

Planetarium

The News Bulletin of the
Planets Programme
Issue 7 – July 2009

www.planets-project.eu/publications

MOVING ON TO IMPLEMENTATION

Dr. Adam Farquhar, Planets Project Coordinator

Throughout its third project year, Planets has made strong progress towards meeting its goal of helping content holders to ensure access to digital material for the long term and enabling a marketplace of digital preservation tools and services.

Several partner institutions are moving towards implementing the Planets technology as part of their internal digital preservation activities. I hinted at this in Planetarium 6. Now we have begun to see some real progress in this area, as discussed in this issue. Of course, the project is learning as much as possible from the experiences of partners as they take the tools and methods out for a test drive. We are fortunate to see these projects move forward in very different implementation contexts including a home-grown digital library system, the Fedora open-source repository software, and commercial products. This will ensure that the integration capabilities are thoroughly tested.

We have completed a major survey to help us understand the readiness of organisations to engage in digital preservation activities and benefit from the Planets technology. The survey confirmed some expectations and produced some surprises. It highlighted the central importance of an explicit preservation policy. Compared with those who don't have one, three times as many organisations with a policy have allocated financial resources for digital preservation and four times as many are either currently investing or expect to within the next six months. Of those with a budget for digital preservation, however, 40% have not provided a revenue budget to support their capital expenditure plans. This suggests insufficient awareness of and planning for ongoing costs.

The survey also suggested organisations are taking a conservative approach to preservation actions. Support for emulation ranked close to the bottom in a list of desired capabilities. We see three critical

roles for emulation to take within an overall digital preservation solution. Firstly, an emulator can be the best way to implement migration! In principle, it is very easy – install an old application into an emulated environment, add content, transform the content using the old application (perhaps via a save-as or print to file), and extract the result. The Planets infrastructure makes it possible to do this in practice. I've heard about commercial projects that have gone to great expense to acquire, assemble, and configure multiple generations of physical hardware to implement this sort of process – it would have been much faster and less expensive using remote online emulation capabilities. Secondly, emulation is critical for quality assurance. Curators can explore the content in its original user environment alongside a modern one to determine if anything was lost by a preservation action. Thirdly, emulation can provide end-users with the original experience of interacting with dynamic content such as games or applications.

Engagement with Planets has deepened my understanding of how to think about preservation. The Planets approach takes the broader decision-making context into account and explicitly recognises the role that risk takes. This is important, because major content holders explicitly use risk management in their strategic planning. It is truly critical for valuing activities that are not primarily revenue oriented. I've also seen my understanding of the term 'obsolete format' mature. I no longer worry about waking up and finding that a format has mysteriously become obsolete, rendering content technically inaccessible. Now, we have a much higher level of control. We regularly review organisational goals as stated in explicit policies; we evaluate content against these policies, the needs of the user community, and expected changes in the external environment. This is a managed process that cautious content holders can love and that Planets is helping to deliver.

INSIDE THIS ISSUE

| | | | |
|---|--------|--|---------|
| Moving on to implementation | Page 1 | GRATE – Emulation made easy | Page 9 |
| Automating the process of comparing digital objects | Page 2 | CEI awards bursaries to attend 2nd Planets training event in Sofia | Page 9 |
| Market calls for practical solutions to preserve escalating volume of digital content | Page 3 | Planets at events | Page 10 |
| Integrating Planets into libraries and archives | Page 4 | Publications and conferences | Page 11 |
| Help! My users can't access their objects! | Page 6 | Background to the Planets project | |
| News round up | Page 8 | (if you are new to Planets, start here) | Page 12 |



AUTOMATING THE PROCESS OF COMPARING DIGITAL OBJECTS

Spotlight on Characterisation

Organisations that hold and provide digital material regularly monitor whether their end-users are happy with the content that they are providing and whether their own goals are being met. Sometimes technology changes will have a big impact – for example, a new compression technique could result in dramatic reductions in storage costs, allowing them to increase the number of objects in their collections. At other times, there might be a change in the user community – for example, university students may be less likely to have software to view or print postscript files. Sometimes, the change may be in the organisation's goals – for example, they may extend their services to include students in addition to professionals. Digital object migration is an essential and popular technique to address these and other challenges. In migration, a digital object in one format is converted into a new format that does a better job of meeting the organisational goals. For example, an organisation might migrate images from the TIF format to JPEG2000 to take advantage of improved compression techniques.

Whenever migration takes place, however, we need to answer an important question: did the migration retain the significant information? In some scenarios, this seems easy. A librarian or archivist can simply check the before and after results carefully to determine if the migration was a success. Consider their plight, however, in the case of a large scale migration process which might easily involve millions of digital objects! The comparison task could easily take years of effort. As a result, it is essential to use highly automated processes to compare and check the results of a migration.

Performing an automated comparison can be pretty hard. It is not enough to check the obvious technical characteristics. Consider an image file A before migration and the result B afterwards. Even when a commercial tool reports that it has successfully converted A to B, things may not be too rosy. The two test images below show a file before and after a migration. The technical characteristics, such as image width and height, colour space, and so on are correct in both. The second image, however, is clearly different! This is the result of an erroneous conversion of the representation of the alpha channel in one format to that of the other. In many cases, it is necessary to compare the pixel colour values across the two images.

An image, before  and after  a migration.
(All basic image characteristics are identical.)

To make matters worse, every file format specification uses its own vocabulary for properties and stores the values for these properties in often idiosyncratic ways. The Planets team at the Universität zu Köln (UzK) is developing techniques that make it possible to compare characteristics across files in different formats. The foundation of their approach is a pair of XML-based languages called the Extensible Characterisation Definition Language (XCDL) and the Extensible Characterisation Extraction Language (XCEL) .

XCDL: a language for describing digital objects

An XCDL provides a way to write down the characteristics of a digital object, for example, the width of an image in pixels, the sequence of words in a paragraph, or the font of a letter. It also provides a way to describe aggregate properties of an object such as the number of pages, the number of images, the number of images on a page, the loudness of an audio stream, or the number of rows in a data table. A major challenge in the design of XCDL was to ensure that it would support a very broad range of content types.

XCEL: a language for producing descriptions

It is great to have a language to write down the characteristics of digital objects – but this is only useful if there are ways of producing these descriptions! The XCEL language helps to solve this problem. Experts use XCEL to specify how a file format encodes information – and each file format uses a different encoding. For example, some image formats encode a property like image width in two bytes at a fixed offset from the start of the file; others might use a list of keyword/value pairs with a keyword, like 'width', to indicate the meaning of the following number.

