ADAS-EU R(13)OP03

ADAS-EU
ADAS for fusion in Europe

Grant: 224607

Hugh Summers

OPEN-ADAS report 3

December 2013

Workpackages : 26-2-3
Category : PU

OPEN-ADAS report 3

Hugh Summers

Department of Physics, University of Strathclyde, Glasgow, UK

Abstract: *The report reviews OPEN-ADAS activities for project months 37-48*

Contents

1 Overview **3**

1.1 OPEN- ADAS user interface 5

1.2 IAEA database accesss page 8

1.3 Work package reports 9

1.4 Work package 26-2-3 9

A OPEN-ADAS sub-contract specification **10**

B OPEN-ADAS sub-contract final report **11**

Preface

The report is third of an originally planned series of four such reports, deliverable under the ADAS-EU project. This report spans the period from month 37 to the end of the project and subsumes report 4.

M G O'Mullane
December 2013

Chapter 1

Overview

OPEN-ADAS is the name of the path for release of ADAS data and support software into the public domain. It is an agreed and shared project between ADAS and the International Atomic Energy Agency, Atomic and Molecular Data Unit, Nuclear Data Section (IAEA) in Vienna. As agreed in the ADAS-EU proposal, OPEN-ADAS was to be used for public domain release of fundamental and derived atomic data which enters the ADAS databases from ADAS-EU activities. The web server is located at the Physics Department, University of Strathclyde and is linked via the IAEA web pages.

The first deployment of OPEN-ADAS was based on v3.0 of ADAS and made available a subset of the ADAS data in a number of ADAS classes. ADAS Data Format (adf) files for fusion relevant modelling and diagnostic interpretation in the following classes were provided:

- adf01 : Charge exchange cross sections
- adf04 : Resolved specific ion data collections
- adf07 : Electron impact ionisation rates
- adf08 : Radiative recombination rates
- adf09 : Dielectronic Recombination coefficients
- adf11 : Iso-nuclear source and power coefficients
- adf12 : Charge exchange effective emissivity coefficients
- adf13 : Ionisation per photon coefficients
- adf15 : Photon emissivity coefficients
- adf21 : Beam stopping coefficients
- adf22 : Beam emission coefficients.

The first 5 classes are fundamental atomic data and the remainder are effective coefficients, mediated via a population model to be used in data analysis of finite density, collisional plasmas. Approximately 4000 datasets, or 1.6Gb, made up the original OPEN-ADAS database.

Technically OPEN-ADAS was a php presentation layer on top of an MySQL databsse. The ADAS datasets were queried to extract meta information and this was used for the searching. Free-form and data class specific search was provided. The final delivery was a complete dataset based on the philosophy of the ADAS system, where a complete set of data, useful for modelling is the starting point. This does demand that the user may need to undertake a post processing step in order to extract a particular piece of data, eg an ionisation rate or a transition probability, since these are delivered as an ensemble of data. It should be noted that computer code for reading each data class is also provided

as a download. Given our resources, and practical experience in using atomic data for modelling, this is a reasonable balance of making data available and not diverting time into marginally useful work.

As discussed in report OPEN2, the original system had a registration requirement. Unfortunately the system was vulnerable to an SQL injection attack and was duly compromised. Although knowledge of the users profiles is useful, it is not strictly necessary since funding is not driven by the usual web metrics of unique users, volume of downloads etc. The cleanest solution was to remove any registration requirement and simply allow free download of the data. The restrictions prohibiting systematic data downloading, on re-hosting OPEN-ADAS sourced data and acknowledging the source of the data remain. OPEN-ADAS was taken down for a period of 6 weeks while the registration part was removed and the web code hardened against SQL injection attacks.

As part of the activity ADAS-EU Support Action, further development of OPEN-ADAS was intended, but loss of a key ADAS staff member (Dr. Allan Whiteford) to the commercial sector delayed this activity. Under Amendment no. 1 to the ADAS-EU specification annex, funds were reassigned for the issue of an extra sub-contract in 2012 for refresh of the OPEN-ADAS website. The contract was placed with the professional web design company (director: Dr. Allan Whiteford) for the re-design of the user experience and to further harden the underlying web code. The new system was deployed in February 2013.

The distinction between fundamental and derived data was made clearer, the former being of more interest to the atomic data producers, the latter specifically designed for use by the plasma modelling and diagnostics community. New data classes were also added, specifically a more elaborated state resolved radiative recombination rate coefficients, photo-excitation and photo-ionisation rate date (adf48, adf38 and adf39). These are of particular interest to astrophysical models and ADAS has always reflected the substantial overlap of interests between the fusion and astrophysical data needs. The data in the existing OPEN-ADAS data classes was brought up to date with v4.0 of ADAS. This included substantial new sets of calculations of excitation data in collision strength and effective collision strength forms. These are calculated in the same distorted wave model. The collision strengths will be used, within ADAS, to advance the population modelling of impurities in plasmas with non-Maxwell electron distributions. Derived coefficients with specified non-Maxwellian EEDFs may be added to OPEN-ADAS but the usability of such data must be studied before a release. The current OPEN-ADAS contains nearly 14000 files and is now 16.5Gb of data. Some highlights of the new data include:

- A generalized collisional-radiative set of data for silicon.
- A set of adf11 data for tungsten assembled for Asdex Upgrade.
- R-matrix collision data for He, Li, F, Na and Ne-like iso-electronic sequences for all elements up to zinc. Note that neutral and single ionised members are not included since these do not exhibit regular sequence-like behaviour.
- Distorted wave adf04 data for all elements up to zinc, with type 1 (collision strength/cross section) and type 3 (effective collision strength / rates) files.
- Iso-nuclear collections of R-matrix adf04 data for argon and neon.
- Revised beam emission coefficients for hydrogen. Note that there is a very modest change in the beam stopping data.
- The adf38, adf39 and adf48 data contains bare ion to Mg-like sequences for hydrogen to zinc and xenon.

