PLANETS NEWSLETTER

Issue 4, May 2008

IN BRIEF

By Dr. Adam Farquhar, Planets Project Director

June 2008 marks the half-way point in the project and has seen good progress towards a number of its end-goals. The team has now made first or even second releases of all the core Planets software components.

In April, Planets presented its preservation planning software tool, Plato, for the first time. The Plato tool will help users to systematically plan to preserve digital objects. Planets is now working to develop models which will make it possible for organisations to devise robust policies that will help to embed digital preservation into day-to-day activity.

The project is engaged in wide-ranging work to support digital preservation. A master-inventory of over 100 file formats found commonly in digital objects is now in place plus an inventory of the tools and actions that are available to preserve them. This is making it possible to identify gaps and create a blueprint of the tools that will be required. Recent work could result in Planets being able to process up to ten times more file formats than we had predicted a year ago.

ABOUT PLANETS

Planets is a four-year pan-European project funded by the European Commission under its Sixth Framework Programme (FP6).

The project was set up in June 2006 to help address challenges with preserving digital information in Europe.

From mid-2010, Planets will provide a common framework that will underpin digital preservation activity, plus downloadable software to help libraries, archives, universities and others preserve their digital content for the future.

The project brings together 16 national libraries and archives, research institutions and leading technology companies in Europe.

Alongside this, the Planets Interoperability Framework will allow third-parties to plug-in tools and services. The second version of the Interoperability Framework will be released in June 2008.

The latest version of the Planets Testbed was released in March, and presented for the first time in April. It will make it possible for users to experiment with strategies and tools to establish empirically the most appropriate for particular content. These developments and tools will be presented at our 2nd Annual Review with the European Commission on 14-15 July 2008.

We are keen to make certain Planets is not digital preservation's best preserved secret! In April, we attended CeBIT 2008 with Caspar, Digital Preservation Europe and Nestor, and were represented at 'What to Preserve? Significant Properties of Digital Objects' - a joint event organised by The British Library, the Digital Preservation Coalition and Joint Information Systems Committee. Also in April, we held our first outreach and training workshop: 'Planning the Future with Planets – a preservation planning tutorial'.

We hope you will find the newsletter informative and it will encourage you to seek out more information and take part in upcoming activities.

PLANETS FOCUS AREAS

Planets is organised in a number of focus areas:

- preservation planning,
- preservation characterisation,
- preservation action,
- interoperability framework,
- and testbed.

The activities in each area are supported by dissemination activities, including workshops, training events, publications and presentations. For more information see www.planets-project.eu