An XCEL specification contains 3 types of elements: items, symbols, and processings.

- Items structure a document into meaningful parts. The file format specification determines how the parts are laid out across a file. For example, a PNG image consists of a fixed-length signature followed by variable length chunks which contain the palette, image data, and so on. Each of these chunks would be represented by an item.
- Symbols correspond to the characteristics that are encoded in the items. For example, the compressed image data that is in the image data chunk of a PNG file is represented by a symbol.
- Processings are used to change the behaviour of a parser, such as the Extractor, during runtime. For example, a processing might specify the need to decompress the image data and transform it into a normalised form before it is stored as the value of a symbol.

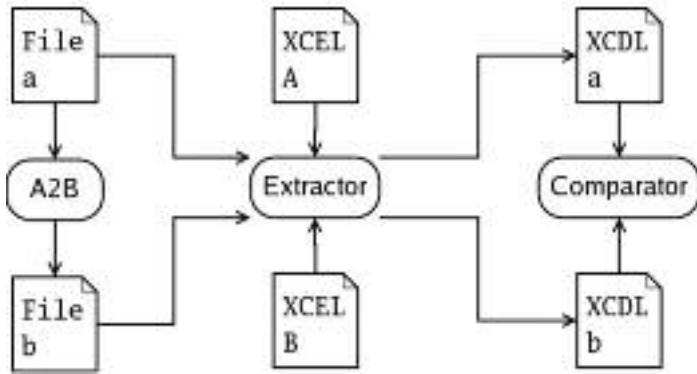
Extractor and Comparator: essential software

The team has developed two key software components to exploit the possibilities of XCEL and XCDL. The Extraction Engine uses the XCEL specification for a format to reach into a digital object, pull out the characteristics that are needed, and write them out in XCDL. The Comparator takes two XCDL descriptions and compares them using a broad range of metrics to determine how similar (or different) two digital objects are.

¹ http://planetarium.hki.uni-koeln.de/public/XCL/Deliverables/planets_pc2-d7_finalxcdspec_ext.pdf

² http://planetarium.hki.uni-koeln.de/public/XCL/Deliverables/planets_pc2-d8_finalxcel specification_ext.pdf

The following graphic shows the relationship between all of these components. A migration service converts File a into File b. The extractor uses the XCEL for the format A and format B to extract descriptions XCDL a and XCDL b. The comparator takes in these two descriptions to produce a measure of how different they are.



The XCL-Ontology

Extracting descriptions from a digital object is very useful – but an important problem remains. Not only does each file format encode information differently, but there is wide variation between the meanings of properties in them. For example, one format may represent image width in pixels, and another may use samples-per-meter. In order to compare characteristics across file formats, UzK has developed an ontology.

The XCL Ontology defines file format properties and the relationships between them. It has been specified using OWL Lite (Web Ontology Language as defined by the W3C) so that it is straightforward to manipulate and implement. The XCL Ontology makes it possible to compare apples and oranges by relating the file format properties to a set of basic information types. Text and image information types are already fully supported.

What's next?

The next major development will be to support composite formats in comparisons. A composite format may include more than one byte stream, such as word processing files that contain images, or presentations with embedded video or sound. The most recent releases of the XCL languages and extractor software introduced support for them. The comparator will provide support in the next major release, scheduled for September 2009. The performance of the extraction engine is being substantially improved. The team is developing an underlying information model that extends Langefors' infological model. This will make it easier to capture and compare differences across subtypes of digital objects such as vector and bitmap graphics. It will also support semantic distinctions that are not called out in the format, but expressed in its rendering. For example, footnotes in PDF files are indicated by font and location, rather than an explicit tag as found in formats such as LaTeX or OOXML.

[3 http://planetarium.hki.uni-koeln.de/public/XCL/ontology/XCLOntology.owl](http://planetarium.hki.uni-koeln.de/public/XCL/ontology/XCLOntology.owl)

READY BUT ABLE? MARKET CALLS FOR PRACTICAL SOLUTIONS TO PRESERVE ESCALATING VOLUMES OF DIGITAL CONTENT

Preliminary Findings of the Planets Long-Term Management of Digital Information Survey

Nearly 200 organisations took part in Planets' Long-Term Management of Digital Information Survey in February and March 2009 to provide a global insight into the readiness and appetite for services to preserve digital content.

The results show that there is a widespread awareness of the need to take practical steps to preserve the rapidly increasing volumes of digital content; however, many organisations are still seeking the technical and practical solutions to support their preservation-specific activities.

Respondents who took part in the survey represented a variety of fields, ranging from digital preservation specialists, librarians and archivists through to IT and project managers. Interest has percolated beyond those charged operationally with managing collections with one in six respondents being institutional Directors, Heads of IT or Heads of Preservation. Digital preservation is emerging as a discipline in its own right: Six per cent of respondents specialise in digital preservation.

Some of the preliminary findings are as follows:

- There is an active awareness, both geographically and across many sectors, of the need to preserve digital information. Ninety-three per cent of respondents said their organisation is aware of the problem of preserving digital content and seventy-five per cent said that they are actively seeking a solution or have one.
- Over the next ten years, the volume of digital content that organisations intend to store is expected to rise twenty-five-fold. 80 per cent of organisations currently hold less than 100 terabytes of data; by 2019, 70 per cent intend to store over 100TB. The proportion expecting to hold more than 1 petabyte now will increase eight-fold from 5 to 40 per cent.
- Half of respondents' organisations have a policy in place to guide their long-term management of information. Commercial institutions (88 per cent), archives (64 per cent) and suppliers and vendors (60 per cent) lead the way in having a digital preservation policy, with libraries (43 per cent) and government departments and public sector organisations (27 per cent) trailing behind.
- The existence of a preservation policy is a crucial first step to developing a solution. Three times as many organisations with a policy (75 per cent) as without one have allocated a budget to finance preservation activities. They are also 50 per cent more likely to incorporate digital preservation into their business continuity, operational and financial planning.
- Organisations with a policy are also likely to be investing sooner. Nearly three times as many (57 per cent) organisations with a policy, as without, are investing now or plan to invest in the next six months. Over twice as many organisations without a policy, as with, will not invest for another two years (34 per cent versus 13 per cent).