Sources and references are in the datasets and a summary can be found in the ADAS release bulletin, available from the ADAS website, *http://www.adas.ac.uk/bulletins.php*.

1.1 OPEN- ADAS user interface

The redesigned internet user interface to OPEN-ADAS is illustrated in some screenshots below. The web address is *http:open.adas.ac.uk*.

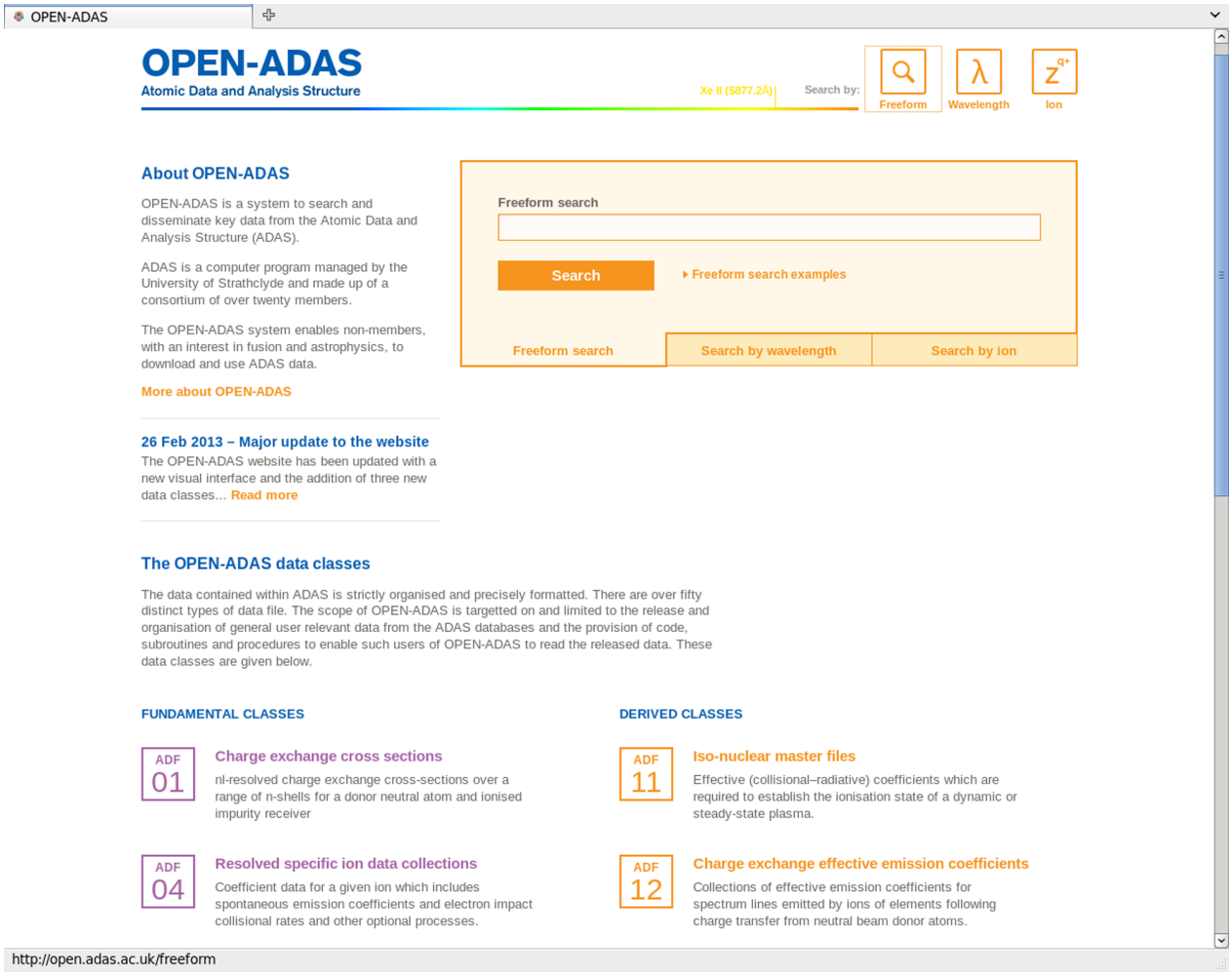

Figure 1.1: Home page of OPEN-ADAS showing freeform selection. Click on the appropriate icon at the top right for selection by *freeform*, *wavelength* or *ion*. Scroll down for the complete set of adf numbers available. Click on a class to show the available directories. The *mauve* shading is used for fundamental data classes (in the collisional-radiative sense and *orange* for the derived data classes.

