

The DPC Award for Research and Development celebrates significant technical or intellectual accomplishments which lower the barriers to effective digital preservation. It is presented to the project, initiative or person that, in the eyes of the judges, has produced a tool, framework or idea that has (or will have) the greatest impact in ensuring our digital memory is available tomorrow.

Four finalists have been selected in this category:

**Collaboration in Data Management Planning: the UK's DMP Online and the US DMPTool, The Digital Curation Centre and partners**



Much research in the arts and sciences is funded from the public purse via grants. A lot of it produces – or relies upon the reuse of – digital data held in a wide variety of formats and in collections of varying size and complexity. Until recently, the long-term stewardship of digital data was largely overlooked by researchers, and important data with the potential to be repurposed or reused was managed in an ad hoc fashion. Data planning was an overlooked, orphan activity, with two key consequences: important and sometimes irreplaceable data stood at risk of being lost forever, and time and money were wasted recollecting existing data, or salvaging it from obsolete formats or storage media.

In many cases datasets were lost completely, not through deliberate neglect or lack of care, but because there was little incentive and no clear guidance on how to manage them, and consequently little planning at the critical early stages of a research project's lifecycle; once lost,

data concerning human and environmental events cannot be replaced.

The situation is now changing, with funding bodies increasingly requiring their grant-holders to create and maintain data management plans (DMPs) to enable future access and provide a solid base for future preservation activities. These plans typically justify the creation of a new dataset, outline how data will be created and managed, and assign clear roles and responsibilities across each stage of the data management lifecycle. However, there is a historical shortfall of understanding in the research sector on how best to care for this digital data, and where to turn for guidance and best practice exemplars during the planning process.

DMP Online is a web-based tool and suite of resources, representing the culmination of the DCC's engagement with the topic of data management planning, work which also covers an ongoing analysis of funder requirements, influential and much-reused checklist, 'How-To' guides and other forms of advice, consultancy reports, journal articles and a successful book chapter. It gathers together guidance from the memory community and from individual funders into a single location to provide a one-stop shop for data management planning.

New data requirements from US funders led to an interest in creating a tool similar to DMP Online, but with a focus on the needs of US researchers. The transatlantic DMPTool consortium formed as a result of conversations at the 2010 International Digital Curation Conference in Chicago, USA. Through these conversations, we came to realise that there were many shared factors in play between our two countries, as well a number of differences which would make a single, shared approach unlikely to succeed.

Our collaboration brought together librarians, IT experts and researchers, and spanned international activities from the very beginning. Recognising the extensive thought, broad consultation and shared effort that went into the development of the first version of DMP Online, the US team immediately sought international collaboration to leverage this existing work.

Each tool offers a variety of functions that have been designed with clarity and ease of use in mind. Features include:

- breadth in scope to plan the management of a very wide range of data types and volumes;
- incorporation of customisable reporting tools for communicating data management issues

to decision makers, thus facilitating awareness-raising of digital preservation in the upper strata of large institutions;

- assistance for users seeking to demonstrate a commitment to data management, enabling institutions to provide crucial assurances – increasingly important as legal and compliance issues continue to grow in prominence;
- inclusion of funder-specific guidance, enabling plans to be tailored to specific disciplinary needs in terms that researchers find familiar and comforting.

The design of these systems makes it possible for the research community to gain insight into the methods and practices of research data management across the entire lifecycle at both a micro and a macro level. They offer value to the individual researcher through a focused data management plan development workflow and just-in-time resource associations, while also offering high-level functionality that makes meta-analysis of data management planning practices across many domains a possibility. They offer the wider research community the opportunity to understand and refine practices for better integration of research data management processes, consequently enabling more interoperable and reusable data.

Shared ownership of the plans cements their place as communication devices between project partners, and the benefits of online hosting underpin our shared vision of DMPs as living, evolving documents which take into account changes in the research activity – and in best practice. Such flexibility is a critical requirement in long-term digital preservation. Similarly, at the macro level, the ability to collect and compare DMPs within institutions and across disciplines provides opportunities for capacity planning and related benefits that offline plans cannot readily achieve.

