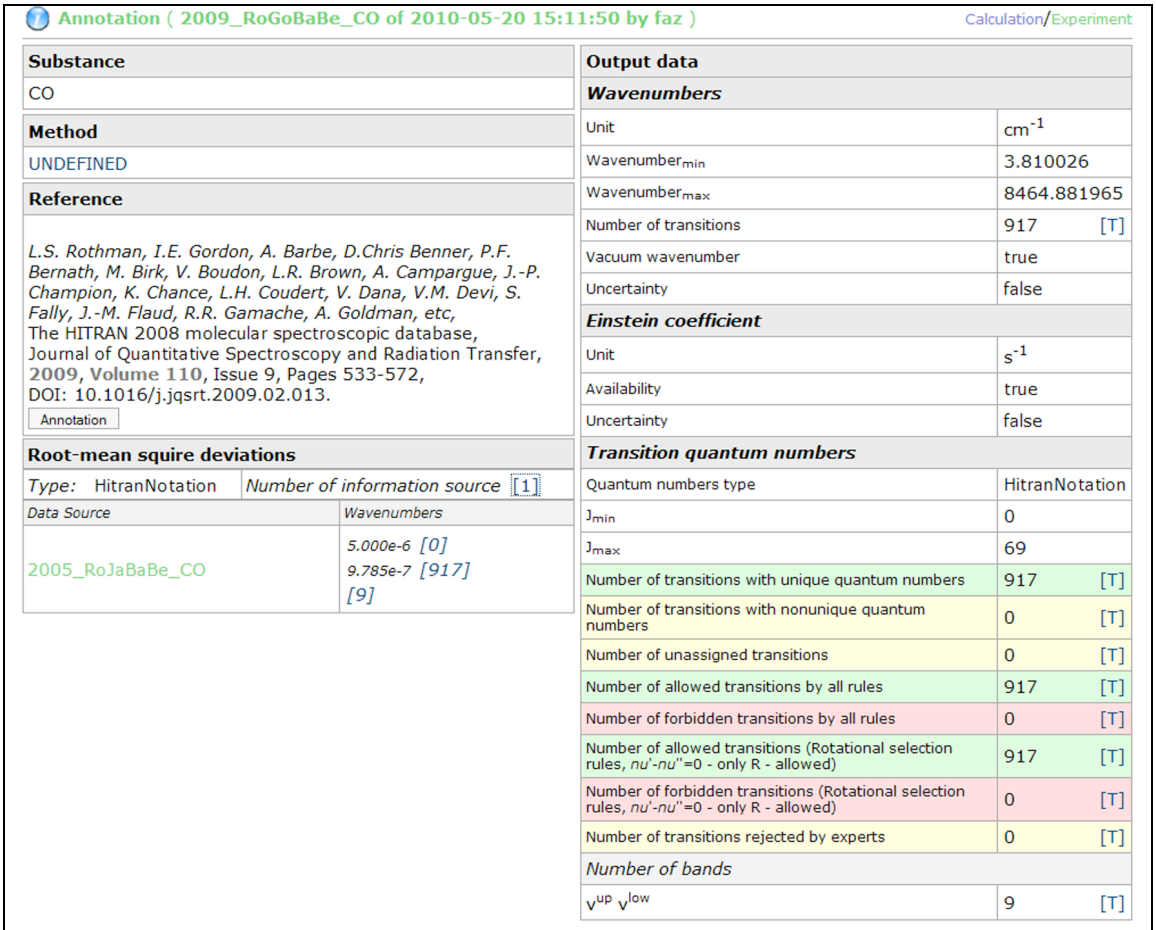


Figure 7. Second and third steps: Construction data schema and upload file.

The properties of the uploaded primary data sources are presented to the users in a form shown on Fig. 8.



Annotation (2009_RoGoBaBe_CO of 2010-05-20 15:11:50 by faz) Calculation/Experiment

Substance

CO

Method

UNDEFINED

Reference

L.S. Rothman, I.E. Gordon, A. Barbe, D.Chris Benner, P.F. Bernath, M. Birk, V. Boudon, L.R. Brown, A. Campargue, J.-P. Champion, K. Chance, L.H. Coudert, V. Dana, V.M. Devi, S. Fally, J.-M. Flaud, R.R. Gamache, A. Goldman, etc,
The HITRAN 2008 molecular spectroscopic database,
Journal of Quantitative Spectroscopy and Radiation Transfer,
2009, **Volume 110**, Issue 9, Pages 533-572,
DOI: 10.1016/j.jqsrt.2009.02.013.
Annotation