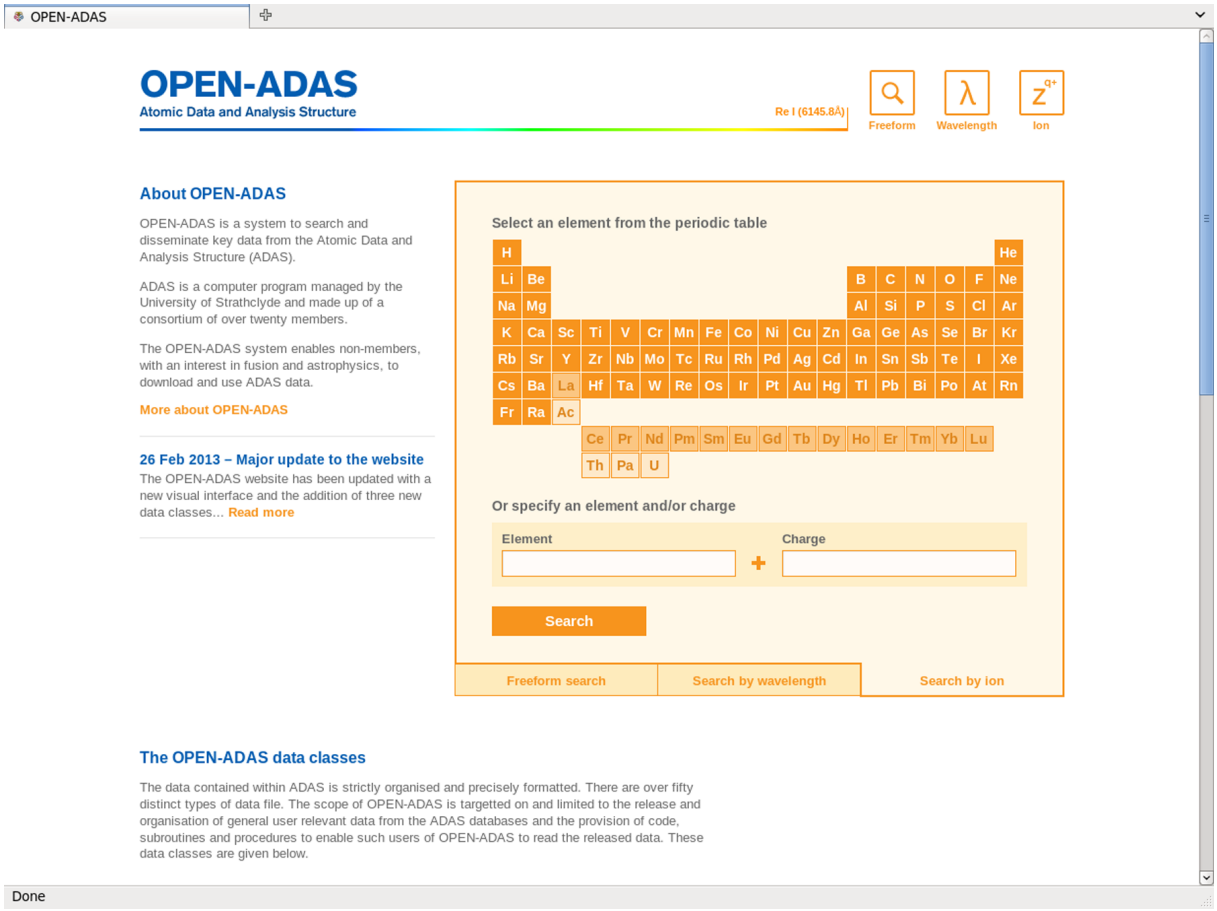

Figure 1.2: Home page of OPEN-ADAS showing selection by ion from the periodic table. Click on the appropriate icon at the top right for selection by *freeform*, *wavelength* or *ion*. Scroll down for the complete set of adf numbers available. Click on a class to show the available directories. The *mauve* shading is used for fundamental data classes (in the collisional-radiative sense and *orange* for the derived data classes.

