

# RDA IG Preservation Techniques, Tools and Policies

## Preserving Scientific Annotation Working Group (PSA-WG)

### Case Statement

---

#### 1. WG Charter

The Preserving Scientific Annotation RDA Working Group (PSA-WG) will precipitate adoption of reliable standards-based preservation *solutions* for both newly-created research contributions which employ annotation of data and documents, and also those redelivered from existing research investments. Annotation of digital resources has emerged as a new research paradigm, extending across scientific domains and offering significant opportunities for improving discovery and preservation of research investment compared with existing workflows and conventional publication processes.

However, realizing long-term benefits from annotation methodologies has remained elusive, due to continuing evolution of both the instruments<sup>1</sup> which create annotations and infrastructures to store them. This has prevented development and operation of stable annotation workflows, and poor awareness of preservation vulnerabilities has led to loss of research investment or the need for costly redelivery to overcome cyclic technology obsolescence. However, this situation can now be remedied, in particular through use of PID strategies<sup>2</sup> for reliably connecting annotation lists, contributing scientists and digital resources in the long-term.

#### 1.1 Multiple Annotation Techniques

Annotation of born-digital texts and datasets, as well as annotations on digitized physical and media objects, can be traced from the ideas of Vannevar Bush in the 1940s and demonstrations at the SRI Augmentation Laboratory<sup>3</sup> founded by Douglas Englebart in the 1960s. Widespread use of Bill Atkinson's HyperCard and its later development by Apple Computer in the late 1980s led to annotation applications linking digital media<sup>4</sup>, and with rapid uptake of the World Wide Web after 1990 these began using HTML. However, it wasn't until 2009 that accessible instruments for creation of application-independent 'stand-off' annotations (represented, for example, in XML) for digitized

---

<sup>1</sup> We use the term 'instrument' throughout this document in the sense of Virtual Research Instrument, rather than implying any specific physical apparatus

<sup>2</sup> <https://www.clarin.eu/content/comparison-pid-systems>, <https://office.clarin.eu/pp/D2R-2b.pdf> "... each individual resource (even an annotation) needs to be referenced, so that we can expect a huge number of PIDs."

<sup>3</sup> <https://www.sri.com/blog/future-augmentation>

<sup>4</sup> <http://www.uni-lueneburg.de/hyperimage/hyperimage/ebsKart.htm>

objects appeared, creating a new mode of research which has entered mainstream practice. New research infrastructures precipitated by developments such as the International Image Interoperability Framework<sup>5</sup> (IIIF) in late 2011 and the emergence in 2012 of nascent standards for representing annotations in JSON and potentially other serializations, have significantly accelerated adoption of annotation in fields such as medical imaging and the social sciences and humanities. IIIF breaches 'silos' of digital resources previously confined by different, incompatible technologies and instead enables delivery of assets from multiple organizations into research workflows using a single consistent interface. Although initially a framework for image interchange, IIIF specifically promotes stand-off annotation—partly because of its close organizational links with the Web Annotation Data Model<sup>6</sup> (WADM) editors. IIIF is expanding to address other media types and multiple projects are already using WADM for annotating 3D objects and time-based media such as movies<sup>7</sup>.

The stand-off style of annotation has advanced, in contrast to the alternative of 'embedding' annotations with digital assets, which requires integration of annotation data with the specific digital representation employed to describe the object. Modification of different digital media formats in this way, and subsequent versioning of assets is generally impractical: for example multiple annotations created by different contributors lead to scalability problems with methods to retrieve and maintain embedded annotations. A particular limitation of embedded annotation approaches is the inability to effectively annotate across multiple objects or distributed corpora, since they are necessarily tied to a particular location. This seriously limits the ability to use annotation to construct narratives or flows through data.

## 1.2 Working Group Focus

PSA-WG will focus initially on stand-off annotation which can be represented using the digital asset-independent WADM from W3C. It will not pursue preservation of embedded annotations except through conversion to stand-off. Although the WADM specification is not currently supported directly by most research instruments, those already in use do generate annotations represented in JSON using various versions of the predecessor to WADM, Open Annotation Data Model<sup>8</sup> (OADM), and this can be converted into WADM. In addition, the need to transform such existing annotations as the WADM specification evolves is also recognized, and can be automated. Moreover, through WADM's incorporation of a linked data approach to classifying annotations, the groundwork has been established for effective re-use of such research. As a result, millions of OADM annotations have been generated since 2016 and there is currently an explosion in the volume of new annotations being created. These developments have provided some important components for infrastructures permitting long-term protection of research investment employing annotation. However other key building blocks, particularly discovery and persistent storage for annotation, remain to be addressed before effective solutions can be provided to guarantee the preservation of both annotations and the resources they target. Consequently, research investment that relies on annotation is currently vulnerable.