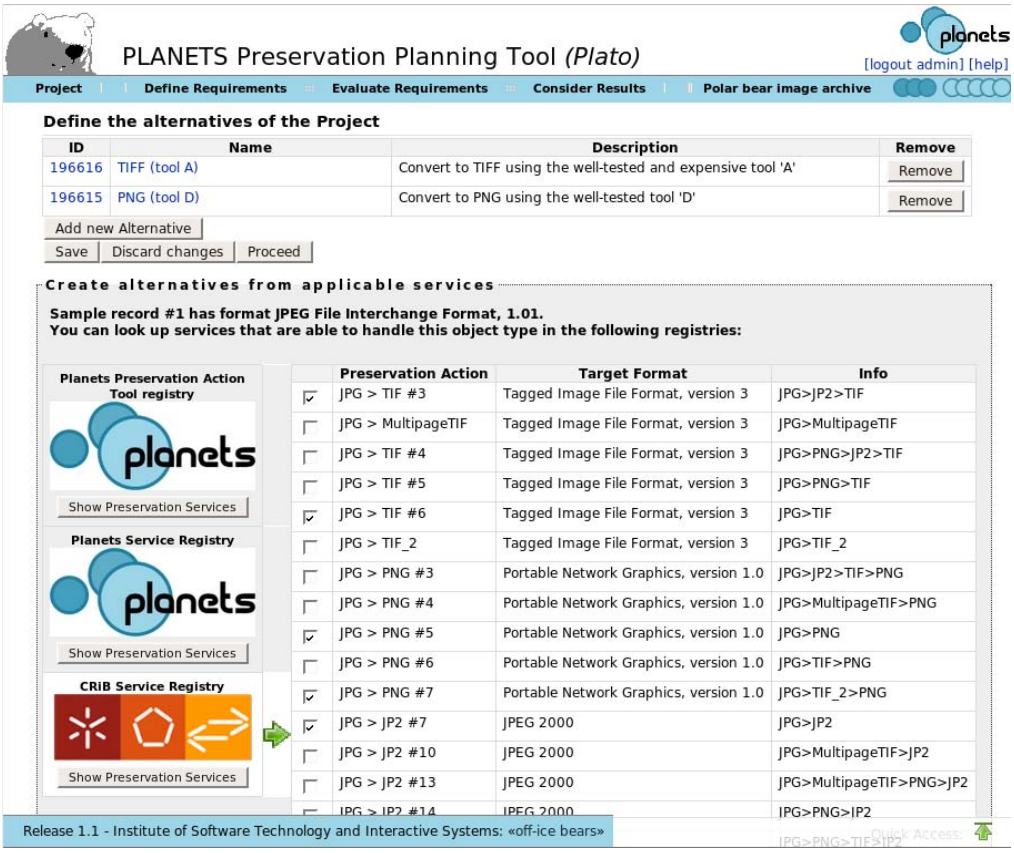
LOOKING INTO THE CRYSTAL BALL: ASSESSING DEMAND FOR DIGITAL OBJECTS IN FUTURE

Determining demand for future access to digital objects is a complex and tricky business. It requires a detailed understanding not only of who will need what, when, and for how long, but also the best way to succeed with different content and formats.

Planets' Preservation Planning team is making substantial in-roads into developing tools and services that will help libraries, archives, universities, government and commercial institutions to plan and preserve their digital collections.

Plato (Planets preservation planning tool) was released in autumn last year. Now, attention has turned to developing a model which will assist organisations in creating preservation policies and provide a framework to embed preservation planning into day-to-day activity. The outcome has been to define more precisely what a preservation plan is, what it embraces, and to arrive at a shared understanding.

This quarter has also seen progress on the creation of two preservation planning tools. The Planets Comparator will make it possible to identify any changes to objects that have occurred during the migration process. A design has also been developed for a second tool, which will allow organisations to profile their collections automatically. The first version of this tool will be in place in July 2008.



Screenshot of Plato. All rights reserved © Planets 2006-2010.

Plato has been updated to better support the preservation planning process. The latest version incorporates new functionality to help define organisational needs at the outset of the process and confirm a preferred preservation plan in the final stages.

The tool was demonstrated at the 'Planning the Future with Planets' tutorial in April, where delegates used it to create plans to preserve simulated Government and private digital archives; computer games for a games museum and private sailboat images collection.

UNDERSTANDING DIGITAL OBJECTS

Preservation Characterisation is a core component of the Planets approach that supports automatic identification of the essential characteristics of the digital materials we wish to preserve. In order to do this, one needs a language in which to express the characteristics, a way of extracting them from the files, and a way of comparing them. Planets has developed two related languages to support this.

The Extensible Characteristics Description Language (XCDL) makes it possible to describe digital objects. Typical characteristics include colour depth of an image, number of images within a textual document, metadata included within a file or the font type of text. The Extensible Characteristics Extraction Language (XCEL) makes it possible to define how characteristics can be extracted from a digital object. Using a software application called an interpreter. Planets can automatically execute the XCEL definitions to extract XCDL characteristics from a file. During the first year of the project, Planets proved this concept in the domain of images. Since then, Planets has applied the concept to the domain of text documents.

- XCEL - a characteristics extraction language
- XCDL - a characteristics description language

Planets has now extended this work by developing a Comparator that enables XCDL characteristics of two files to be compared automatically. This is an essential building block for digital preservation. For example, consider the case of a repository that has a policy to create a PDF version of every document that it archives. This is a special case of a general migration strategy. The repository manager needs a way to determine whether or not the migration has been successful – and this can be a challenging task. It may be possible to check some documents by hand, but in many settings this would be implausible.