- Regardless of the existence of a policy, 85 per cent of organisations are currently assessing their requirements using consultancy or a prototype, or are involved in a tender process, or already have a solution in place.
- Organisations must deal with a wide range of file formats, most commonly documents, images, databases, audio, websites, video, and e-mails. They must also accept content from a wide range of sources such as: file systems, document scanning programmes, the internet, electronic data management systems, e-mail, electronic records management systems and media digitisation programmes.
- There are differing perceptions about organisations' ability to influence the formats in which digital information is provided to them. National archives are ten times more likely than national libraries to say that they have complete control over the formats in which content is delivered to them, while national libraries are three times more likely to say that they have little or no control over formats.
- Approaches to implementation vary and are split equally between those who contract with a third-party to develop a solution and those who develop one in-house. The majority (64 per cent) has chosen to integrate digital preservation components into a custom solution. The remainder will develop a custom solution or use an off-the-shelf package. Respondents are also combining these approaches.
- Approaches to the source of software to manage digital information differ. Almost three-fifths of respondents use a mix of proprietary and open-source software, and will continue to do so. Thirteen to fourteen per cent will rely solely on open-source, whereas reliance solely on proprietary software will decrease sevenfold from 14 to 2 per cent.
- There is broad consensus about the capabilities required of a digital preservation solution. Respondents agree that the top eight capabilities of such a solution should have are: the ability to

maintain the authenticity, reliability and integrity of information; to check records are undamaged; to plan the preservation of content by dealing with obsolescence; to comply with established data and digital information management standards; to ensure records are accessible for up to 50 years; to perform migrations to deal with technical obsolescence; to store many different types of content, and to handle a wide variety of file formats.

- Scalability is critical. Seventy per cent of respondents rated scalability to high content levels (petabytes of data) as critical or very important, while national libraries are three times more likely than national archives to regard scalability of content as critical.
- Interest in Planets is high; over three quarters of respondents would like to receive electronic updates about Planets. Over sixty per cent of respondents definitely see Planets as a potentially valuable resource for preservation planning tools (70%), characterisation tools and methodology (63%), and tools to transform the format of digital objects (62%).
- The results show that interest reaches beyond the project's tangible outputs to the services it could offer. Over sixty per cent of respondents definitely see Planets as a potentially valuable resource for information about the latest developments in the field of digital preservation (68%), and certification of preservation tools (65%).

Findings from the survey will be used to inform both the development of Planets tools and services during the final year of the project, and a framework to sustain and build on Planets outputs after May 2010. The report on findings will be made available on the Planets website at www.planets-project.eu/publications from autumn 2009. Contributors will be given earlier sight of the report.

Our thanks go to the organisations and the individuals who took part in the survey and to contribute to the development of practical solutions to support the long-term management of digital information. Special thanks go to Tessella, which conducted the study as a member of the Planets Consortium: www.tessella.com

INTEGRATING PLANETS INTO ARCHIVES AND LIBRARIES

"A library is not a last resting place for the books contained there but a place where information and ideas live and breathe in new minds. To continue to do so, the materials collected – whether books or web pages – must themselves be alive and fresh, in forms and formats that preserve their character and make them accessible to new readers."⁴

In 2000, the Library of Congress in its Digital Strategy began to consider the impact of digital information and networks on its core practices. It concluded that, in addition to defining roles and responsibilities in relation to preservation work, it must enhance its technical capacity and expertise in digital preservation.

From the outset, Planets has recognised that, as a result of their legal obligations, National Libraries and Archives – whether in Europe or beyond – are uniquely placed to safeguard digital information and provide enduring access to digital cultural and scientific knowledge.

This need was a, if not the, main driver behind the project. Planets was expected to help develop understanding, as well as tools and services, to enable Consortium members and other such institutions to address the direct challenge of digital preservation.

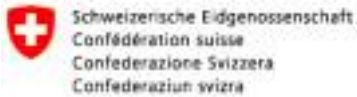
As Planets moves into its fourth and final year, work is underway to integrate Planets components into the operations of National Libraries and Archives within the Planets Consortium. In February 2009, the Integrating Planets into Libraries working group (IPIL) was formed to enable Consortium partners in libraries to exploit Planets' outputs to meet the challenges of preserving their digital collections. This initiative, renamed IPAL (Integrating Planets into Archives and Libraries), has now expanded to include parallel activities by Archives to assess how Planets can help to address archive-specific challenges.

The following page provides examples of how IPAL members might best take advantage of Planets capabilities.

⁴ A Digital Strategy for the Library of Congress, Committee on an Information Technology Strategy for the Library of Congress, Computer Science and Telecommunications Board, National Research Council, 2000.

The British Library holds a huge, complex and heterogeneous digital collection comprising 80 terabytes of digitised newspapers, 60 terabytes of web-harvested information, e-journals and books, 25 million pages of digitised C19 literature, broadcast television, digital video and digital maps. Its goal is to store its entire digital collections in a single repository – the Digital Library System – by 2016. It also plans to construct a new Digital Research Centre – effectively a reading room dedicated solely to accessing its collections by digital means.

Work has already begun by its Digital Preservation Team to incorporate Planets into activities to preserve the Library's digitised newspaper collection. The British Library has used Planets Plato tool to create a plan to reduce threats to the collection, Planets characterisation languages XCDL and XCEL to identify and ensure important features are retained and, after experimentation in the Planets Testbed, the JMagick tool to automate the process of large-scale migration of the collection from TIFF to jpg 2000 format. The British Library intends to adopt Planets as a key technology component of its digital preservation solution.



The Swiss Federal Archives (BAR) is legally responsible for preserving documents – paper records, photos, audio and film

recordings – of the Swiss government, the parliament and the federal administration. Its digital material amounts to roughly 13 terabytes, increasing by an estimated 1 terabyte in 2010. From 2012 on, all federal administrative institutions will be asked to deliver master records to the archives in digital form only. Consequently, BAR anticipates very large digital data deliveries, including databases, in the next two decades.

BAR established SIARD as a mandatory file format for database deliveries in 2007 and is integrating the SIARD tool for database conversion in the Planets Framework. It has been testing Planets tools for preservation planning, characterization and migration on databases, videos and text data such as files and emails.