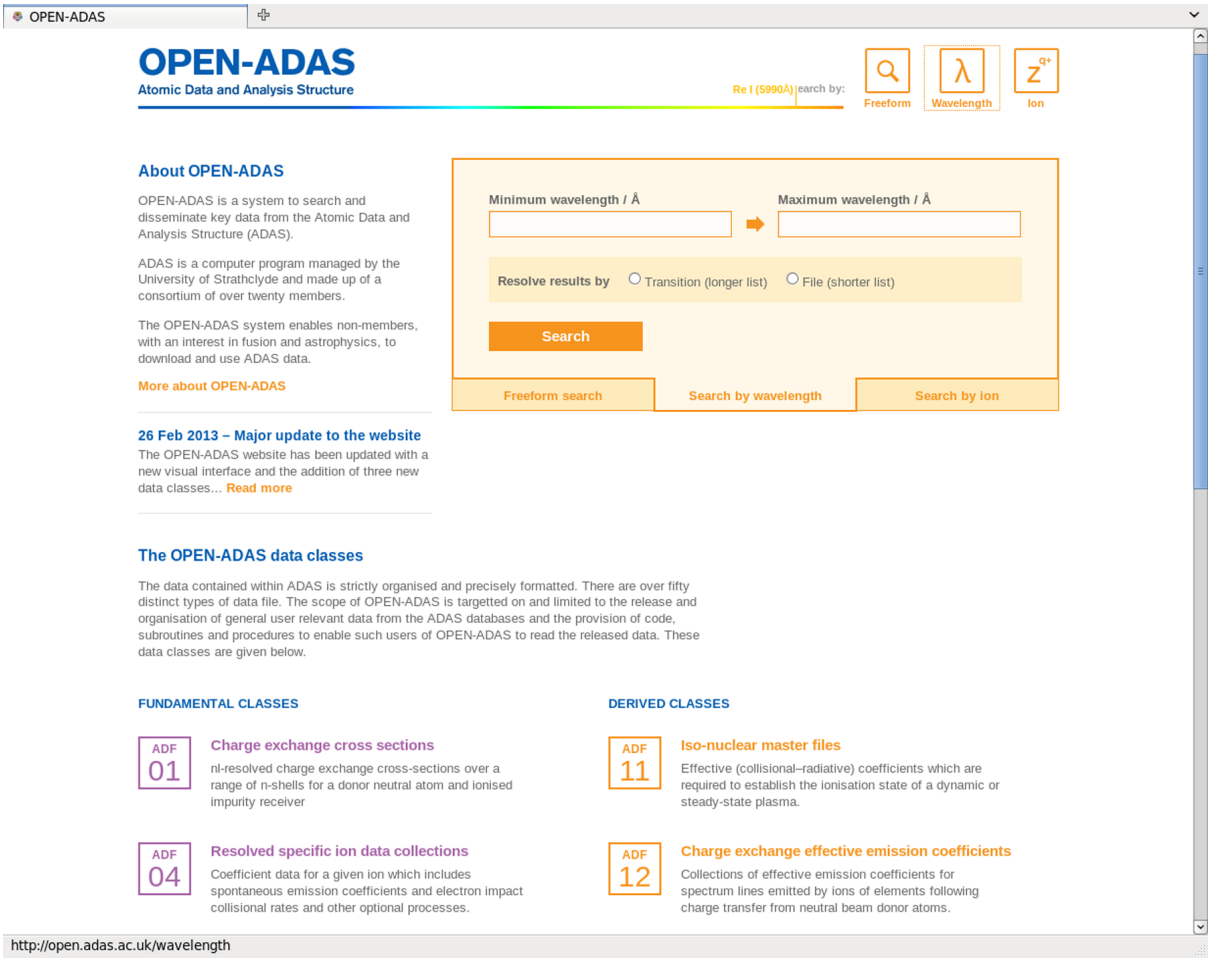

Figure 1.3: Home page of OPEN-ADAS showing selection by wavelength. Click on the appropriate icon at the top right for selection by *freeform*, *wavelength* or *ion*. Scroll down for the complete set of adf numbers available. Click on a class to seow the available directories. The *mauve* showing is used for fundamental data classes (in the collisional-radiative sense and *orange* for the derived data classes.

OPEN-ADAS allows download of datasets and utility subroutines/procedures in different languages (FOTRAN, IDL, PYTHON) . The bottom of the home windows, as shown in figure 1.4, gives access to these extended capabilities. The new version of OPEN-ADAS does not maintain statistics of usage, instead relying on IAEA AMDIS to maintain these (see section1.2).

Figure 1.4:

1.2 IAEA database accesss page

As explained in the introduction to chapter 1, the OPEN-ADAS development was originally jointly funded by the International Atomic Energy Agency, Atomic and Molecular Data Unit and the ADAS Project. The Atomic and Molecular Data Unit Information System, AMDIS, (see *https://www-amdis.iaea.org/*) supports a link to OPEN-ADAS. The AMDIS home page is shown in figure 1.5. Unfortunately, The IAEA AMDIS website has itself been subjected to and brought down by a Denial of Service attack connected to MYSQL vulnerabilities. The site was offline for nine months in 2013. Also statistics up to March 2013, the date of attack were corrupted and not recoverable.

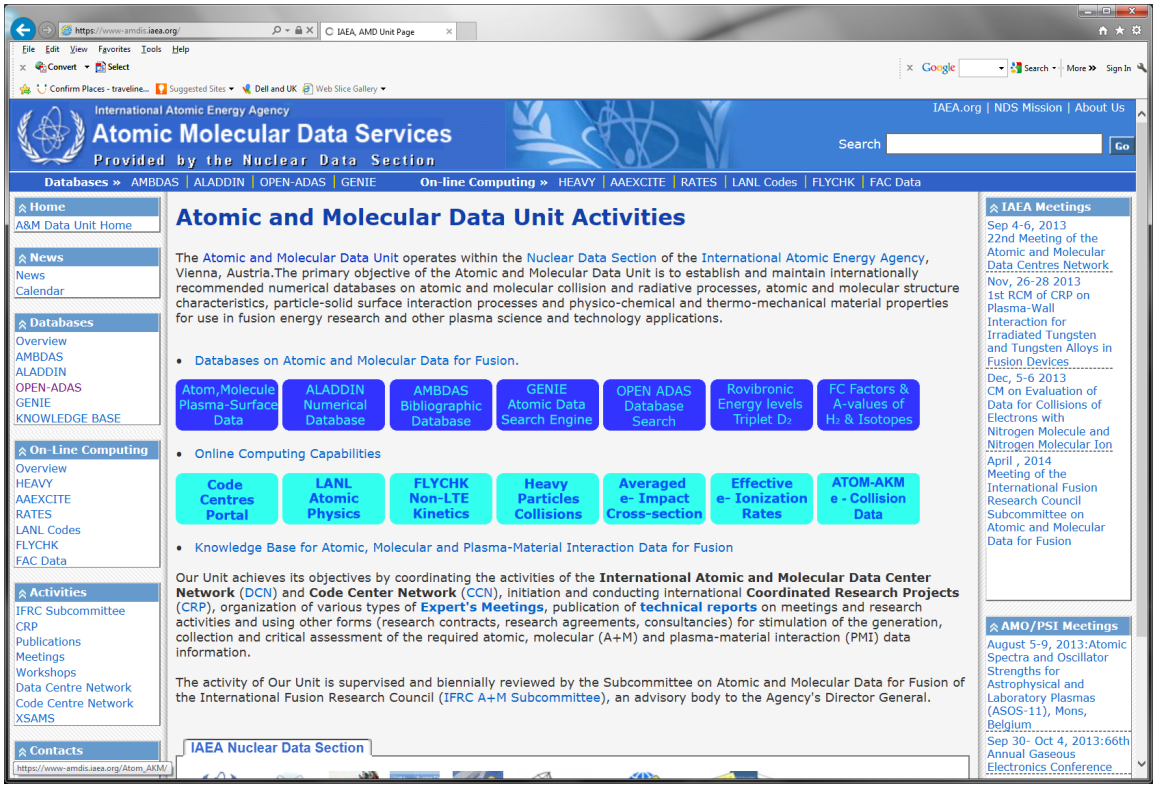

Figure 1.5: Shows the IAEA link to ADAS data

1.3 Work package reports

1.3.1 Work package 26-2-3

The work package task comprises the preparation of this report.