Since their respective launches in April 2010 and August 2011, DMP Online and DMPTool have been used to create more than 3700 plans, and uptake patterns have mirrored the journey from specialist need to wider use among our stakeholder communities. Between us, our work has attracted interest from the Southern Hemisphere as well as continental Europe, and conversations are underway as to how we can continue to foster this spirit of international cooperation.

The DMPTool project has also begun to include contributing partners beyond the initial group, starting with the Inter-University Consortium for Political and Social Research. More widely, the team has also received enthusiastic interest from a number of other institutions and government agencies, including the US National Science Foundation, the US Forest Service, the US Geological Survey, and the US National Oceanic and Atmospheric Administration, as well as several for-profit organisations that specialise in support of the research management process.

We continue to liaise with various stakeholders to take this forward as a shared, community-driven effort.

The ongoing intention is to continue adding contributing partners, conducting structured user testing on new subjects, and growing the community to ensure the broader sustainability of the service and more expansive integration of the system data and functionality with other research platforms and resources.

**Note - DCC (nominee), UK LOCKSS Alliance and HATII Glasgow University (proximity) and JISC (funder) may not vote for Data Management Planning Tools**

**PLANETS Preservation and Long-term Access through Networked Services, The Open Planets Foundation and partners**



The Planets Project set out to enable organisations including national scale libraries and archives to ensure the long-term preservation of their valued digital content. To accomplish this, it brought together a team of sixteen major libraries and archives, large technology companies, smaller enterprises, and leading university research teams.

As the project completed in 2010, it was able to deliver a fully integrated, open framework of tools and services to support every stage of the Digital Preservation Lifecycle in a way which would not only reflect the OAIS model, but also enabled third party archiving solutions to be able to connect simply and seamlessly to either the whole or individual parts of the Planets system.

The project contributed to advances in key digital preservation challenges including characterising digital items, planning to address the risks to digital content, using emulation technology, preserving databases, and providing a basis for evaluating the success of

preservation actions. It also helped to establish a conceptual framework for digital preservation as well as establish a testbed approach to conduct non-destructive repeatable experiments with either one's own digital objects or ones selected for their specific properties from reusable corpora.

The legacy of Planets continues today. The project established the Open Planets Foundation (OPF) to further develop the international technical and practitioner community in digital preservation. The OPF has become a key meeting place for deep discussion and problem solving as well as providing a way to sustain open source software contributions. Technologies and approaches developed in Planets have spurred national and international projects to carry them forward. Examples include the SCAPE project to build a scalable preservation environment, the KEEP project to further pursue emulation techniques. Future Hacka-thons were fostered through the existence of both the developer community and tool basis that had been developed in Planets. The conceptual framework has contributed to international standards such as PREMIS.

The methodological and technical developments enabled organisations such as the British and Dutch national libraries to improve their approach to ensuring long-term access to digital content.

Planets combined its own innovative developments with existing software and research. Planets created a suite of tools in the areas of Preservation Planning, the identification (Characterisation) of digital objects and the creation of a Preservation Testbed in which Preservation Tools can be tested by conducting non-destructive, repeatable experiments with either one's own digital objects or ones selected for their specific properties from Corpora of various formats.

Alongside Planets' own work, the project also integrated and supported the further development of tools developed by Planets partners, as well as 'wrapping' existing preservation tools to enable them to be invoked and managed automatically within sometimes-complex workflows.

At its completion, the Project had achieved its main objectives and established publicly accessible instances of its PLATO Preservation Planning and Testbed applications connected via the Planets Interoperability Framework to over 50 Preservation Services as well as to characterisation and format comparison services.