KB

Koninklijke Bibliotheek

In the Netherlands, the National Library (KB-NL) accepts copies of every print publication in the Netherlands as part of a voluntary

deposit scheme. In addition, KB-NL holds 12 million digital objects, primarily scientific e-journals in PDF. The Library loads two million items of digital content into its e-Depot, based on IBM's archiving system DIAS, each year. It is also undertaking major digitisation and web-archiving activities including the Digital Daily Newspaper project. Consequently, the volume of content is rising rapidly. At present, in mid-2009 the e-Depot holds 11TB. By 2012 they project that it will hold 290TB. Some of the content is complex, comprising web-sites and multi-media. KB-NL anticipates that it will use Planets to support its preservation planning, to characterise content and to perform migration and emulation activities.



DET KONGELIGE BIBLIOTEK

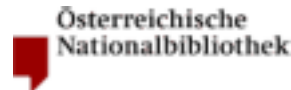
The Royal Library in Copenhagen, is the National Library of Denmark, (KB-DK) and the largest of the libraries of the Nordic countries. It has a legal obligation to preserve Danish cultural heritage for the future. It shares a storage facility with the State and University Library (SB), the library for the University of Aarhus, the National Archives

and government institutions. KB-DK holds 20TB of digital data such as images, books, digital negatives. This is increasing by 7TB per year. Both KB-DK and SB are engaged in archiving Danish internet content. The KB-DK's web-archive, Netarchive, stores 100TB and this is expected to grow by 80 to 100TB each year. KB-DK is planning mass digitisation projects starting in 2010. KB-DK and SB plan to integrate Planets core preservation planning and characterisation tools, the Planets Core Registry and also tools to perform automatic large-scale characterisation and migration activities.

nationaal archief

The National Archives of the Netherlands (NA), is responsible for

managing the archives of the Netherlands and its overseas dominions of the past. It accommodates born-digital material like electronic records and digital camera pictures, and digitised copies from analogue originals such as maps. Due to the nature of its content and its digitising efforts, the NA anticipates managing petabytes of data in the near future. It plans to implement the whole Planets suite of tools. It has already used Plato to compare three different strategies for regaining access to old emails and text files generated on legacy systems which are now obsolete.



The Austrian National Library (ONB) has, within the last five years, digitised around 6 million newspaper and book

pages, pictures and pieces of papyrus. The library has also digitised a significant number of its audio recordings as well as items from special collections such as incunabula, posters and flyers. The ONB also collects online publications and as a result of new Legal Deposit Law in Austria, is now responsible for harvesting the Austrian web domain. ONB will migrate its repository system to Ex Libris' Rosetta solution in 2010. ONB is working closely with Ex Libris to incorporate Planets preservation capabilities during pre-ingest to characterise digital objects, create preservation plans, carry out actions to transform objects into accessible formats and validate that those actions have been successful. The Austrian National Library is planning to use the Planets Testbed to decide on appropriate file formats for image preservation, taking into account measurable properties such as rendering efficiency, robustness and error resilience.



IPAL's work indicates that Planets has an important role to play in helping organisations to plan and to start to take early action to preserve sizeable and complex collections.

Max Kaiser, Head of Research and Development, Austrian National Library says: "The digital preservation community expects a lot from Planets. It has an especially important role to play in preservation planning."

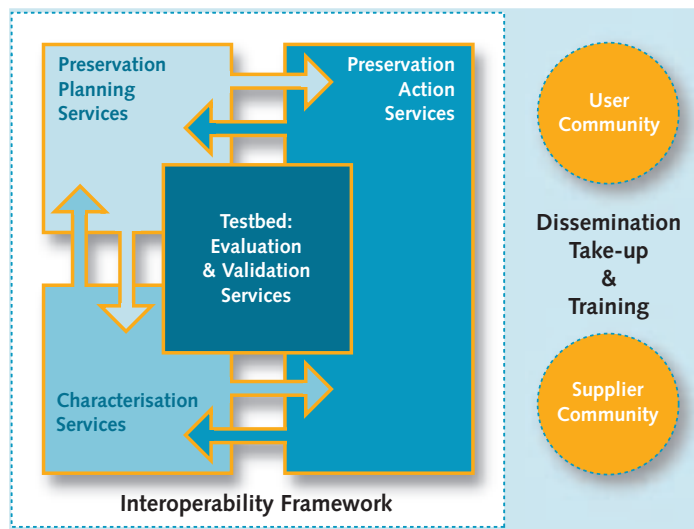
HELP! MY USERS CAN'T ACCESS THEIR OBJECTS!

“My users are reporting problems working with content in my repository. They say they don't have access to the applications they need to view the content. What can I do to help them?”

Here we describe some of the work Planets is doing on tools to make digital objects accessible and the services it will provide.

Planets will provide an integrated set of tools to assist curators of digital materials at each stage of the digital preservation process. It provides preservation planning workflow, via the Planets tool Plato, for:

- preparation of a risk-based preservation plan
- ability to identify, test, apply and validate preservation action tools against digital collections in a controlled environment without putting the collections at risk (via the Planets Testbed)
- execution of selected preservation actions.