Appendix A

OPEN-ADAS sub-contract specification

1. Make the OPEN-ADAS site as secure as possible in line with current best practice. Review the mysql and php settings and eliminate the errors that are seen in the logs, even if they are only warnings due to deprecated features. In particular make sure that the mysql settings are secure with separate processes for different aspects. Also isolate the web servers into containers if deemed necessary.
2. Review the functional flow of website - at an early stage of the contract establish the look of the site in line with current expectations. The basic functionality is to remain, ie full datasets are returned and a lightweight, mostly pure HTML web page delivery is retained but enhanced CSS styles could be used to improve the user experience. The addition of new adf data classes may necessitate a more hierarchical webpage structure.
3. Document and streamline the back-end database generation. The current system is a series of perl scripts from which the metadata for the mysql is derived.
4. Add new ADAS data classes - adf00, adf38, adf39, adf48.
5. Add a news and privacy section and ensure the site's use of cookies is compatible with current legislation.
6. Make funding contributors, particularly FP7 and IAEA more prominent.
7. Attribution of effort in increasingly important if OPEN-ADAS is to be positioned as a primary publication vehicle. Improve the back-end extraction of this data and make recommendations for changes to the ADAS workflow and comment inclusion.
8. References such as 'data from ADAS', or the slightly better 'data downloaded from OPEN-ADAS July 2012' show that it is not sufficiently fine grained. The document object identifier system (doi) is a way to make a permanent link to an object, in this case an ADAS dataset. The doi system is not confined to academic publishing, although is has established an authoritative position there. Explore the cost, both initial and ongoing, for tagging each ADAS file with a doi. This will make referencing simple and acceptable to the journals. If this proves affordable add a doi resolver box to OPEN-ADAS. If the cost is prohibitive, on an ongoing basis, develop our own unique url identifier scheme.
9. Reinstate the statistics which was removed after the hacking incident. Anonymous provision of data removes effort in maintaining the system but the valuable knowledge of where the data was used was lost. Add a system to parse the access logs for unique access and reference to geo-location.
10. Oversight of the contract will be by phone and video conference but two formal meeting, one early on to approve the design aspects and a wrap-up meeting are planned.
11. Contract supervision and completion approval by Dr. Martin O'Mullane for ADAS-EU.
12. Duration 3 months - 1 Oct. 2012 to 31 Dec. 2012.
13. Deliverables are (a) Working OPEN-ADAS with updated code (b) Summary report on update and guide for users.
14. Cost £10,000

Appendix B

OPEN-ADAS sub-contract final report

OPEN-ADAS

Upgrade of OPEN-ADAS Online Infrastructure

Final Report

Last updated: February 21, 2013

Executive Summary

The present document details changes made to the OPEN-ADAS system as part of a sub-contract awarded to 256 Kelvin Limited from the ADAS-EU project.

The primary purpose of the sub-contract was to enhance and reinforce the online distribution mechanism for OPEN-ADAS. A significant security review was performed and the re-write of front-end infrastructure gave opportunity to also address some of the visual aspects of the site.

Allan Whiteford
allan@256kelvin.com
07974 196 941

Andy Armstrong
andy@256kelvin.com
07449 309 989

Contents

1 Background **3**

2 Site security and Infrastructure **4**
2.1 SQL Injection Attacks 4
2.2 Removal of user database 5
2.3 URL Structure 5

3 Visual design and user experience **6**
3.1 Design aesthetics 6
3.2 Navigation scheme 8
3.3 Home page 8

4 Expansion of data coverage **10**
4.1 ADF38 10
4.2 ADF39 10
4.3 ADF48 10
4.4 DOI 11

5 Back end data processing and organisation **12**
5.1 File tagging 12
5.2 SQL population 12
5.3 Code organisation 12

6 Other considerations **12**
6.1 News 12
6.2 Privacy policy 12

1 Background

OPEN-ADAS is the name of the path for release of ADAS data and support software into the public domain. It is an agreed and shared project between ADAS and the International Atomic Energy Agency, Atomic and Molecular Data Unit, Nuclear Data Section (IAEA) in Vienna. As agreed in the ADAS-EU proposal, OPEN-ADAS is used for public domain release of fundamental and derived atomic data which enters the ADAS databases from ADAS-EU activities. The web server is located at the Physics Department, University of Strathclyde and is linked via the IAEA web pages. The development of OPEN-ADAS was substantively finished early in the ADAS-EU project with the addition of free-form search capability. During the first 18 month reporting period the number of users increased steadily each month and came from very many countries. The principal aims of the project had been realized and it had settled down to high-availability but low-maintenance web service.

