Preservation Health Check (PHC) – iPRES 2012 Workshop Report

The PHC-workshop was held during the PREMIS Implementation Fair 2012, at iPRES (October, Totonto).

Introduction to the pilot
Titia van der Werf introduced the Preservation Health Check (PHC) Pilot, which is a joint initiative of OCLC Research and the Open Planets Foundation (OPF). The main research questions being considered are: what does my preservation metadata tell me about digital preservation risks to which my digital collections are exposed? How effective are standards, tools and practices in addressing risk control? The pilot will collect preservation metadata recorded by repositories and analyze it as to risk assessment. This work will involve mapping the SPOT risk model with PREMIS and then apply this as a protocol to analyze real life metadata and report back results in the form of Health Check Reports. The pilot will involve 3 pilot sites (not all 3 have been identified yet) and the duration of the pilot is set to 1 year. More information is available at: http://www.oclc.org/research/activities/phc.html

Standards, Risk analysis, Common sense and Evidence
Bram van der Werf explained that the use of the term "risk assessment" in the context of digital preservation is confusing and that what the DP-community really talks about is "quality assurance": the ongoing monitoring and improvement of the preservation process itself, which is all about treating and handling digital objects to stop or slow down their deterioration, obsolescence and loss of quality (renderability, functionality, etc.). Quality assurance requires continuous monitoring (PHC-dashboard concept) as opposed to one-off risk assessments or certifications. The PHC-dashboard visualizes the data gathered by sensors and triggers preventive/corrective actions. The pilot is looking at PREMIS to identify such sensors. For tool-building, the PREMIS standardization process should not be a moving target.

Presentation by the Bibliothèque Nationale de France (BnF), a prospective PHC Pilot Site
Sébastien Peyrard shortly presented BnF's expectations of the PHC-pilot and why it is interested in becoming a pilot site. He explained how the need for regular monitoring (the last assessment they did was in 2007) was becoming more urgent, in particular to evaluate the quality and usefulness of the metadata and how it might be improved. He suggested that external evaluation might trigger internal improvements and be helpful in securing activities and staff effort to improve the preservation metadata. He envisioned how the PHC would help to formulate functional requirements for preservation metadata and how the dashboard would improve the curator's data management interface.

Presenting the SPOT Model
Priscilla Caplan gave a short introduction to the SPOT Model which is a lightweight framework for addressing threats to the digital content of repositories. It explicitly does not address other threats found in more encompassing risk-assessment frameworks, such as governance, financial or legal threats to the sustainability of repositories. The threats identified in the SPOT Model have been carefully chosen on the basis of criteria such as appropriate and comparable levels of granularity and comprehensiveness. The SPOT approach was guided by the desire for practical application in the field. The Model is published in: http://www.dlib.org/dlib/september12/vermaaten/09vermaaten.html

Preservation metadata as an evidence base for risk assessment
Brian Lavoie explained how mapping SPOT to PREMIS might lead to identifying preservation metadata that supports evidence-based risk assessment. The mapping exercise would lead to a protocol, ideally a standardized and widely-applicable protocol for threat-control, addressing both types of threats: those that have already happened and those that can potentially happen. The focus of the mapping will be on actionable intelligence and automated analysis. Brian gave some practical examples and concluded that the mapping exercise could highlight gaps and inform repositories (the gap between necessary metadata and metadata actually recorded) and advance the further development of threat models and of the PREMIS Data Dictionary.

Plenary Discussion
The presentations led to practical questions and several more elaborate discussion threads.

The relationship with other threat models: TRAC and DRAMBORA are pretty high-level and they are not linked to evidence information that is maintained in the repository. The community needs regular assessments and the potential for automating these and making it a routine. The PHC-pilot is an appealing project because it is trying to do this. It would be useful to position the PHC-exercise in the bigger picture of risk assessment and certification.

The impact question: it would be interesting to understand how the PHC-results would impact our organisation and operations. How would this inform preservation managers in terms of preservation strategies or facilitate the process of becoming a trusted repository? Will the pilot provide input to justify the development of health check tools? How would these fit in existing tools and workflows? Where in the metadata cycle would this health check take place?

The relationship with preservation policies: repository mission and policies are essential for establishing context for risk-assessment. What is the relationship between policies, preservation metadata and risk factors? How will the SPOT-PREMIS mapping take policies into account?

Digital lifecycle and control over metadata: repository managers do not necessarily have full-control over the preservation metadata they maintain. Metadata might be produced by other stakeholders in the digital lifecycle. The PHC-pilot might need to take the digital life-cycle into account when mapping the metadata and the threats.

Wrap-Up
Titia wrapped up the session by thanking the participants for their active and useful contributions. She stressed the fact that the PHC-pilot was only in its very early stages of defining the research methodology and identifying pilot sites. She encouraged those interested in participating to contact Bram or herself. The community will be able to follow progress via the various communication channels of OCLC Research and OPF.