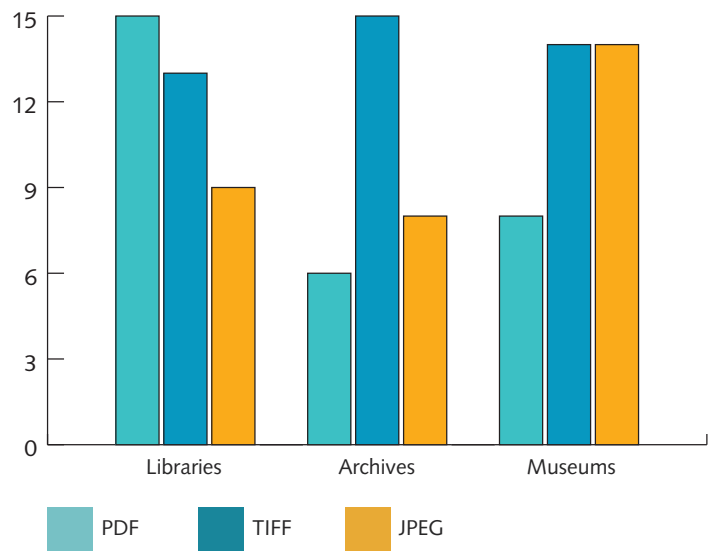
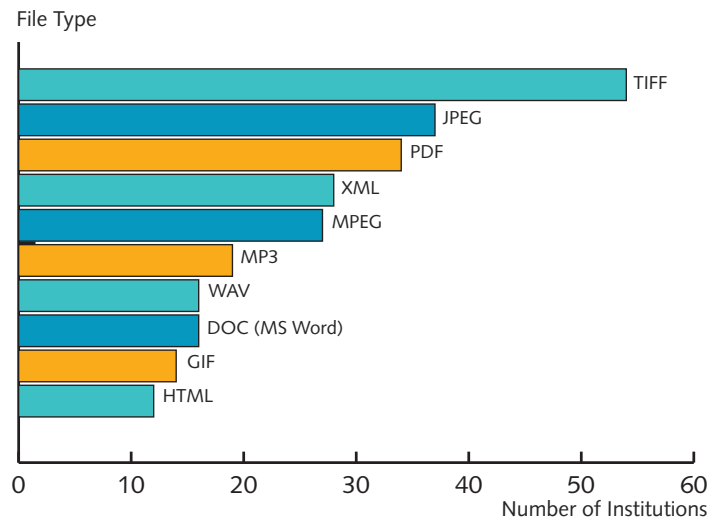
“Does a tool exist to enable me to convert a high-risk file format to a current one and still retain the characteristics which make it fit for purpose?”

In order for Planets to know which preservation action tools it will need to provide, it has to know which file formats are in use by the organisations who are potential users of Planets tools and services. One can then study which tools are available for these formats and determine whether any tools are missing. Planets is defining a Blueprint for the development of preservation action tools, in other words, a specification of the functional requirements to which new migration and emulation tools should adhere.

Planets has therefore conducted a survey to determine the file formats in use by various types of cultural heritage institutions within Europe and also the United States. Each institution which took part was questioned about which file formats they archive for the long term. The survey focussed on the range of formats rather than the quantities stored in each.

The questionnaire was returned by 76 institutions from thirteen countries, most of which were archives, libraries or museums. The inventory of file formats which resulted contains 124 file formats, where different versions of a format are counted as one. Of these, 67 were identified as ‘niche formats’, i.e. in use in only one institution.

The chart below shows the top ten formats found in the survey.



As can be seen from the second graph, there are some significant differences between the occurrences of the top three file types in different kinds of institution.

The results of our survey were compared with other existing, but differently focussed, research, namely NESTOR, which surveyed German museums, and ROAR, the Registry of Open Access Repositories, which contains information about open access e-print archives. These studies broadly support our findings.

Now that we are confident of which file formats are in use by cultural heritage institutions, we can look into whether tools are available to preserve these formats.

This is where the Planets Core Registry (PCR) helps.

The PCR⁵ is a registry which holds details of file formats and also the preservation action tools available to manipulate them. (The Project has already identified a wide range of 3rd party preservation action tools available under a variety of licensing agreements capable of

⁵ The Planets Core Registry (PCR) is a combination of the earlier Planets Preservation Characterisation Registry (based on PRONOM) and Preservation Action (Tool) Registry.

being invoked by the Planets framework) The PCR will be integrated with Plato and the Planets Testbed towards the end of 2009. Work done so far shows that there are tools to preserve the top three formats in use, plus a number of others. Nine out of the top ten formats (as well as many of the less used ones) are addressed by preservation action migration tools already wrapped by Planets, and the remaining one does not require any tools. (Wrapping a tool makes it capable of being deployed as a service within the Planets Testbed – subject to appropriate licensing.)

The list of input and output sources and formats is being continually updated and currently includes the following:

Input: CGM, DOC, DOCX, DVI, DXF, EPS, GIF, HTML, JPG, MDB, ODP, ODS, ODT, PDF, PNG, PPIX, PPT, PS, SVG, TIFF, VDX, WMF, WPD, XCT, XLS, XLSX and MS SQL Server, Oracle, SIARD

Output: CGM, DOCX, DXF, EPS, GIF, HTML, JPG, JP2, ODP, ODS, ODT, PDF, PDF/A, PNG, PPTX, PS, SVG, TIFF, TXT, VDX, WMF, XENA, XLSX, XML and MS SQL Server, Oracle, SIARD

An alternative route to preservation may involve emulation:

- to provide access using software that only runs on out-of-date hardware (this might then enable migration to a new format)
- to provide end-users with the experience of interacting with dynamic content such as applications or games
- to provide quality assurance capabilities (compare with migrated object)

There are a number of emulation tools available within Planets:

- Dioscuri emulates MS DOS 4.0, 5.0, 6.22, MS Windows 3.0, ELKS 16-bit Linux;
- QEMU currently emulates X86, ARM, SPARC, MIPS, MIPS64, N68k (Coldfire) and will be able to emulate SPARC64, POWERPC, POWERPC64, SH-4, Alpha, CRIS in the future;
- Universal Virtual Computer (UVC)

By using GRATE⁷, Planets users do not need to install the emulators separately but have remote access.

However, knowing that a tool exists to migrate/emulate a particular file format does not necessarily mean that it addresses the preservation need of a particular organisation for digital collection items in that format. One needs to consider the output format as well as the input format and whether the properties of the objects significant to the organisation are adequately preserved (as well as any other requirements the organisation may have specified). This is where the preservation plan and the Planets Comparator and Testbed help. The Planets Comparator is invoked by Plato and makes it possible to identify changes that may have occurred to the characteristics of a file during a migration process. These might include, for example, the font type of text, or the colour depth of an image. See Planets newsletter issue 4 for discussion of characteristics of files – Understanding Digital Objects.

The Planets Testbed is 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, to discover which tools best address the specific organisation's preservation requirements.

Work on the Planets framework is ongoing, and the Testbed corpora is to be developed with examples of files in the most commonly used formats. Testing will be done, using the PCR, and guided by the file format inventory from the survey, to establish whether the existing tools address the likely specific preservation needs of organisations in migrating specific formats. This work will help to identify where, if any, there are real gaps in tool provision.

There is still the onus on the institution wishing to ensure the longevity of its digital collections to risk-assess the formats being used for storage of its digital items – will they be preservable?

“So, are there tools available for me to convert my digital objects?”

Only you can tell for sure, because it depends on your specific preservation needs. But once you have determined these, Planets can help you to find an answer!