---

<sup>5</sup> <https://iiif.io/>

<sup>6</sup> <https://www.w3.org/TR/annotation-model/>

<sup>7</sup> DOI: 10.1109/VSM.2017.8346274 [https://www.researchgate.net/publication/324785534\\_I-media-cities\\_a\\_searchable\\_platform\\_on\\_moving\\_images\\_with\\_automatic\\_and\\_manual\\_annotations/download](https://www.researchgate.net/publication/324785534_I-media-cities_a_searchable_platform_on_moving_images_with_automatic_and_manual_annotations/download)

<sup>8</sup> <http://www.openannotation.org/spec/core/>

There are also stand-off annotation techniques and activities which PSA-WG will not address at this time. For example [hypothes.is](https://web.hypothes.is/)<sup>9</sup> an open source project enabling annotation of web resources, which continues an approach advanced by Google before 2010 (Sidewiki<sup>10</sup>). [hypothes.is](https://web.hypothes.is/) addresses annotation of text on webpages, rather than elements of independent digital assets—for example features depicted in digital images themselves, or on solid models or in movie frames—and as soon as the webpage is altered the annotations are lost. Because they do not select targets on digital resources which can be maintained independently of the evolving internet environment we do not address this family of technologies here, although it could be an important future activity of the WG to do so. [hypothes.is](https://web.hypothes.is/) currently supports Google Chrome and has announced plans for a similar Firefox development. Separately, Apache Annotator <sup>11</sup>, is an ASF Incubator project supporting ‘Web annotation in Web browsers, Web publication readers, and the servers that serve them’, which has set out a broader roadmap of annotator.js-based projects and plugins.

### 1.3 Summary of Outcomes

Anticipated impacts of the WG are discussed in Section 2 and in Section 4 the deliverables are described. The task of the PSA-WG is to tackle preservation of annotation in a substantive manner and its anticipated outcomes can be summarized as follows:

- communicating vulnerabilities: a campaign to raise awareness of preservation risks and halt the continuing loss of investment in research which employs annotation
- overcoming roadblocks to creation of end-to-end solutions for preserving stand-off annotation: identifying essential but currently incomplete preservation mechanisms and precipitating the implementation of new functionality by repositories and annotation instruments to fill these vacua
- developing preservation use-cases in collaboration with partner organizations in multiple domains of research activity, which demonstrate exemplary strategies for preserving annotation and communicating these effectively to the broader community
- influencing on-going standards developments to ensure robust and efficient preservation solutions for the long term and delivering new benefits from stand-off annotation such as annotation store-based discovery services, which cannot be realized until preservation issues have been resolved

Short-term priorities will include, for example, PSA-WG working with ORCID<sup>12</sup> to implement an Annotation Work Type marker for contributor attribution and working with CERN to implement a Zenodo<sup>13</sup> Annotation Collection data resource type. These activities will produce a core identifier scheme—ORCID/URN/Zenodo-DOI—which will be evaluated in pilot projects and incorporated in a Technical Report within the planned 18-month activities of the WG. The WG will later pursue the

---

<sup>9</sup> <https://web.hypothes.is/>

<sup>10</sup> <https://chrome.googleblog.com/2009/10/bringing-google-sidewiki-goodness-to.html>

<sup>11</sup> <https://annotator.apache.org/>

<sup>12</sup> Activity commenced at PIDapalooza 2019

<sup>13</sup> <https://www.openaire.eu/zenodo-is-launched>

preservation of relationships between annotations and the digital resources they target using multiple identifier schemes, but discussions originating at PIDapalooza 2019 have led to an ORCID/URN/Zenodo-DOI implementation proposal which could be evaluated in pilot projects and subsequently delivered within the planned 18-month activities of the WG. The WG will later consider an instrument-agnostic OADM implementation guideline to promote display and maintenance of annotations multiple instruments and improve migration to WADM.

---

## 2. Value Proposition

### 2.1 Overcoming Roadblocks to Preservation of WADM Annotations

Stand-off scientific annotations using WADM are packaged as discrete digital entities—residing in databases, cloud services, etc—implicitly related to a file in a digital resource: the annotated asset. Annotations comprise 'bodies' of information associated with 'targets' defined on the asset—the latter remains unmodified. The target definition is stored in the annotation, and uses 'selectors' depending on the asset type to identify the feature which the annotation body refers to. For example a movie clip or just one frame can be defined by a 'time code' referred to the end of the leader and start of discrete exposed frames (the 'in point'), plus an 'out point'. Media types determine how such assets can be targeted—a PDF produced by a scanner without embedded OCR can't be targeted with a text selector in the same way that a born-digital PDF or Word file might be. However, the frames of a movie or the page boundaries of a document containing a snippet of text nevertheless 'anchor' annotations to the respective assets in one-way relationships constituted by 'target selectors'.