In addition, manual inspection can be very error-prone itself. Some changes may be acceptable. For example, it may be acceptable to change the hyphenation in the application that displays or prints the document.

Other changes may be hard to detect, but important. For example, the repository policy might be to change some original fonts into similar fonts that are available without any external dependencies. The number of pages a text has, *may* be an indication of its authenticity, but if page breaks are assigned during a layout step of text processing, then the number of pages depends on the settings of the individual user's text processing tool. The only solution is to provide a broad palette of characteristics upon which to base judgement and a flexible tool to compare them. This is exactly where the Comparator can help.

PARTNER HIGHLIGHT
Austrian Research Centers GmbH
www.arcs.ac.at

ARC is a national research institution, structured as an industrial limited liability (the "GmbH" part) company, owned in majority by the Austrian federal government. ARC receives approximately 40% of its budget as basic funding from the Federal Ministry for Transport, Innovation and Technology (BMVIT) and has the mission to acquire the rest through competitive research proposals (e.g. projects like Planets) and industrial commissions. A total of around 1,000 employees located at 14 sites across Austria, with the main facilities in Seibersdorf, Lower Austria, and in Vienna. ARC is a valued technology provider and innovation partner for society, industry and commerce.

The research activities of ARC are quite diverse; the company is presently organised into four strategic divisions which reflect the main research areas: Health Technologies, Materials Technologies, Information Technologies, and Mobility & Energy. In Planets, ARC is primarily active in the Testbed and Interoperability Framework, involving staff from the "Digital Memory Engineering" (DME) department of the ARC Information Technologies division. DME is presently involved in another EU project (TELplus - supporting The European Library) and is also working toward commercialisation of Planets results.

Planets is also working on:

i. Converting the output of JHOVE to XCDL. This will mean that whichever file formats JHOVE can work with can also be used with the Planets comparison software. JHOVE is an extensible framework which provides functions to perform format-specific identification, validation and characterisation of digital objects. It is well-known and widely adopted by the digital library community.

ii. Translating the characteristics extracted by commercial tools such as ImageMagick to XCDL. ImageMagick is an open-source software suite to create, edit and compose bitmap images. It can read, write and characterise images in a variety of formats (over 100).

iii. Investigating the possibility of developing conversion tools for extracting XCDL descriptions from binary Office formats. These tools would be based on tools which Microsoft is building for transformation of older, binary Office formats into OOXML.

If these activities are successful, they will enable Planets to process five to ten times as many file formats as had been anticipated a year ago.

MAKING PRESERVATION POSSIBLE

Specific actions must be performed on digital objects, if they are to continue to be accessible for the future. Typically, these actions fall into two categories: those which transform the object itself and those which transform the technical environment required to access it.

Migration involves taking a digital object and converting it into a new format so that it can still be read. Emulation involves providing a tool which replicates the operating environment, in which the software was created, so that users can continue to interact with digital content.

To support these actions, Planets has completed desk research and surveys with prospective users in Denmark and United Kingdom to identify the file formats most commonly found in digital archives. The outcome has been to create an inventory which contains 133 file formats.

Planets' Preservation Actions team has also surveyed the tools that exist to characterise and preserve each file format and conducted a gap analysis. This has made it possible to identify where tools do not exist or are insufficient and put in place a blueprint for tools that will need to be created.

MEET THE STAFF

What is your name and education?
My name is Annette Balle Sørensen. I hold an MSc in biology and a PhD in molecular biology, University of Aarhus, Denmark.

Where do you work?
I work at the State and University Library, Aarhus, DK, as a Senior Advisor to library users in molecular biology and medicine. Before that, I worked for several years as project leader in a pharmacological contract research organisation and as a researcher at the Department of Molecular Biology at the University of Aarhus.