References

Planetarium 5, p2 PLATO – a not so philosophical approach to preservation planning

Planetarium 6, p2 – Close Encounters of the Digital Preservation Kind: Spotlight on the Planets Testbed

NESTOR – http://www.langzeitarchivierung.de/downloads/mat/nestor_mat_02.pdf

ROAR – <http://roar.eprints.org/>

Planets Newsletter 4, p3 – Understanding Digital Objects

⁶ Dioscuri developed by Tessella Technology & Consulting, The National Archives of The Netherlands and the National Library of the Netherlands

⁷ See page 9; 'GRATE: Emulation Made Easy'

NEWS ROUND UP

Planets Takes to Europe

The volume of digital content that organisations expect to hold will increase fifty-fold in the next decade. Yet the average digital object will be obsolete in less than the same period of time.

While information can be generated more quickly and shared more easily online, this means that time and effort must be invested if that information is to be available ten years from now.

To this end, Planets is hosting a series of five three-day training events, 'Digital Preservation – The Planets Way', in locations around Europe between June 2009 and May 2010. The first took place at the stunning Black Diamond building in the Royal Library, Copenhagen on 22–24 June 2009 and which coincided with the Midsummer festivities and bonfires in Denmark's capital.

Day one introduced CEOs, Heads of IT and Conservation/Preservation and digital preservation staff to the case for investing in long-term management of digital collections as part of business risk management; the issues involved in preserving digital content

400 People Sign-up to Planets E-Bulletin Service and User Community

Four hundred people around the world have registered to join the Planets new e-bulletin service and User Community launched.

The new service which was launched in mid-March will help to keep organisations informed about Planets outputs during the final year of the project and in the run up to the release of technology components.

Subscription is the fastest and easiest way to learn about Planets without repeatedly visiting the Planets website.

By registering to the Planets user community and e-bulletin service, subscribers receive news updates about developments in the project including:



Black Diamond building, Royal Library, Copenhagen

and practical solutions. The National Library of the Netherlands demonstrated how it is preserving its digital collections.

On days two and three, digital preservation staff (digital librarians and archivists, repository managers, software developers, policy managers) saw practical demonstrations and gained hands-on experience of Planets working with sample collections. Attendees had an opportunity to talk to and learn from international experts in the field of digital preservation and the architects and developers of Planets tools and services. The event included speaker panels and interactive discussion sessions.

For more information and to see the programme, visit: <http://www.planets-project.eu/events>

- releases of reports and papers
- the Planetarium newsletter
- tools and services
- advance notice of training and outreach workshops
- invitations to test preservation tools such as the Planets Testbed

The service is especially relevant to researchers, librarians, archivists, curators, preservationists, IT staff, analysts and others whose work involves caring for digital collections. It will also be relevant to providers of services such as digital library systems, consultancy, educational publishing and training.

To subscribe, please visit: <http://www.planets-project.eu/community>. You can unsubscribe at any time by sending an e-mail to info@planets-project.eu.

Upcoming events in the Planets Testbed

In the upcoming months, the Planets Testbed will be opened up to a broader community. Currently, the Testbed-Beta version allows Planets partners to perform migration and characterisation experiments using a corpus of test files.

By autumn, experiments on emulation tools will be possible and external institutions and organisations will be invited to perform experiments within the Testbed. External institutions will also be invited to take part in large-scale trials of the technology at the end of this year.

The first opportunity for institutions to try out Planets' scientific testing environment was the Planets training event at the Royal Library, Copenhagen on 22–24 June. Attendees at the Copenhagen training event obtained practical, hands-on experience of using the Planets suite of tools and services, including the Testbed, to gauge how the technology could best be used by their individual institution or company.

The Planets Testbed provides a controlled laboratory environment in which users can test preservation strategies and tools on particular types of digital objects to assess their suitability. It means that users can select the tools and services to preserve particular content based on the results of scientific experiments and empirical evidence.

For information about new Testbed developments, register at the Planets Website: <http://planets-project.eu/community/>

For questions about the Testbed, ask our Helpdesk: helpdesktb@planets-project.eu

Try out the Planets' Testbed, at Planets training events. The next one is to be held in Sofia, Bulgaria, 21-23 September 2009.

For more information about the Testbed environment, see our previous issue of Planetarium. <http://www.planets-project.eu/docs/newsletters/NewsletterIssue6.pdf>

GRATE – Emulation Made Easy

Hankering after Aztec; missing Mighty Magic I-II; bored without Bad Street Brawler; fed up there's no Firefox, pining for Pitstop 2, lost without Lunar Lander and Leader Board, spaced out without Space Invaders?

Digital objects need soft and hardware environments to be accessed. But, over time environments change and the objects that ran on them become inaccessible. Converting old file formats into newer file formats (migration) is one solution, but migration is not suitable for all digital objects, especially interactive ones such as video and computer games.

Planets has released the first version of its Global Remote Access to Emulation Services (GRATE) tool – a unique preservation service which makes it possible to access 'aged' digital objects such as historic computer games or text processing files by replicating their original environment.



It means that users of games or other ancient files can access tools such as MESS for ancient home computer systems, DOSBox for 16-bit applications, or the flexible multiple platform emulator QEMU allowing access to a wide range of ancient or modern platforms and environments. They can do this using standard Internet technology and any Java-enabled web browser (platform independent), provided the appropriate software elements are pre-installed on the GRATE server.

In turn, they can view and edit e.g. with several document formats like WordPerfect 5.1, AmiPro or interact with computer games in C64 and Atari without the need to install emulation software, target operating systems, applications, drivers or software. Unlike most of the commercial remote desktop applications, the tool and the software components it depends on are open-source.

Although currently only a Linux version is available, in principle the GRATE server could be executed on operating systems like Windows in the future.

Objects can automatically be identified using services such as DROID and the Planets Core Registry.

They can be transported from a digital repository into the emulation environment so that they can be viewed or edited. Keystrokes can be sent to the emulated environments to control software running 'inside' the emulated operating system (eg. to shut down the emulated machine). Modified objects can be saved back to the local system. The GRATE server is responsible for session management (establishing and terminating sessions, transporting the object, executing emulators and delivering meta-information.).

GRATE has been developed by PLANETS partner the University of Freiburg.

For more information about GRATE, visit: http://www.planets-project.eu/docs/reports/Planets_PA5-D7_GRATE.pdf.