In June 2010 email attacks began. The webserver logs showed a few SQL injection attacks every few weeks which is not unexpected on the public internet. However around the end of May 2011 the volume rose sharply. The reasons are unclear but news of a successful exploit probably circulates quickly in the anarchistic world of web breaking. Ameliorating steps were taken. The usual way of ameliorating such an attack is via a captcha. Before implementing such a system a simpler text challenge field was tried: requesting the correspondent to type 'not spam' into the appropriate box. This proved to be successful initially so the more invasive captcha system was not deployed. The OPEN-ADAS web site required a simple sign-up resulting in an account with a password. The name and email address were not verified before creating the account so a degree on anonymity was possible. This enabled us to produce accurate usage statistics since the number of hits is a very poor metric. Reports that a list of the OPEN-ADAS user names and passwords were available on the 'pastebin' web site came to us in the first week of June 2011. A serious security breach had occurred.

It should be noted that the actual attack vector was suspected to be another site hosted on the same server but that it is still deemed necessary at this point to do a thorough security audit of OPEN-ADAS.

It was decided that the most effective way to prevent a further occurrence was to remove the registration requirement. All data from OPEN-ADAS can now be downloaded anonymously. The removal of the code for registration, sanitizing the MySQL queries and trapping all input resulted in the OPEN-ADAS website being unavailable until 4 September 2011 — a two month interruption which also displaced other activities. The disadvantages in not knowing the OPEN-ADAS users are the loss of statistics and the ability to inform people who downloaded data of any subsequent improvements. Given the time pressures on the staff, this was an acceptable compromise.

2 Site security and Infrastructure

The public-facing website was given a thorough overhaul as a primary objective of this work. The underlying logic on searching for data has remained largely unaltered. Although it should be noted that there is a significant change in display (see section 3.1) and navigation (see section 3.2) along with enhanced data coverage (see section 4).

2.1 SQL Injection Attacks

A complete re-write of the way in which input data is processed was undertaken. This primarily involved rationalisation of data into distinct classes with well defined input and output characteristics. Whereas before each data class had it's own custom code, each class now has a file associated with it which transfers information about search parameters etc. to a more generalised searching mechanism.

This allows much less repetition of code and, critically for the security of the site, means that there orders of magnitude less places for security vulnerabilities to occur. Previously a file such as `adf15.php` had code like:

```
@$element   = $_GET['element'];
@$charge    = $_GET['charge'];
@$wave_min  = $_GET['wave_min'];
@$wave_max  = $_GET['wave_max'];
@$resolveby = $_GET['resolveby'];
```

it would have been necessary to modify **each** line to contain a call to `mysql_real_escape_string` and then repeat this exercise for **every** data class.

There is now a single place of code where search parameters come in which looks like:

```
$searchparams=$adf->searchparams();
foreach(array_keys($searchparams) as $p)
{
        if (isset($_REQUEST[$p]))
            $searchparams[$p]=mysql_real_escape_string($_REQUEST[$p]);
}
```

Such an approach makes it much less likely for an unescaped string to be able to be passed into SQL and also has the advantage of meaning there is less code so that future maintenance of the system will be much more straightforward.

2.2 Removal of user database

The original OPEN-ADAS system consisted of two MySQL databases:

- A read only database containing metadata about each file
- A read-write database containing user information

The need for login and capture of user information was removed some time ago but part of this upgrade fully cleansed any reliance on the user database existing at all.

2.3 URL Structure

The URL structure of the site was changed significantly, urls such as:

- `http://open.adas.ac.uk/adf08.php`
- `http://open.adas.ac.uk/filedetail.php?id=16300`
- `http://open.adas.ac.uk/aboutoa.php`

were changed to:

- `http://open.adas.ac.uk/adf08`
- `http://open.adas.ac.uk/detail/adf08/radrec/lilike/lilike_c3ls.dat`
- `http://open.adas.ac.uk/about-open-adas`

There is no inherent security advantage in this approach but it does obscure somewhat the technology being used and the fact there is a database with 'id's etc.

The two main advantages are that it is much better suited for search engine optimisation and also that links to files are persistent. Previously the file with an id of '16300' would change between versions of the ADAS database. By mapping the URL to a pathname similar to that of the underlying database URLs can be shared and stored which are more predictable. Note that because the character '#' has a special meaning in a URL it cannot be used in file URLs. Unfortunately, ADAS makes extensive use of the '#' character so in URLs we replace this with the string ']['.

3 Visual design and user experience

The overall user experience of the OPEN-ADAS site was upgraded significantly. Emphasis and attention was given to user flow both for the non-expert and expert user.

3.1 Design aesthetics

The design aesthetics of the site were completely redone. Figure 1 shows a screenshot from the new site. There are a number of specific issues to note versus the old design which are drawn attention to here.

Colour has been used with blue representing the site generally along with headings etc. Orange has been used for derived data and purple for fundamental data. Blue was chosen as the base colour to match current ADAS branding but the shade of blue was changed very slightly to be less imposing on the screen and easier on the eyes. Orange is then a natural companion to blue. There was an attempt made to make the purple shade match H_α emission but this was, unfortunately, not aesthetically appropriate.

It should also be noted that orange has been used for general searching (top right) even though these searches (with the exception of wavelength) search both fundamental and derived data. This is to subtly shift the emphasis of the site to derived data.

Note also the line at the top of the screen stretching from the blue logo to the orange buttons. The ‘spectrum’ band between the two colours corresponds to an inverse linear dispersion of 3.2 Å / px converted to an RGB representation using Bruton’s algorithm with a gamma value of 0.8. The orange and blue used in the design do not correspond exactly to a spectral colour (pure spectral colours do not work well in web design as they tend to be too bright and reduce contrast when against a white background). To deal with this problem and to prevent a sharp jump in colour there are ten pixels used at the short wavelength and five at the long wavelength to blend to the design colours .