With stand-off annotation these relationships extend no further than the contents of the annotations themselves, and unless precautions are taken to insure that the digital assets which they target can be located in the long-term then they become vulnerable. Management of research data constituting annotations is subject to local IT infrastructure policies and invariably different to the policies and patterns of investment affecting sustainability of the various digital resources which they target. Consequently, any changes affecting either infrastructure which compromise unique identification of digital assets, using annotations' targets selectors alone, renders investment in those annotations useless. This situation is complicated further by versioning and the existence of multiple representation formats for digital assets; for example images may be compressed using JPEG for internet use, compared with lossless representations produced at the point of digitization. Annotation target selectors will, in general, be valid for only one version of a digital asset file, even if variants have the same resolution, so vulnerabilities currently arise when certainty of identifying the correct asset file using only the information in an annotation target selector cannot be maintained over time.

Moreover, annotations created using one instrument cannot currently be presented using another, without conversion and potential loss of information.

Lastly, although evolving annotation standards make provision for identification of creating agencies—whether directly by a researcher or indirectly by software algorithm—none of the existing annotation instruments yet support ascription mechanisms, such as ORCID<sup>14</sup> contributor identifiers.

---

<sup>14</sup> <https://orcid.org/>

## 2.2 Benefits for the Research Community

In summary, PSA-WG will address principal roadblocks to preservation of WADM annotation through the following activities:

- evaluating and producing recommendations on the use of persistent identifiers for annotated resources, to ensure long-term resolvability of annotation targets—additionally, to address evolving resources so that annotations continue to reference the correct version
- recommendations for the use of WADM (and OADM in the short term) so that annotation tools can interoperate and share annotations without conversion
- develop attribution and credit mechanisms to ensure that annotations can be properly treated as scholarly activity—additionally, to provide aggregation mechanisms so that scholarly contributions can be managed at effective levels of granularity

Success with these developments would also enable the WG to contribute to formalizing ‘research activity’ identification<sup>15</sup>, and contributing significantly to better discovery of research investment employing annotation. Research activities could be connected definitively to Annotation Collections<sup>16</sup>—themselves comprising annotation lists containing both identifiers of the resource targeted and institutional or original contributor identifiers. Such connectivity would allow contributors as well as annotated digital resources to be identified by research activities with fine granularity, as well as permitting automated verification of continued accessibility of digital resources in the long-term.

## 2.3 Key Impacts

The volume of existing research investment incorporating annotation which is vulnerable, and strategies for its protection are assessed in Section 2.4. However, it is evident that the lack of robust standards-based infrastructure for preservation of annotation, as of the date of this document, is the most pressing concern. Without this, the probability of saving historic investment and mitigating the significant costs of redelivering research which is currently in progress or being planned are low. The roadmap set out in Section 4 to deliver annotation preservation solutions to the research community within the 18-month program of PSA-WG is not unrealistic, since new technologies will not have to be developed: effective persistent identifier and long-term repository components have been in use for several years. To provide preservation infrastructures for annotation, these components have to be assembled into practical solutions, and this could be achieved by the WG through selection from available components, publication of recommendations and detailed use cases. However, effective communication of vulnerabilities as well as recommendations and development of early adoption projects in specific sectors of the research community will be a WG priority. Key impacts of these activities can be summarized as follows:

- sharp reduction of loss in research investment using annotation currently being planned, through effective communication of vulnerabilities to the research community

---

<sup>15</sup> e.g. <https://pub.uni-bielefeld.de/record/1972842>, <https://www.raid.org.au/>, <http://www.researchobject.org/>

<sup>16</sup> see section 5.0 of W3C Recommendation 23 February 2017, <https://www.w3.org/TR/annotation-model/>

- overcoming roadblocks to creating workable solutions for long-term preservation of annotation using available technologies where possible
- creation and communication of guidelines and use cases leading to uptake of preservation solutions for annotation and improved rates of recovery of historic material at risk of loss
- elimination of cyclic costs of redelivery or loss of investment in the future arising from lack or failure of preservation strategies for annotation
- create new potential for FAIR digital resources incorporating annotation

## 2.4 Scale of Research Investment using Annotation

Stand-off annotation is already a mainstream mode of research—increasingly, scholarly information is being made available as annotations and comments created as part of the discursive process, rather than via conventional publication. Many of the assertions so made are individually not significant enough to warrant publication unless combined with many others and rewritten to form an overarching narrative. This may not happen immediately, and annotations often remain the only instances of such contribution for long periods. However, in the long run, the information content of accumulated annotations often constitutes a significant online resource. For example, Early Modern Letters Online<sup>17</sup> is primarily a catalogue resource but contains over 50,000 comments made by historians and literary scholars that greatly enhances its utility.

