***Name of Proposed Interest Group: Preservation Tools, Techniques and Policies***

***(Note pending name change to: REsearcher’s Tools for AttaIning Knowledge preservatioN  (RETAIN))***

***Introduction*** *(A brief articulation of what issues the IG will address, how this IG is aligned with the RDA mission, and how this IG would be a value-added contribution to the RDA community):*

The Preservation Tools, Techniques, and Policies (PTTP) IG provides a forum to bring together domain researchers, data and informatics experts, to discuss such issues as:

* What data/software/artifacts/documentation (hereafter referred to as “knowledge products”) should be preserved for sharing, re-use, and reproducibility for a given research domain? For other domains?
* What tools are available for researchers to preserve these elements in a manner that does not obstruct or hinder their research?
	+ What are the strengths and weaknesses of these tools?
	+ Are there common features that could allow tools from one domain to be re-used elsewhere?
	+ Are there tools that archives/repositories could provide that could make preservation much easier for researchers?
	+ What are the longer-term development goals of each of these tools?

Through the course of these discussions, the PTTP IG acts to strengthen the dialogue between domain researchers and the data community by focusing on how researchers are enabled to use previously generated and preserve new results. This enhanced engagement amplifies the voice of the research community within the fabric of RDA.

Given that one must preserve knowledge products before one can (usefully) share them, the mechanisms by which this preservation happens is primarily in the hands of the researcher and should be a critically important element to the mission of the RDA. The quality of the data and the information relevant to their creation can only be guaranteed by the researcher who produces the data. Thus, it is in the RDA’s best interest to consider this an integral part of its progress.

This group has obvious synergies with the Reproducibility IG, the Provenance IG, the Active Data Management Plans IG, the Archives and Records Professionals for Research Data IG, the Virtual Research Environment IG, and the Preservation e-Infrastructure IG, as well as the domain-specific IGs. It has no overlap and is entirely complementary in charter/focus with the existing Preservation e-Infrastructure IG.

***User scenario(s) or use case(s) the IG wishes to address*** *(what triggered the desire for this IG in the first place):*

Largely absent from formal RDA deliberations thus far have been discussions of *how* researchers can interact with repositories in order to preserve their findings. Most researchers do not consider preservation as part of their research workflow, and, when confronted with an unfamiliar repository interface for data ingestion, are unable to provide the information required. They do not, as a matter of course, use tools that allow the automatic generation of the necessary metadata and other information that is necessary for preserving the knowledge behind their research results. In fact, for many researchers, the situation is represented by the (re-purposed) familiar cartoon, below:



The PTTP IG will operate in the “overlap space” that is (somewhat unfairly) represented as empty in the above figure. The IG will explore how knowledge preservation is currently being done, what tools exist, what are their strengths and shortfalls, and how policy and other considerations are (if at all) driving preservation strategies, preservation tool development, and preservation tool adoption. These discussions are quite timely given the impending implementation of “open data” policies from all US funding agencies and the corresponding move in the EU in this same direction. The knowledge preservation tools for most researchers are either inadequate or woefully under-adopted. RDA is the only global forum that currently provides an opportunity for discussions of these issues by the primary stakeholders. By encouraging increased participation by domain researchers in these important discussions, the PTTP IG can make significant contributions to solving one of the most important issues around data and research.

***Objectives*** *(A specific set of focus areas for discussion, including use cases that pointed to the need for the IG in the first place.   Articulate how this group is different from other current activities inside or outside of RDA.):*

Following from the issues listed above, the PTTP IG will:

* Catalogue available preservation tools, including capabilities, compatibilities, and rates of adoption, and make this information available to researchers and archivists. This catalogue will serve as a basis for discussion of tool development and deployment in order to better meet the needs of diverse research communities.
* Undertake outreach activities, including holding sessions at RDA plenaries and conducting outside workshops to engage researchers and archivists in preservation tool specification and, potentially, adoption.
* Engage related RDA IGs: Metadata, Provenance, Reproducibility, Active Data Management Plans, Archives and Records Professionals for Research Data, Preservation e-Infrastructure; and domain-specific IGs in a wider dialogue around preservation needs, tools, and the specifications thereof.
* Survey preservation policies across countries, funding agencies, and research areas to assemble a comprehensive view of researcher and archive responsibilities
* Engage representatives of funding agencies, either at RDA plenaries or at other workshops, in order to involve them in the details of these discussions.