Root-mean squire deviations

Type: HitranNotation | *Number of information source* [1]

Data Source	*Wavenumbers*
2005_RoJaBaBe_CO	*5.000e-6 [0]* *9.785e-7 [917]* *[9]*

Output data

Wavenumbers		
Unit	cm^{-1}	
$Wavenumber_{min}$	3.810026	
$Wavenumber_{max}$	8464.881965	
Number of transitions	917	[T]
Vacuum wavenumber	true	
Uncertainty	false	
Einstein coefficient		
Unit	s^{-1}	
Availability	true	
Uncertainty	false	
Transition quantum numbers		
Quantum numbers type	HitranNotation	
J_{min}	0	
J_{max}	69	
Number of transitions with unique quantum numbers	917	[T]
Number of transitions with nonunique quantum numbers	0	[T]
Number of unassigned transitions	0	[T]
Number of allowed transitions by all rules	917	[T]
Number of forbidden transitions by all rules	0	[T]
Number of allowed transitions (Rotational selection rules, *nu'-nu''*=0 - only R - allowed)	917	[T]
Number of forbidden transitions (Rotational selection rules, *nu'-nu''*=0 - only R - allowed)	0	[T]
Number of transitions rejected by experts	0	[T]
Number of bands		
$v^{up}\ v^{low}$	9	[T]

Figure 8. Metadata characterized the properties of the primary data source.

The data structures were created in DBMS MySQL. The PHP-scripts for data upload were created.

- Preparing the prototype of the upload data system for the atomic data (the atom transitions and line profiles data structures) from primary data sources. The scheme of the database for the storage of 2 types of atomic spectroscopy tasks' solutions (line profiles and transitions for an isolated atom) was developed.
- Preparing the prototype of the tabular presentation system of molecular data (CO molecule transitions and line profiles data structures) from primary data sources.

The software for tabular representation of molecular data was developed. Below the examples of data tabular representation are presented.