What are your interests in Planets?
Preservation Planning with a focus on how collections and archives of research libraries and data centres are used and the needs and requirements expressed by the users. This work will ultimately define appropriate schemes of how to preserve such collections and archives. Dissemination and Take-up, especially focused on user field studies of communication within academic communities and government agencies.

How did you become involved in Planets?
Planets is an important component in the digital preservation research and development carried out at the State and University Library. Some of my working hours are reserved for research activities within molecular biology field and are conducted with University researchers. This brought me into contact with Planets' work relating to user field studies. Later, I joined the research field in Planets which deals with usage of collections and archives of research libraries and data centres. Being involved in Planets has definitely broadened my knowledge of the world of digital preservation!

Planets' blueprint has been expanded to incorporate emulation tools. Planets is working to integrate second versions of Dioscuri and Universal Virtual Computer (UVC). Remote access to the emulators is provided through the Grate tool, which has been developed by the University of Freiburg.

The Preservation Actions team has worked with organisations such as the British Library, National Library of the Netherlands, IBM, Nationaal Archief Netherlands and University of Freiburg to identify the impact of emulation tools.

These case studies included: running legacy map software using VirtualBox and Virtual PC; accessing an old website using Dioscuri; running computer games using MESS, VMware and QEMU; rendering WorkPerfect files using Dioscuri and preserving interactivity with digital objects using the emulation tool UVC.

The work indicates that emulation can be a viable approach for preserving digital objects. Further work will be required into aspects such as finding hardware drivers for legacy operating systems and accessing the results of rendered objects if the original environment is no longer available.

PULLING IT ALL TOGETHER

The Planets Interoperability Framework (IF) provides shared functions and integrates the Planets tools and services into an easily managed preservation system. Its extensibility allows plug-in of third party tools and services. The first software release of Planets' Interoperability Framework was made available for Planets partners in September 2007. The Planets IF team has completed a second iteration of requirements gathering and design specification. The second release is expected in June 2008.

The team is also working on guidelines for wrapping preservation tools for integration into Planets. The term 'wrapping' refers to the provision of a web service interface so that the tool can be incorporated into preservation workflows. Planets provides an environment that allows software developers to quickly wrap tools (for example, the DROID file format identification tool, or a tool that converts Word docs to PDF's).

MEET THE STAFF

What is your name and education?
My name is James Carr and I hold a BSc in Theoretical Physics from the University of Newcastle, UK, and a PhD in Theoretical Atomic Physics from the University of Nottingham, UK.

Where do you work?
I work at Tessella Support Services plc, a company which is a supplier of software development and support services to the Scientific and Engineering community. I have worked for Tessella for the last ten years, and I currently work in the Digital Archiving Group as a Team Leader and Software Architect.

What are your interests in Planets?
My interests in Planets include preservation characterisation, especially frameworks for high throughput characterisation using tools such as DROID and Jhove etc.

My roles in Planets include being the Preservation Characterisation representative on the Technical Coordination Committee (TCC), which guides, supports and makes recommendations concerning technical issues across Planets.

How did you become involved in Planets?
I became involved in Planets through my work for the UK National Archives SeamlessFlow project. Since then I have been fortunate enough to visit not only the UK archives but also the National Archief in Den Haag and the Schweizerisches Bundesarchiv in Bern.