***Participation*** *(Address which communities will be involved, what skills or knowledge should they have, and how will you engage these communities.  Also address how this group proposes to coordinate its activity with relevant related groups.):*

The PTTP IG seeks to connect domain researchers with data scientists and data-handling professionals in order to improve, first, the communication between them, and second and more important, the tools the researchers use to preserve the knowledge inherent in their research results. If this is taken as a primary goal, the communities involved are quite broad: all researchers on one hand, and all those whose goal is to provide the platforms for preservation and access on the other. This second community is currently the main constituents of RDA, and they are quite engaged already. Emissaries of the research community are also enthusiastic RDA members. The PTTP IG will prevail upon this smaller cadre for outreach (or in-reach) to other researchers who should be involved in the discussions. Several means will be exploited to achieve additional in-reach or outreach to broader communities, including attendance at scientific society meetings or at domain-specific conferences. Additional, smaller workshops outside of RDA plenaries may be a more targeted way to attract additional participation and dialogue. Extensive networks of interested researchers exist; the task is to bring more of them into the conversation.

In terms of coordination, members of the proposed IG have already coordinated two joint sessions at the 8th RDA Plenary in Denver with two of the related groups, the Reproducibility IG and the Active Data Management Plans IG. IG leaders from the Provenance IG and several of the discipline-specific groups were also attendees at those sessions. At the 10th Plenary in Montreal, joint discussions were held in our IG meeting and in the meeting of the Archives and Records Professionals for Research Data IG, which is very complementary in focus. Thus, coordination is already occurring. Maintaining open lines of communication and combining meeting sessions when appropriate should be relatively straightforward.

***Outcomes*** *(Discuss what the IG intends to accomplish.  Include examples of WG topics or supporting IG-level outputs that might lead to WGs later on.):*

A. As mentioned above, the PTTP IG intends to produce a catalogue/review:

1. A taxonomy of preservation tools, including their features, strengths and weaknesses, and their rates of adoption in various research domains

B. A primary outcome of this IG will be better communication between the data scientist/archivist realms and that of the domain researcher

C. A primary outcome of this IG will be wider adoption of preservation tools by domain researchers

D. A primary outcome of this IG will be the delineation of desirable characteristics and features of scientific preservation tools for future tool development

A potential WG project would be to choose a pilot research domain with no good preservation tools and to solve that particular problem in a manner that isn’t completely domain-specific.

***Mechanism*** *(Describe how often your group will meet and how will you maintain momentum between Plenaries.):*

The PTTP IG will have meetings every month to six weeks, partially reflecting on issues raised by previous plenary meetings, and planning for the next round of plenary sessions. Significant attention will also be devoted to accomplishing the goals of the IG, namely greater researcher participation and the development of the proposed catalogues of tools and policies. Planning of additional workshops, when appropriate, will certainly keep the level of engagement high.

***Timeline*** *(Describe draft milestones and goals for the first 12 months):*

Goals:

1. Organize at least one breakout session at the 11th RDA Plenary in Berlin, including researchers who have not previously attended RDA.
	1. Timeline: months
2. Compile preliminary list of available preservation tools
	1. Timeline: 3 months
3. Compile preliminary tool taxonomies (features, strengths, weaknesses, adoption)
	1. Timeline: available for 12th Plenary (12 months)

***Potential Group Members*** *(Include proposed chairs/initial leadership and all members who have expressed interest):*