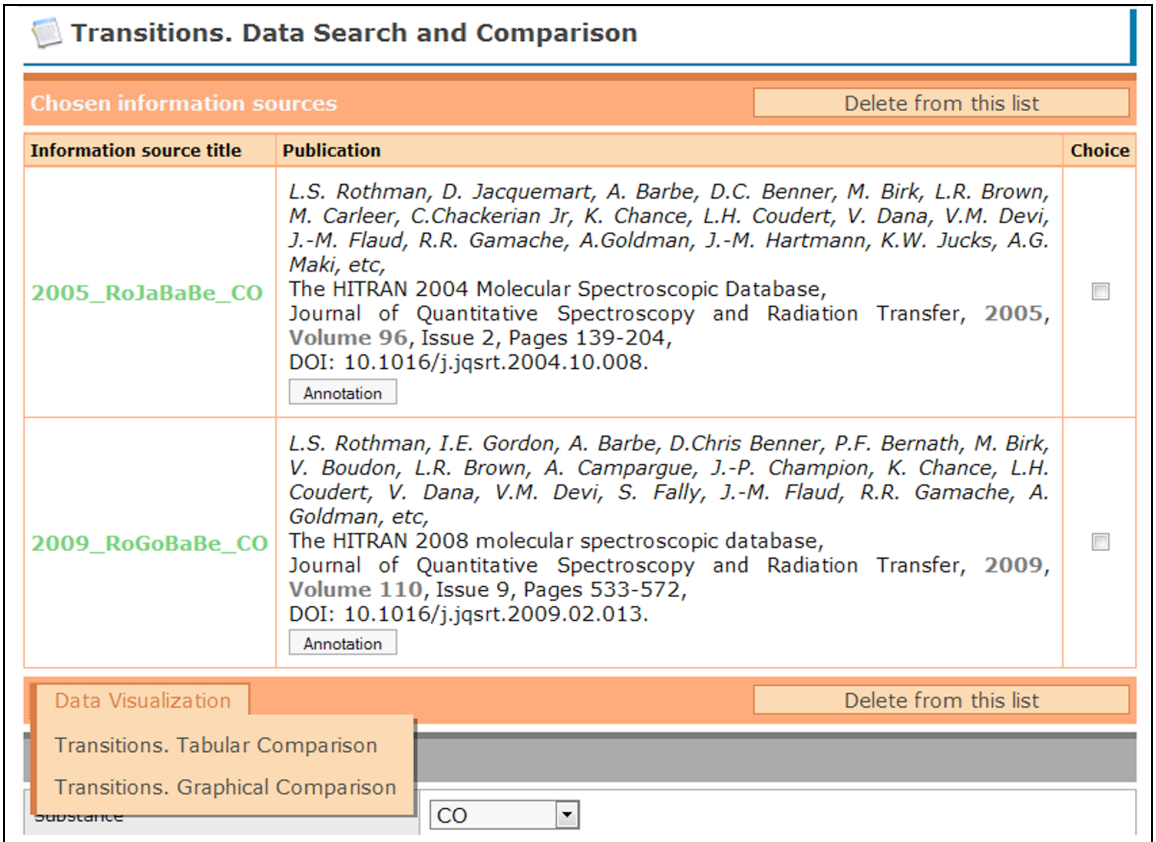


Figure 9. Step 2: Choice of data visualisation type

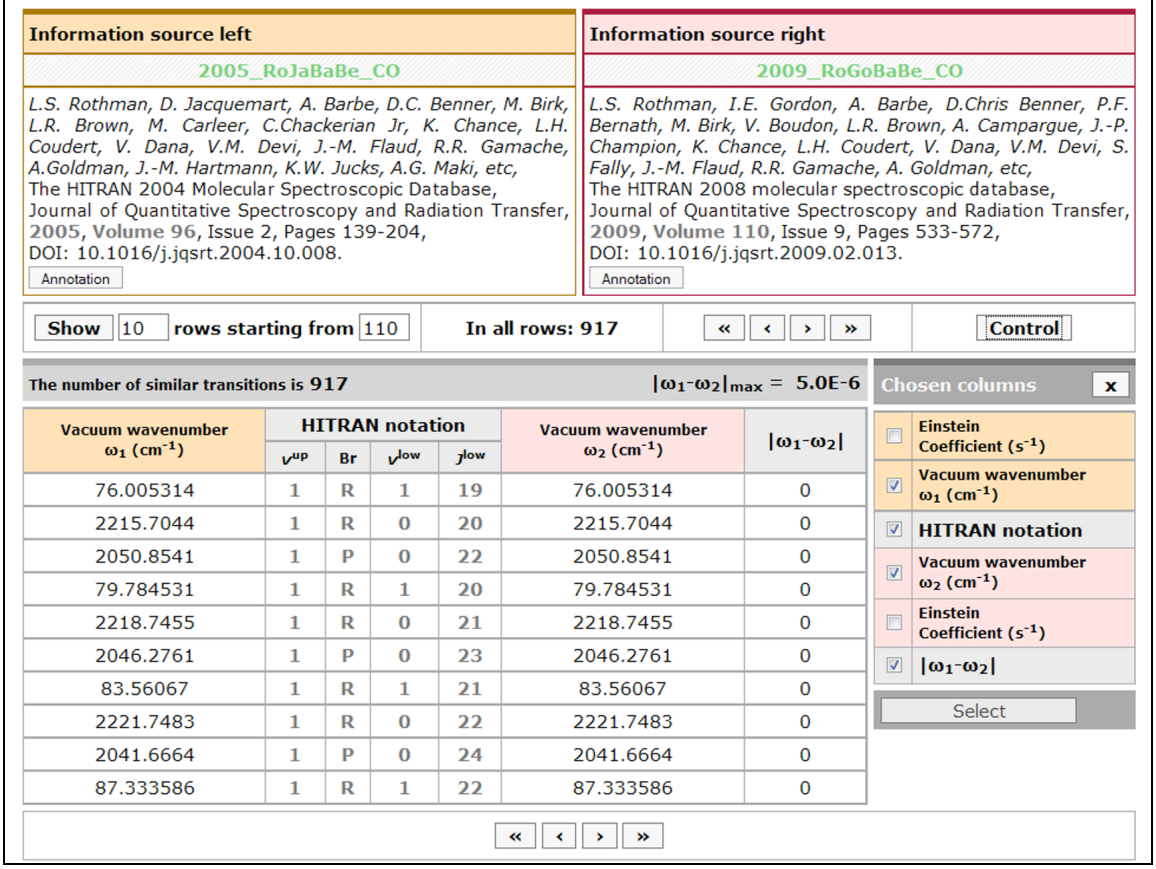
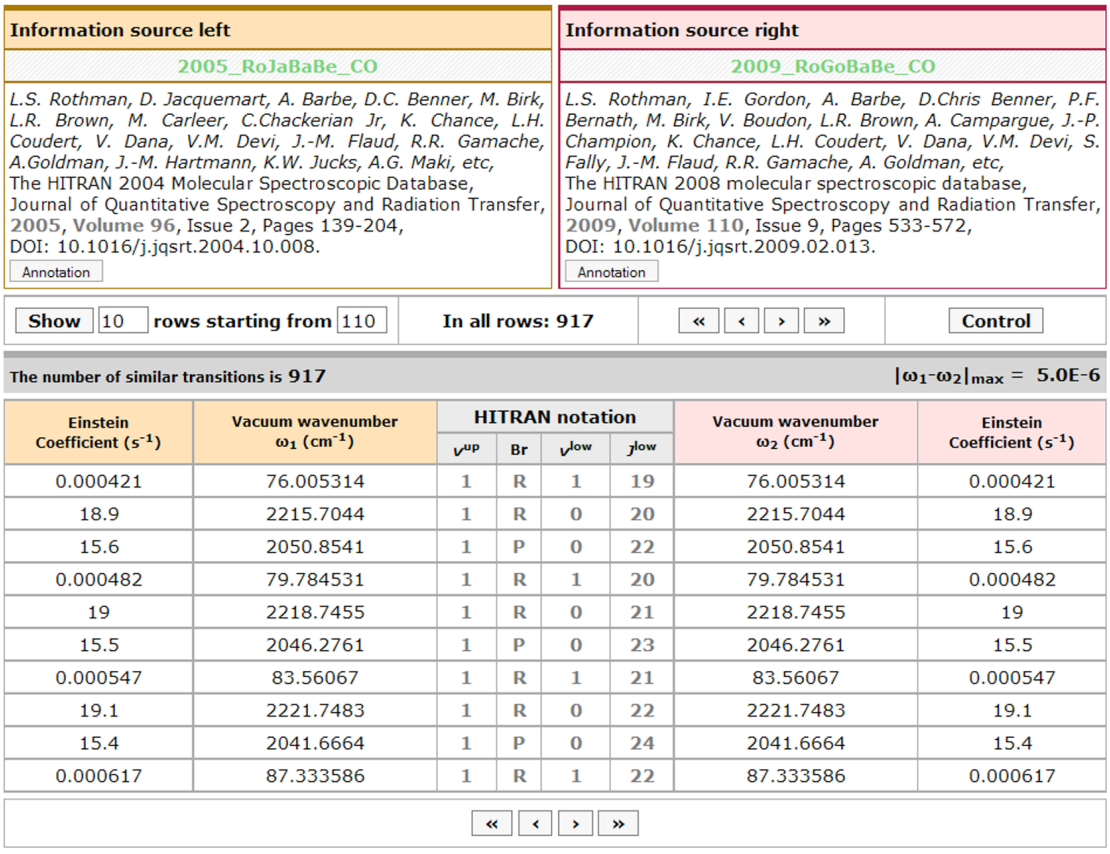