LEARNING WHAT WORKS

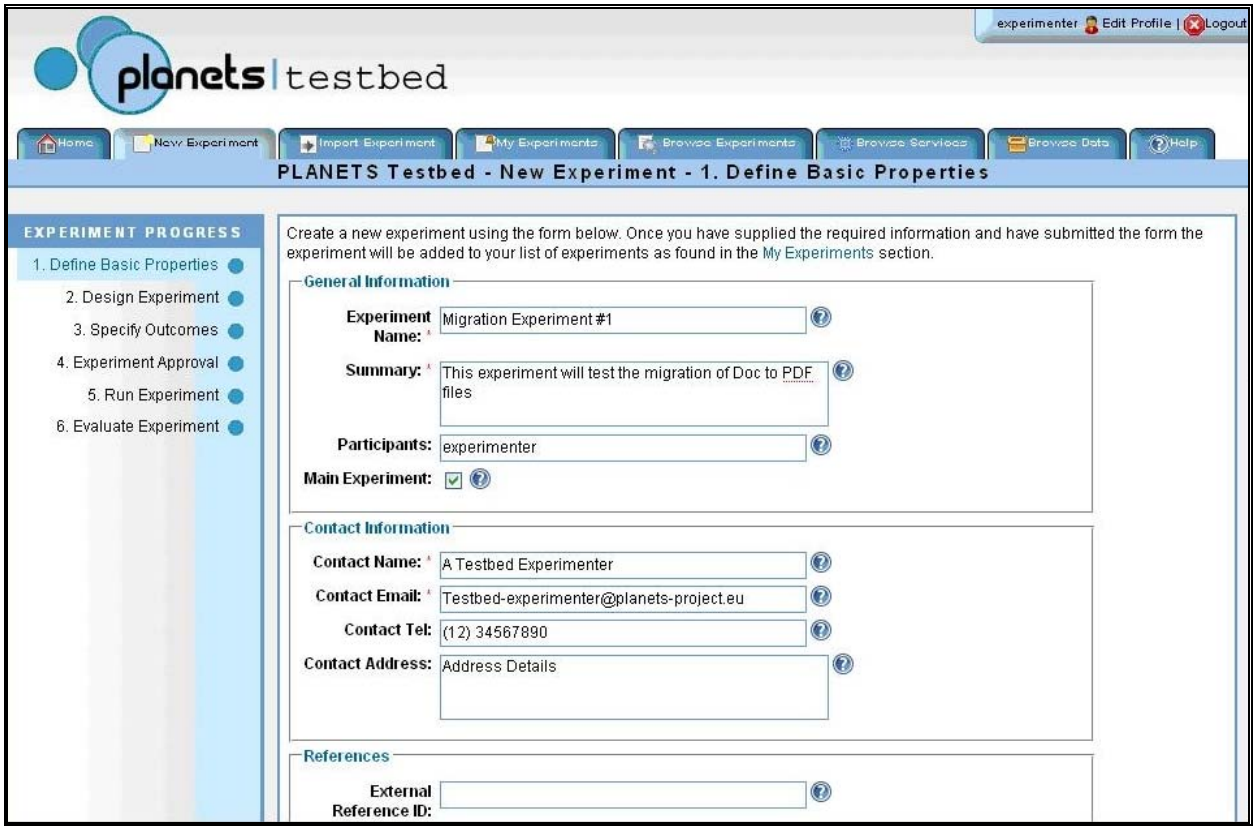
How do we know which preservation strategies and tools will work effectively for different types of content and benchmark these against the experience of others and best practice?

This is the work of the Planets' Testbed, which is being designed to help identify the most appropriate tools and strategies based on the characteristics of specific content. Combining hardware, software and sets of sample data, the Testbed makes it possible to experiment with preservation strategies and tools and analyse the outcome objectively.

The Planets Testbed was released to Planets partners in March 2008. It will shortly be supported by a helpdesk and guidelines to allow partners carry out experiments. A central instance of the Testbed will soon be available in Planets for systematic testing.

The Planets Testbed was demonstrated publicly for the first time at the "Planning the Future with Planets - a preservation planning tutorial", held in April 2008 in Vienna. The next phase of implementation was launched, at a meeting of Testbed staff in London in April, where plans for building the Testbed corpora were established. These groups of reference documents will act as benchmark content, making it possible to ascertain how far Testbed results are reliable and can be reproduced against well-known sets of digital content.

Planets aims to open up the Testbed to external researchers and institutions in spring 2009.



Screenshot of Planets Testbed. All rights reserved © Planets 2006-2010.

FORTHCOMING EVENTS

The following events, organised by Planets partners, are open for participation by anyone with a responsibility for, or interest in, digital preservation in their organisation. To find out more about Planets representation at international events, visit the Planets website.