Proposed Chairs: Ruth Duerr, Michael Hildreth, + someone from outside the US

|  |  |  |
| --- | --- | --- |
| First Name | Last Name | email |
| Kana  | Asano  | k2asano@jst.go.jp |
| Kevin  | Ashley  | kevin.ashley@ed.ac.uk |
| Donatella  | Castelli  | donatella.castelli@isti.cnr.it |
| Harrison  | Dekker  | hdekker@berkeley.edu |
| Tim  | DiLauro  | timmo@jhu.edu |
| Devan  | Donaldson | drdonald@indiana.edu |
| Dave  | Dubin | ddubin@illinois.edu |
| Ruth  | Duerr  | Ruth.duerr@ronininstitute.org |
| Kim  | Edwards | kindrew@jlab.org  |
| Dom  | Fripp  | dom.fripp@jisc.ac.uk |
| Matt  | Fry  | mfry@ceh.ac.uk |
| Yolanda  | Gil | gil@isi.edu |
| Helen  | Glaves | hmg@bgs.ac.uk |
| Siddeswara  | Guru | s.guru@uq.edu.au |
| Joe  | Hand  | joe@hand.email |
| Kari  | Heffner  | heffner@jlab.org |
| John  | Helliwell  | john.helliwell@manchester.ac.uk |
| Michael | Hildreth | hildreth.2@nd.edu |
| Rob  | Hooft  | rob.hooft@dtls.nl |
| Cynthia  | Hudson-Vitale | chudson@wustl.edu |
| Yanni  | Ionnidis  | yannis@di.uoa.gr |
| Reyna  | Jenkyns  | reyna@uvic.ca |
| Dessi  | Kirilova | Dessi.Kirilova@syr.edu |
| Jens  | Klump  | jens.klump@csiro.au |
| Yusuke  | Komiyama  | komiyama@nii.ac.jp |
| Rebecca  | Koskela | rkoskela@unm.edu |
| Inna  | Kouper | inkouper@indiana.edu |
| Chris  | Lenhardt  | clenhardt@renci.org |
| Mary  | Leverance | mleveran@uark.edu |
| Ye  | Li  | yeli@mines.edu |
| Tobin  | Magle  | tobin.magle@colostate.edu |
| Brian  | Matthews | brian.matthews@stfc.ac.uk |
| Rudolf  | Mayer  | rmayer@sba-research.org |
| Eva  | Mendez  | emendez@bib.uc3m.es |
| Tomasz  | Miksa  | tmiksa@sba-research.org |
| Reagan  | Moore | rwmoore@renci.org |
| Kieran  | Murphy | kieran.murphy@ecodev.vic.gov.au |
| Linda  | Naughton | linda.naughton@jisc.ac.uk |
| Yasushi  | Ogasaka  | ogasaka@jst.go.jp |
| Ray  | Osborn | rosborn@anl.gov |
| Ellas  | Papadopoulou | Ellas.Papadopoulou@ec.europa.eu |
| Hans  | Pfeiffenberger  | hans.pfeiffenberger@awi.de |
| Line  | Pouchard  | pouchard@purdue.edu |
| Andreas  | Rauber  | rauber@ifs.tuwien.ac.at |
| Terrell | Russell | tgr@renci.org |
| Andre  | Schaaff  | andre.schaaff@astro.unistra.fr |
| Frank  | Schluenzen | frank.schluenzen@desy.de |
| Soren  | Scott  | sorenscott@gmail.com |
| Natasha  | Simons  | Natasha.simons@ands.org.au |
| Victoria | Stodden | vcs@stodden.net  |
| Aaron  | Sweeney  | aaron.sweeney@colorado.edu |
| Kosuke  | Tanabe  | TANABE.Kosuke@nims.go.jp |
| Ville  | Tenhunen  | ville.tenhunen@helsinki.fi |
| Erica  | Veil | ericaveil@ericaveil.com |
| Joachim  | Wackerow | joachim.wackerow@gesis.org |
| Simon  | Waddington  | simon.waddington@kcl.ac.uk |
| John  | Watkins  | jww@ceh.ac.uk |
| Malcolm  | Wolski | m.wolski@griffith.edu.au |
| Lesley  | Wyborn  | lesley.wyborn@anu.edu.au |
| Carlo  | Zwölf  | carlo-maria.zwolf@obspm.fr |

+ others who have attended the last sessions at Barcelona and Montreal.