Figure 10. Comparison of Hitran versions (Hitran 2004 & 2008)



Information source left	Information source right
2005_RoJaBaBe_CO	2009_RoGoBaBe_CO
L.S. Rothman, D. Jacquemart, A. Barbe, D.C. Benner, M. Birk, L.R. Brown, M. Carleer, C.Chackerian Jr, K. Chance, L.H. Coudert, V. Dana, V.M. Devi, J.-M. Flaud, R.R. Gamache, A.Goldman, J.-M. Hartmann, K.W. Jucks, A.G. Maki, etc, The HITRAN 2004 Molecular Spectroscopic Database, Journal of Quantitative Spectroscopy and Radiation Transfer, **2005**, **Volume 96**, Issue 2, Pages 139-204, DOI: 10.1016/j.jqsrt.2004.10.008.	*L.S. Rothman, I.E. Gordon, A. Barbe, D.Chris Benner, P.F. Bernath, M. Birk, V. Boudon, L.R. Brown, A. Campargue, J.-P. Champion, K. Chance, L.H. Coudert, V. Dana, V.M. Devi, S. Fally, J.-M. Flaud, R.R. Gamache, A. Goldman, etc,* The HITRAN 2008 molecular spectroscopic database, Journal of Quantitative Spectroscopy and Radiation Transfer, **2009**, **Volume 110**, Issue 9, Pages 533-572, DOI: 10.1016/j.jqsrt.2009.02.013.
Annotation	Annotation