To test the GRATE software service, visit: <http://planets.ruf.uni-freiburg.de>

CEI Awards Fifteen Bursaries to Attend 2nd Planets Training Event in Sofia!

The Central European Initiative (CEI) has joined forces with Planets to award 15 training bursaries to attend the second Planets training event in Sofia, Bulgaria, on 16–18 September 2009.



Thanks to the CEI, the bursaries, each worth up to almost E 600, will entitle fifteen delegates from nine designated countries to attend the 2nd Planets training event free-of-charge. The bursaries will cover the costs of tuition, accommodation and all meals.

The bursaries are intended for staff who can demonstrate that they are engaged in, or about to engage in, work to preserve large collections or collections that are nationally, culturally or scientifically significant. Eligible countries are: Albania, Belarus,

Bosnia and Herzegovina, Croatia, Macedonia, Moldova, Montenegro, Serbia, Ukraine.

To apply, please send an email to show how you meet the selection criteria, with your CV, an outline of your current job role and responsibilities, (500 words max.) and a separate short supporting email from your senior manager or Head of organisation, to the bursary administrator, Edith Michaeler, Austrian National Library, at: sofia-grants@planets-project.eu. The closing date for applications is 27th July.

For enquiries please e-mail: sofia-grants@planets-project.eu. For more details about the 2nd Planets Training event, selection criteria and how to apply, please visit: <http://www.planets-project.org/events>.

The CEI is an Italian-based forum founded in 1998 to foster co-operation between 18 Central and Eastern European countries and their integration into the European Union. In its 2007-2009 Action Plan, it has identified a 'Digital Divide' in which more and more digital content is being created but, without systematic action to preserve it, it could be lost. Bursaries totalling E9000 support the CEI's work to promote the use of digital technologies as a foundation for economic and social development and to preserve member states' growing digital cultural and scientific heritage.

PLANETS AT EVENTS

DPE/Planets/CASPAR/nestor Joint Training Event
'The Preservation challenge: basic concepts and practical applications'
University of Barcelona, Institute of Catalan Studies,
Barcelona, Spain, 23–27 March 2009

Thirteen speakers and thirty-three delegates gathered together at two venues in Barcelona, on 23–27 March 2009, to address the preservation challenge at the 2009 WePreserve Annual Training Event.



Delegates participating in break out sessions.

The week-long training event attracted participants from academic institutions, libraries, archives, local government and commercial companies. Fifteen lectures gave delegates an insight into issues around preserving digital content and solutions emerging out of publicly funded digital preservation projects. These were supported by case studies illustrating of the practical application of the tools developed by DPE, Planets, CASPAR and Nestor. Delegates participated in break-out sessions where they could review the topics discussed during the lectures and receive one-to-one advice from the experts. Pre-course and post-course online training was available to supplement their learning.

Manfred Thaller (University of Cologne) opened the event by asking what should we preserve and how. Subsequently, he examined which file formats are useful for long-term preservation. Max Kaiser (Austrian National Library) and Brian Aitken (Humanities Advanced Technology and Information Institute)

Digital Preservation – The Planets Way is a series of five three-day training events to be held in locations around Europe between June 2009 and May 2010. The second event, in Sofia, organised by the Austrian National Library and the Humanities Advanced Technology and the Information Institute at the University of Glasgow, will equip up to 60 staff such as archivists, librarians, preservation, project managers, IT to preserve digital collections with the knowledge and practical skills to initiate or advance work in their organisation.



Hannes Kulovits presenting Plato.

presented the Planets Testbed. Angela Dappert (British Library) considered the role of metadata for the effective management, retrieval and re-use of digital information. Hannes Kulovits (Vienna University of Technology) presented Planets preservation planning tool, Plato. Hans Hofman (National Archives of the Netherlands) examined specific approaches for approaching long-term management of digital information.

Other topics included: standards; the OAIS Reference Model; authenticity; intellectual property rights; repositories; trust, audit and assessment methodologies, including PLATTER (Planning Tool for Trusted Electronic Repositories) and the DRAMBORA toolkit.

All delegates said the event had met their expectations. Over 70 per cent of delegates believed the event had been well-organised, well-structured and the teaching and speakers had been effective.

WePreserve is led by the Humanities Advanced Technology and Information Institute based at the University of Glasgow. It is a collaborative forum for three projects funded under European Commission Framework Programme 6: Digital Preservation Europe (DPE); Cultural, Artistic and Scientific Knowledge for Preservation, Access and Retrieval (CASPAR) and Planets plus Nestor, the German preservation network and the Digital Curation Centre. The forum provides a joint web-portal and bulletin board and educational activities.

To download presentations, please visit: www.planets-project.eu/publications

PUBLICATIONS AND CONFERENCES

PRESENTATIONS

Foundation Rinascimento Digitale tutorial and round table workshop

'Preservation Metadata: Implementation Strategies'

5–6 February 2009, Rome, Italy

<http://www.rinascimento-digitale.it/PREMIS-workshop.phtml>

'PREMIS Implementations at the British Library & PREMIS and the Planets Project'

Angela Dappert, British Library

www.planets-project.eu/docs/presentations/Dappert_Rome2009.pdf

Joint DPE/Planets/CASPAR/nestor Training Event

The Preservation challenge: basic concepts and practical applications

23–26 March 2009, Barcelona, Spain

www.wepreserve.eu/events/barcelona-2009/programme/

'Introduction to Digital Preservation'

Manfred Thaller, University at Cologne

http://www.planets-project.eu/docs/presentations/Thaller_Introduction.pdf

'File Formats and Significant Properties'

Manfred Thaller, University at Cologne

http://www.planets-project.eu/docs/presentations/Thaller_SignificantProperties.pdf

'A Testbed for Preservation Planning'

Max Kaiser, Austrian National Library, and Brian Aitken, University of Glasgow

www.planets-project.eu/docs/presentations/Kaiser_Testbed.pdf

'Digital Preservation Metadata'

Angela Dappert, British Library

www.planets-project.eu/docs/presentations/Dappert_PreservationMetadata.pdf

'Preservation Planning with PLATO'

Andreas Rauber and Hannes Kulovits, Vienna University of Technology

www.planets-project.eu/docs/presentations/Kulovits_PLATO.pdf

'Digital Preservation Process: Preparation and Requirements'