ECDL2008, European Conference on Research and Advanced Technology for Digital Libraries

14-19 September 2008, State and University Library, Aarhus, Denmark

ECDL is the major European conference on digital libraries and associated technical, practical, and social issues. It brings together researchers, developers, content providers, and users in the field.

More information: please visit the ECDL2008 website.

iPRES 2008, The Fifth Conference on Preservation of Digital Objects

29-30 September 2008, British Library Conference Centre, St Pancras, London, UK

iPRES 2008 is the fifth in the series of annual international conferences and will bring together practitioners and experts internationally to explore trends, developments and applications in digital preservation. It is the longest running conference devoted to digital preservation.

More information, please visit the iPRES 2008 website.

The Third WePreserve Annual Conference

29-30 October 2008, Nice, France

The conference will showcase European digital preservation initiatives and feature demonstrations of CASPAR, DPE and Planets tools and resources.

More information, please visit the WePreserve website

RECENT EVENTS

Planets has organised or taken part in the following events. For more information on Planets activities, please visit the events calendar.

Planning the Future with Planets: a preservation planning tutorial

14-15 April 2008

Forty-five delegates from National Libraries, Archives, and Museums, research institutions, commercial software houses, repositories and publishers assembled at the Austrian Computer Society, on 14-15 April 2008, for the 'Planning the Future with Planets' preservation planning tutorial.

The workshop explored the process of preserving digital content using the Planets approach and technology and was the first in a series of events planned for 2008 and 2009.

"Successfully combined theory and practice"

Andreas Rauber, Vienna University of Technology, introduced challenges to preserving digital objects and Planets technology. Ross King, Austrian Research Centres GmbH, presented an overview of Planets' architecture. Eleonora Nicchiarelli, the Austrian National Library, demonstrated the Planets Testbed. Hans Hofman, Nationaal Archief, Netherlands, led discussions on the requirements for robust preservation plans. Christoph Becker, Vienna University of Technology, presented the Planets planning tool (Plato), approaches to workflow, and to the characterisation of digital objects.

Day Two saw delegates move from theory to practice. Using Plato, delegates devised plans to preserve simulated government and private digital archives; computer games for a games museum, and digital images in a private yacht collector's sailboat collection.

"Everybody was friendly, motivated and enthusiastic".

To download presentations and for more information about future events, visit the Planets website.

Delegates experimenting with Plato

Planets Stars at Cebit 2008
3-7 March 2008

Planets and colleagues from CASPAR, DPE and NESTOR exhibited at CeBIT 2008, the world's biggest Information Communication Technology (ICT) exhibition, in Hannover, Germany.

Like Planets, DPE and CASPAR are strategic projects co-funded by the European Union under its Sixth Framework Programme. NESTOR, the German network of expertise in long-term digital preservation, is a cooperative project funded by the German Federal Ministry of Education and Research.

Under the umbrella of "WePreserve - European Digital Preservation Initiatives", the joint stand was organised to raise awareness about fragility of digital information and showcase solutions under research and development.

Demonstrations of the first prototypes of tools PRONOM and Dioscuri and explanations of the goals of Planets attracted a high degree of interest among business, the general public and the German government.

The event was a great success for the four projects. Almost 1,000 people (165 per day) visited the stand. Among them was Bern Neumann, German Minister of State to the Federal Chancellor and Federal Government Commissioner for Culture and the Media (see photo below).

German Minister of State to the Federal Chancellor and Federal Government Commissioner for Culture and the Media

Some of the visitors were so enthusiastic they could not wait until the end of the Planets project in spring 2010 to make full use of the tools and services: "Great work! Where can I download the software suite?"

What to Preserve? Significant Properties of Digital Objects
7 April 2008

Identifying the characteristics of digital objects is essential to preserving them for the long-term.

This was the theme of the joint workshop presented by The British Library, Digital Preservation Coalition (DPC) and the Joint Information Systems Committee (JISC) at the British Library Conference Centre on 7 April.