Show 10 rows starting from 110 | In all rows: 917 | « ‹ › » | Control

The number of similar transitions is 917 | $|\omega_1-\omega_2|_{max}$ = 5.0E-6

Einstein Coefficient (s^{-1})	Vacuum wavenumber ω_1 (cm^{-1})	HITRAN notation				Vacuum wavenumber ω_2 (cm^{-1})	Einstein Coefficient (s^{-1})
		v^{up}	Br	v^{low}	j^{low}		
0.000421	76.005314	1	R	1	19	76.005314	0.000421
18.9	2215.7044	1	R	0	20	2215.7044	18.9
15.6	2050.8541	1	P	0	22	2050.8541	15.6
0.000482	79.784531	1	R	1	20	79.784531	0.000482
19	2218.7455	1	R	0	21	2218.7455	19
15.5	2046.2761	1	P	0	23	2046.2761	15.5
0.000547	83.56067	1	R	1	21	83.56067	0.000547
19.1	2221.7483	1	R	0	22	2221.7483	19.1
15.4	2041.6664	1	P	0	24	2041.6664	15.4
0.000617	87.333586	1	R	1	22	87.333586	0.000617

« ‹ › »

Figure 11. Interface for control over the representation of columns.

- Preparing the prototype of the tabular presentation system of atomic data (transitions and line profiles data structures) from primary data sources. Tabular representation of atomic data is implemented with the use of the technology described in the previous subtask.
- Formulating the requirements for publishing tools (analysis of the relations between the data owner and person reviewing the data). Preparing specification of these relations and implementing them in a prototype of software for the publishing procedure.

The principal scheme of persons' actions (user, editor and reviewers) under data source publication is shown below.

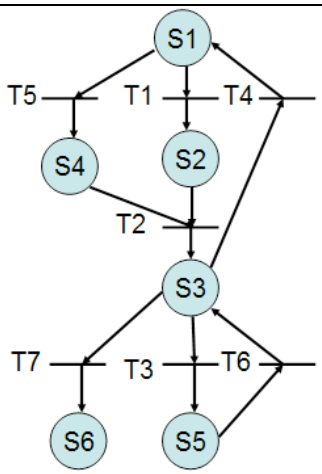


Fig.12. General Petri net for description of publishing procedure

States

S1 – «User (person)»

S2 - «User has new data source in information system (IS)»

S3 – «Editor (Person)»

S4 - «User has modified data source in IS»

S5 – «Reviewer (Person)»

S6 – «Published data source»

Transitions (Actions)

T1 – creation of a new user's data source in the IS

T2 - send address to Editor

T3 – send data to Reviewers

T4 – send letter to user (your data are declined or your data have to be modified)

T 5 – upload modified data source

T6 – send review to Editor

T 7 – data source publication

Restrictions (execution environment)

Transition T1 – data are not uploaded in the IS
Transition T5 - data are uploaded in the IS

Transition T4 – Editor got data source or reviews

Transition T7 – data source has to be published

The software and interfaces which provide the interrelations between user, editor and reviewers in the information system are implemented. This software is an integrated part of the middleware.

- Formulating the requirements for the "standard" middleware based on the analysis of the atomic and molecular data and metadata structures and constructing the first version of middleware.

The first version of middleware for the formation of molecular and atomic data upload system was created. The middleware includes the facilities for the development of interfaces providing multilingual interfaces and representation of the interfaces in many forms as well as supporting a reference system, a glossary and the means of applications integration into the information system.

Developed software use the typical freware environment related to web information system (Linux, Apache, MySQL and PHP).

1. *De Roure D ., Jennings N ., Shadbolt N* . A Future *e* -Science Infrastructure // Report commissioned for EPSRC/DTI Core e-Science Programme. 2001. 78 p.

Deviations from the contract (Annex I) and reasons for them (if applicable)

NOT APPLICABLE

Failures to achieve critical objectives and/or not being on schedule and reasons for them (if applicable)

WP7 Plan for Period 1 has not followed exactly Annexe I separation of Tasks. It has tried to re-define what would be useful to produce during Period 1 considering that WP7 started from scratch. No deviation from P1 WP7 Plan is observed apart from Task 2 which is postponed to Period 2. For Period 1, registry is manually filled with resource records for databases, legacy access services, DSA/Catalog TAP service installations.

Task 2 was not on schedule because it requires more work around specifications for registries and a more stable version of XSAMS -

Proposed corrective actions (if applicable)

Task 2 is postponed to Period 2 –
Period 1 WP7 has been extremely creative with publishing interfaces and allowed to better see what needs to be provided as generic tools for data producers. Plans for Period 2 come back to Annexe I description of tasks -

(approximate length of Part 1: 2 pages)