Hans Hofman, The Dutch National Archives

www.planets-project.eu/docs/presentations/Hofman_DPPProcess.pdf

WePreserve Forum 2009

27 March 2009 – Barcelona, Spain

www.wepreserve.eu/events/forum-2009/programme/

'Introduction to Planets'

Hans Hofman, The Dutch National Archives

www.planets-project.eu/docs/presentations/Hofman_Introduction.pdf

'The Planets Preservation Planning workflow and the planning tool Plato'

Hannes Kulovits, Vienna University of Technology

www.planets-project.eu/docs/presentations/Kulovits_PLATO-wePreserveForum.pdf

DigCCurr 2009: Digital Curation Practice, Promise and Prospects

1–3 April 2009, Chapel Hill, NC USA

<http://ils.unc.edu/digcurr2009/>

'Preservation Planning with Plato'

Andreas Rauber, Christoph Becker, Hannes Kulovits, Michael Kraxner, and Riccardo Gottardi, Vienna University of Technology

www.planets-project.eu/docs/presentations/DigCCurr2009Rauber_Plato.pdf

'Home Office Painless Persistent Long-term Archiving (HOPPLA) – Digital Preservation Support for Small Institutions'

Andreas Rauber, Florian Motlik, Petar Petrov, and Stephan Strodl, Vienna University of Technology

www.planets-project.eu/docs/presentations/DigCCurr2009Rauber_Hoppla.pdf

'Personal Archiving: HOPPLA Assumptions and Design Decisions'

Andreas Rauber, Vienna University of Technology

www.planets-project.eu/docs/presentations/DigCCurr2009Rauber_PersonalArchiving.pdf

'Dioscuri – the emulator for digital preservation'

Jeffrey van der Hoeven

www.planets-project.eu/docs/presentations/DigCCurr2009Hoeven_Dioscuri.pdf

INTENSIVE 2009

20 – 25 April, 2009, Valencia, Spain

www.iaaria.org/conferences2009/INTENSIVE09.html

'A Service for Data-Intensive Computations on Virtual Clusters'

Rainer Schmidt, Christian Sadilek, and Ross King, Austrian Research Centers

www.planets-project.eu/docs/presentations/intensive09_schmidt.pdf

PUBLICATIONS

'Systematic Characterisation of Objects in Digital Preservation: The eXtensible Characterisation Languages'

Christoph Becker, Andreas Rauber, Volker Heydegger, Jan Schnasse, Manfred Thaller

In: J.UCS, Journal of Universal Computer Science, vol. 14/Issue 18

www.jucs.org/jucs_14_18/systematic_characterisation_of_objects

'Requirements modelling and evaluation for digital preservation: A COTS selection method based on controlled experimentation'

Christoph Becker, Andreas Rauber

www.ifs.tuwien.ac.at/~becker/pubs/becker_requirements.pdf

Presented at the 24th Annual ACM Symposium on Applied COMPUTING (SAC'09)

In: Proceedings of the ACM Symposium on Applied Computing (SAC'09), Track 'Requirements Engineering'. Honolulu, Hawaii, USA, March 9–12, 2009

'Significance is in the Eye of the Stakeholder'

Angela Dappert, Adam Farquhar

www.planets-project.eu/docs/papers/Dappert_Significant_Characteristics_ECDL2009.pdf

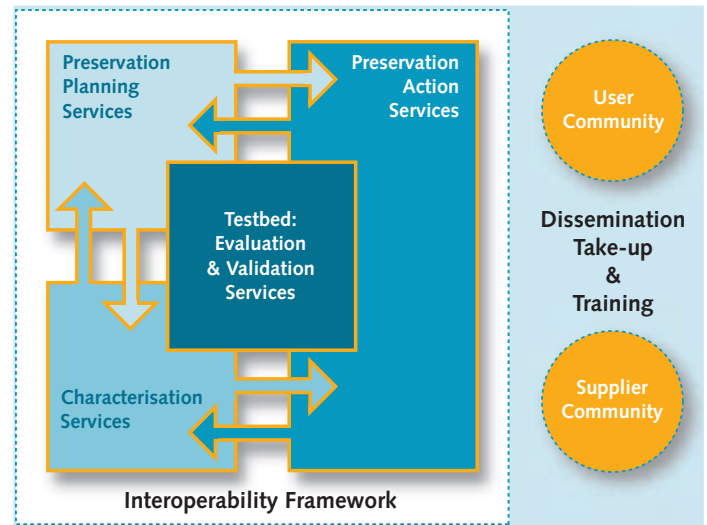
Will be presented at the 13th European Conference on Digital Libraries, ECDL 2009 and published in the proceedings from the conference, the LNCS series by Springer

BACKGROUND TO THE PLANETS PROJECT

Preservation and Long-term Access through Networked Services (Planets) is a European joint-venture for research and development in the field of digital preservation. The project is being delivered by sixteen institutions in Europe and coordinated by the British Library. It is co-funded by the European Commission under Framework Programme 6. (IST-033789).

Planets' is delivering a framework and set of practical tools and services that will enable institutions in Europe to manage and access digital collections for the long-term:

- **Models** to enable you to identify the digital preservation needs of your organisation, decide what information you want to keep and in what format you want to store it. They also help you to understand the ways in which end-users work with these collections.
- **Tools** to allow you to build, shape, compare and execute preservation plans in accordance with your preservation needs. Planets and third-party tools enable you to identify the significant properties of your collection; recommend and execute migration to convert old file formats into new ones or emulation to replicate the original environment; and to compare objects before and after preservation actions have taken place to quality assure and document the outcome.
- **Registry** containing descriptions of file formats, descriptions of available preservation actions preservation and their suitability when applied to particular object types.
- **Architecture** providing access to Planets and third-party preservation tools through a single web-based application.
- **Testbed** and laboratory environment in which you can test these tools and services using real data in a secure environment, so you can make decisions based on scientific evidence.



Planets Partners are:

The British Library
The National Library of the Netherlands
Austrian National Library
The Royal Library of Denmark
State and University Library, Denmark
The National Archives of the Netherlands
The National Archives of England, Wales
and the United Kingdom
Swiss Federal Archives
University of Cologne
University of Freiburg
HATII at the University of Glasgow
Vienna University of Technology
Austrian Research Centers GmbH
IBM Netherlands
Microsoft Research Limited
Tessella Support Services Plc

For more information, visit www.planets-project.eu or email info@planets-projects.eu