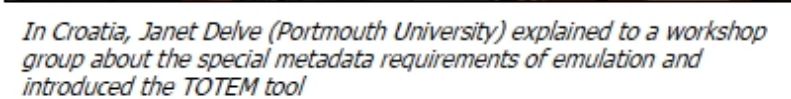
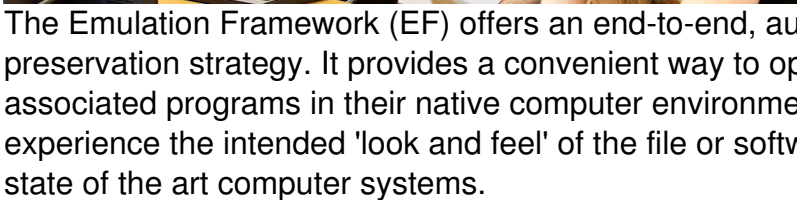
The Planets suite integrates with and adds capabilities to repository management systems such as Fedora, Rosetta, ePrints, or internally developed systems such as the British Library's DLS or the Dutch National Library's eDepot through a repository manager adaptor. A single interface is all that is required to connect any third-party system to all Planets tools and services, irrespective of location. Full specifications have been published to enable further interfaces to be developed, and experience gained in connecting Fedora suggests that less than 10 programmer-weeks' effort is required to complete this work.

Planets has raised awareness of issues surrounding Digital Preservation in the global archives and library community. Via structured training courses held across Europe, Planets provided practical training to 320 delegates and published audio-visual self-study material. The Planets newsletter has a regular readership of approximately 1,250 and 600+ people have registered to receive regular e-bulletins.

Planets also helped to raise awareness of digital preservation challenges more broadly. It commissioned, a series of awareness-raising short films published on YouTube, including 2 cartoons featuring DPE's 'Digiman' that continue to garner attention and have been seen thousands of times. The unique Digital Time Capsule campaign during which we deposited a digital time capsule in the ultra-secure Swiss Fort Knox data repository, reached an audience independently estimated to exceed 17.5 million readers, listeners and viewers around the world.

**Note - British Library (host) and British Library Preservation Advisory Centre (proximity) may not vote for PLANETS**

**TOTEM Trustworthy Online Technical Environment Metadata Registry, University of Portsmouth and partners**

[illegible]

To achieve this the EF combines existing emulation technology with a sophisticated workflow that automates steps of defining and configuring hardware and software dependencies. As a result, the end user doesn't need in-depth technical knowledge of the original computer system or software to be able to render the digital object.

Key functionality of the EF is the automated workflow for defining, configuring and rendering the emulated environment. The following steps are automated:

- identifying the type of digital file that the user has selected;
- finding the required software and computer platform for the file;
- matching these dependencies against the available software and emulators;
- configuring the emulator and preparing the software environment;
- injecting the digital file into the emulated environment;
- giving the user control over the emulated environment

The EF builds on existing work done in digital preservation: instead of re-inventing the wheel, the EF project team reused the current state of the art developments in emulation and file identification technology. For identification, the Harvard File Information Tool Set is used; instead of developing new emulators or modifying existing ones, the EF incorporates a number of open source emulators. Emulators such as QEMU and Dioscuri have a proven track record in reliably rendering specific computer environments, and have been selected to emulate x86 hardware. Other platforms included in the current release include the Commodore 64, Amiga, Amstrad CPC, BBC Micro and Thomson, but the spectrum of potential computer platforms and applications that can be supported is practically unlimited.

Another key feature of the EF is its architecture. The EF can be used as a standalone tool, but has been designed so that the solution can be scaled in an organisational setup. To achieve this, the EF has been split into three components which run independently of each other:

- EF engine (including graphical user interface)
- Emulator archive service
- Software archive service

While the EF engine runs locally, both the emulators and the required software are stored centrally in an emulator archive and a software archive, respectively. Both are designed to be contacted via web services allowing them to be used across the network within an organisation,



or even worldwide. This improves manageability of available software in the archives and lowers IT maintenance. The single archive locations mean that extending them with emulators and software provides immediate benefit for all EF users connected.

The EF is available as open source software under the Apache 2.0 license and comes with a point-and-click installer, a basic set of seven emulators and open source operating systems and application software. With this suite, users can immediately render more than 30 file formats, such as PDF, TXT, XML, JPG, TIFF, PNG, BMP, Quark, ARJ, EXE, disk/tape images and more.

The set of supported emulators, operating systems and application software can be extended using wizards for adding new emulators and software to the web services, further supporting any number of file formats.

The software is developed in Java with cross-platform compatibility in mind, and runs on all versions of Microsoft Windows, Mac OS and Linux that support the Java Runtime Environment version 1.6 or higher.

**□ Note - University of Portsmouth (proximity) □ may not vote for the Emulation Framework**