As the mouse is moved by the user, a spectrum line moves along the top of the screen corresponding to a line at the wavelength of that pixel position. The movement is reasonably gentle and doesn’t cause much distraction but the implication is that horizontal spectroscopy is so trivial that it’s been relegated to a neat graphical effect so that the serious business of vertical spectroscopy can be done.

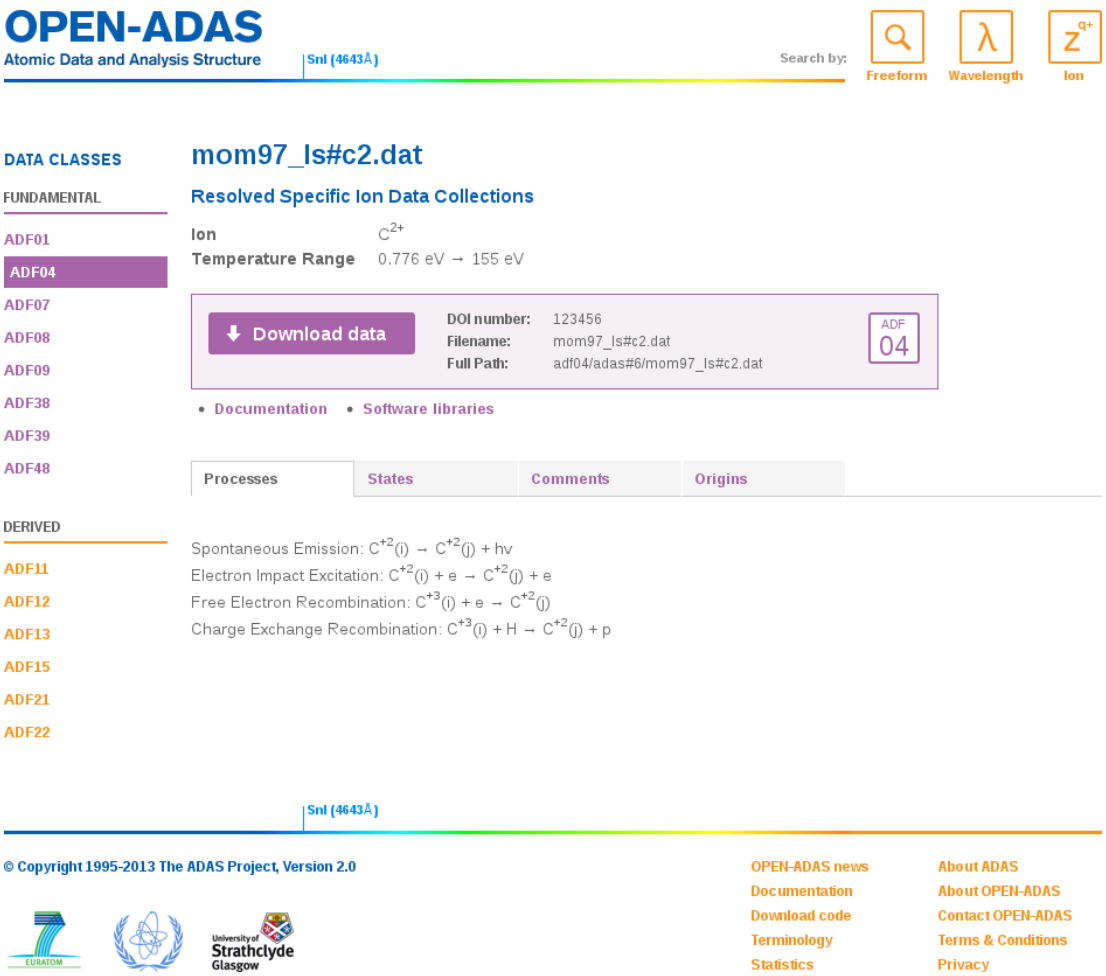

Figure 1: New design of OPEN-ADAS site

3.2 Navigation scheme

The navigation scheme was also re-done completely. In particular there are now three parts to the navigation:

1. General search boxes at the top right in a toolbox style arrangement.
2. All of the data classes are present down the left hand side of the screen.
3. Ancillary links have been moved to the footer.

A number of things have been achieved by these changes. Specifically the ADAS data classes now play a more front and centre role to the navigation, previously they were hidden behind a ‘search by dataclass’ link. The links themself have also become (intentionally) ambiguous. Rather than saying explicitly that they are for searching, the contents behind the link now gives some information about the data class and additionally allows one to search.

The fact the link is now simply the data class name allows unambiguous highlighting of it after a user navigates to it irrespective of the route they chose to get there. The previous version of OPEN-ADAS suffered in that if a user navigated to a file using freeform search then they would end up in a state where ‘search by dataclass’ was highlighted and the dataclass appearing on the tabbed navigation even though they had never explicitly chosen this route or option. The new navigation scheme, by design, eliminates these problems entirely.

The links at the left serve to remind the non-expert user about the extent and organisation of OPEN-ADAS and they also allow the expert user to quickly search for a specific data class.

3.3 Home page

The home page of the OPEN-ADAS system was given particular attention so that the non-expert user can quickly understand what the system is for, The previous home page was simply a freeform search box and some navigation. The new homepage is shown in figure 2.

This new home page frames the idea of OPEN-ADAS and has some positioning paragraphs for the new user. It also puts front and centre the three different ways of cross-class searching (freeform, wavelength, ion).