Chaired by Chris Rushbridge, of the Digital Curation Centre, and by Neil Grindley, Programme Manager, JISC, the conference brought together 140 delegates who spanned records managers, digital archivists, policy-makers, librarians and academics.

Planets was represented by Adrian Brown (pictured right), of The National Archives UK, who described Planets work on characterisation. Other topics included: Digital Objects Semantics (Stephen Rankin), Significant Properties (Carl Lee, UNC Chapel Hill) and case studies presented by JISC, the INSPECT Project, the SCARP Project and Barclay Wealth. Andrew Wilson, National Archives of Australia and Frances Boyle provided comments and closing remarks.

Determining the properties of digital properties which must be preserved and tools and services to characterise them has been the focus of research conducted by the JISC and the European Union. It was clearly a topic for which there is a hearty appetite.

Frances Boyle, Executive Director, DPC said: “A key message which we all took away from the event was that ‘significant properties’ is a hot topic”, and “The range of delegates’ designations and delegates’ destinations signals the pervasiveness of the topic across many areas.”

For programme information and photographs, please visit:
www.dpconline.org/graphics/events/080407workshop.html

For commentary about the event, please visit:
www.jisc.ac.uk/whatwedo/programmes/programme_preservation/2008sigpropswrkshp.aspx

PLANETS AT EVENTS
Active attendance by Planets partners at major international conferences and workshops is increasing Planets’ profile across the digital preservation community. Planets was seen and heard at the following major events during the first months of 2008:

- CeBIT 2008
  A joint stand with CASPAR, DPE and nestor (see report in this newsletter)
  Presentation: Planets at CeBIT 2008

- SAC 2008 - The 23rd Annual ACM Symposium on Applied Computing
  Presentation: A generic XML language for characterising objects to support digital preservation

- What to preserve? Significant Properties of Digital Objects
  A joint workshop organised by Joint Information Systems Committee (JISC), the British Library and the Digital Preservation Coalition (DPC)
  Presentation: Preservation Characterisation in Planets

- Planning the future with Planets: a preservation planning tutorial
  A combined workshop and training event organised by Planets:
  Presentations:
  Define Tools, Run Experiments and Evaluate Results
  Hands-on Preservation Planning with Plato
  Identification of Institutional Setting and Selection of Sample Records
  Planning the Future with Planets – Introduction
  Objective Trees
  The Planets Interoperability Framework
  The Planets Testbed
  Preservation Planning Process: An Overview
  Preservation Planning: What is a Preservation Plan?
  Preservation Planning: Gathering requirements (discussion session)
  The Preservation Planning Workflow
  Planning the Future with Planets – Summary and Outlook (1) & (2)

PUBLICATIONS
New scientific papers produced by Planets participants include:

- "A Generic XML Language for Characterising Objects to Support Digital Preservation"
  Christoph Becker, and Andreas Rauber (Vienna University of Technology, Austria), Volker Heydegger, Jan Schnasse, and Manfred Thaller (University of Cologne, Germany)
- "Considering the user perspective: research into usage and communication of digital information"
  Kellie Snow, Perla Innocenti, and Seamus Ross (Humanities Advanced Technology and Information Institute, HATII, University of Glasgow, UK), Bart Ballaux, and Hans Hofman (Nationaal Archief, The Netherlands), Birte Christensen-Dalsgaard, Jens Hofman Hansen, Michael Poltorak Nielsen, Jørn Thøgersen (Statsbiblioteket, Denmark)
- "Preservation Planning in the OAIS Model"
  Stephan Strodl and Andreas Rauber, (Vienna University of Technology, Austria)

ABOUT THE NEWSLETTER
Planets will publish a newsletter throughout the four year life of the project. Each issue details recent project activities, describes the practical tools and services developed by the project, and highlights one or two partners involved in the Planets project.

CONTACT PLANETS
Sign up for Planets newsletters via the RSS feed at: **www.planets-project.eu**

For more information on the project, please contact: **info@planets-project.eu**