Finally, it explains each data class and what they are, encouraging the user to click on a data class of interest — this gives a new user an immediate overview of the coverage of OPEN-ADAS in terms of types of data.

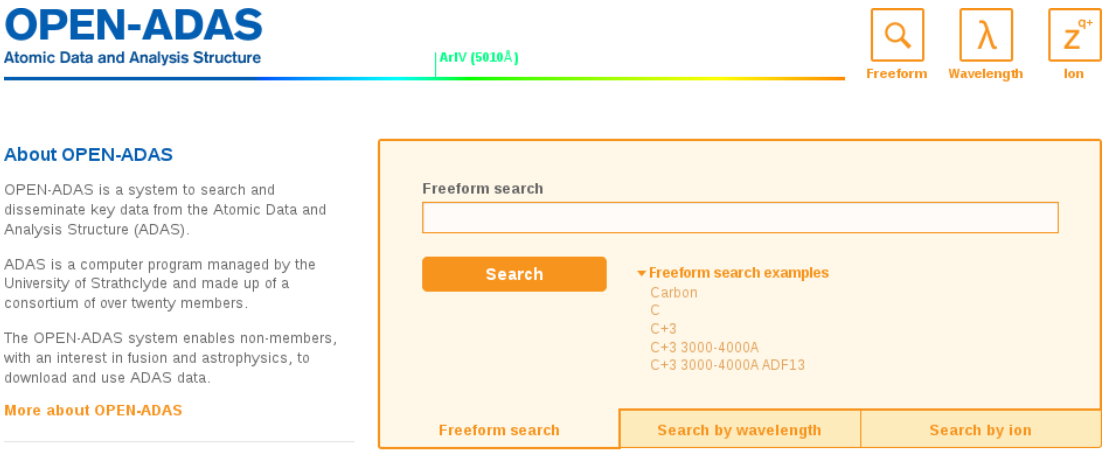

Figure 2: Home page of OPEN-ADAS site

4 Expansion of data coverage

As part of this project, a number of new data classes were added to OPEN-ADAS, these are detailed below. Most of these files are quite similar in structure (from the point of view of parsing the data) and searching parameters (from the point of view of the user searching for data) to ADF09 files.

4.1 ADF38

The basic process is

$$1\mathrm{s}^2 nl + \gamma \rightarrow 1\mathrm{s}nln'l' \rightarrow 1\mathrm{s}n''l'' + \mathrm{e} \tag{1}$$

A separate file is provided for each n , which contains data for all n'> n . Note however that

$$1\mathrm{s}nln'l' \leftrightarrow 1\mathrm{s}2n'l' + \gamma \tag{2}$$

has to be considered consistently at the same time to avoid multiple counting. Separate files are given for term (LS) coupling and intermediate (IC) coupling.

4.2 ADF39

The basic process is inner shell photo-ionisation:

$$(n_c l_c)^q \, nl + \gamma \rightarrow (n_c l_c)^{q-1} \, nl + \mathrm{e} \tag{3}$$

A separate file is provided for each n shell. Separate files are given for term (LS) coupling and intermediate (IC) coupling. Outer shell photo-ionisation is included for completeness. Where inner shell photo-ionisation is present, the atomic structure is optimised for the inner-shell processes.

4.3 ADF48

ADF48 files contain partial final-state resolved radiative recombination rate coefficients from both ground and metastable levels.

4.4 DOI

DOI numbers are issued by a number of different registries, the two which are most relevant for OPEN-ADAS are (descriptions taken from official DOI website):

- **CrossRef**
  - Scholarly and professional research content.
  - Journal articles, books, conference proceedings, etc.
  - Reference linking and searchable metadata database.
- **DataCite**
  - DataCite is focused on improving the scholarly infrastructure around datasets. There will be a set of activities around establishing and sharing best-practices, identifying and solving some of the unique issues that arise with datasets.
  - DataCite is focused on working with data centres and organisations that hold data. The details of their business models, workflows, and other requirements do not appear to be identical to those of publishers producing traditional journals.
  - DataCite has a business model that meets the needs of non-commercial and sometimes smaller organisations; larger national-scale organisations (e.g., TIB, BL) carry the basic infrastructure costs and will reclaim where appropriate within their domain.

It was decided, on balance, to firstly pursue the DateCite registry. An initial approach was made to the British library who are the UK body responsible for administering DateCite (typically participants do not approach DOI or even DateCite directly). Initial discussions were positive that DateCite via the British Library was an appropriate route to go down. These discussions are ongoing with the British Library.

5 Back end data processing and organisation

5.1 File tagging

A system was added whereby files could be excluded from OPEN-ADAS. This has initially been used to exclude older ADF38 and ADF39 files where levels and rates were split across separate files.

5.2 SQL population

The process of taking tag files and moving them into the relational (i.e. MySQL) database was increased by several orders of magnitude as part of this work. This was not a objective of the work but proved worthwhile to do in order to speed up testing of processing new files.

5.3 Code organisation

The code was reorganised and moved into a subversion repository. A detailed set of instructions for installing and updating OPEN-ADAS was also produced.

Most of the code reorganisation was done to the PHP front-end of OPEN-ADAS as necessitated to make it easier to add future data classes and also to deal with security concerns (see section 2 for more details).

6 Other considerations

6.1 News

Infrastructure for a news item was added to the site, this is an optional panel on the home page in addition to a link in the footer to see previous news items.

6.2 Privacy policy

A privacy policy page was added to the site, the specifics of this page are not finalised at the time of writing.