Individual research endeavors organized around annotation are also multiplying, as standards for defining annotations emerge and new instruments for creating them precipitate innovative methods and inflect research agendas. This evolution is distributed across research domains, from the natural and life sciences to the social sciences and humanities. For example, one of the long-standing editors of the current WADM annotations standards group is based at Massachusetts General Hospital<sup>18</sup>; the Digital Imaging and Communications in Medicine (DICOM<sup>19</sup>) group represents hundreds of institutions and medical equipment manufacturers, some of which are in the process of adopting WADM. Heidelberg's Excellence Cluster for Transcultural Studies<sup>20</sup> produced less than 100,000 annotations between 2010 and 2015 across multiple projects, whereas a single contemporary research activity at Europa Institute Basel has already produced more than 900,000.

Annotation investment by the Heidelberg Cluster would have been lost in 2017 when personnel supporting its software infrastructure were reassigned. Urgent data forensic work on several projects led to creation of Invenio repositories and IIF services and redelivery of tens of thousands of complex annotations as OADM targeting these new resources. However, mechanisms to insure connection between annotations and these resources remain to be finalized. The current Europa Institute project has developed its own standards-based repository infrastructure to be able to guarantee long-term

---

<sup>17</sup> [emlo.bodleian.ox.ac.uk](http://emlo.bodleian.ox.ac.uk)

<sup>18</sup> Paolo Ciccarese, Massachusetts General Hospital <https://www.w3.org/TR/annotation-model/>

<sup>19</sup> <https://www.nema.org/Standards/Pages/Digital-Imaging-and-Communications-in-Medicine.aspx>

<sup>20</sup> <http://www.asia-europe.uni-heidelberg.de/en/hcts.html>

accessibility of its outputs. These are high cost activities, for which support through conventional applications for research funding would be unlikely.

---

### 3. Engagement with Existing Work and Adoption Plan

PSA-WG will engage with existing data preservation work and with the research community in three distinct activities. First it will work with identifier management authorities and repository services to develop identifier schemes for preservation of annotation data, including organizations developing ‘research activity’ identification mechanisms (as already discussed in Section 2.2). Secondly, it will develop partnerships with multiple research communities to evaluate effective identifier schemes for the specific digital resources and annotation instruments which they employ, leading to publication of use-cases and recommendations. Thirdly, PSA-WG will also work with groups engaged in on-going development of standards relating to preservation of annotation, including:

- developers of instruments supporting display, maintenance and creation of annotation
- standards developers contributing both to convergence of existing usage of annotation standards and their future evolution

Additionally, PSA-WG will engage with RDA ESIP/RDA Earth, Space, and Environmental Sciences IG, RDA PID Kernel Information WG and W3C WADM early in its activities in order to consult over development of pilot projects, use-cases and standards.

#### 3.1 Collaboration with Identifier Authority and Repository Organizations

Section 4 addresses outreach activity work-packages, and describes the WG’s strategy of creating and testing a *core identifier scheme* for preservation of annotation as the basis for development of schemes tailored to the needs of specific research communities.

This approach builds on preliminary work after RDA Plenary-11 on preservation of annotation pilot projects conducted by The Bodleian Libraries, CERN Repositories Section and Data Futures LBG. These activities redelivered vulnerable research in the humanities employing complex annotation; transforming annotation in legacy formats into OADM targeting new IIF digital resources. However they did not implement a persistent identification scheme to connect the annotations and targeted resources. Since RDA Plenary-12 PSA-WG has worked with Zenodo and ORCID, as summarized in Section 1.3, to implement support for annotation-based research activities, leading to implementation of an Annotation Work Type marker for contributor attribution and a Zenodo Annotation Collection data resource type.

In the first months of its work-plan after being formally established, PSA-WG will implement an ORCID/URN/Zenodo-DOI annotation identifier scheme for the previously redelivered humanities corpora and additionally implement this scheme in a new humanities research project (see WP2). However the WG recognizes that IIF digital resources (and consistent implementation of URN) and

the Mirador<sup>21</sup>-flavored OADM employed in these projects is not representative of preservation requirements of other research communities. PSA-WG will work with partners in the life and physical science communities (current discussions summarized in Section 3.2) who are already planning development of annotation workflows, in order to develop effective identifier schemes for those domains. This will require collaboration with multiple digital resource PID authorities, research activity developers, such as Research Object Consortium, and with repository services and developers. Accordingly, Bodleian, CERN and Data Futures will pursue discussions with a new tier of organizations, such as California Digital Library and Duraspace from Plenary-13 onwards.

### 3.2 Engagement with Research Activities in the Life and Physical Sciences

Multiple attendees of the PSA-WG Birds-of-a-Feather meeting at Plenary-12 expressed interest in developing preservation of annotation pilot projects in disciplines other than the humanities. As a result discussions are now in progress with the Earth sciences, bioinformatics and medical imaging communities. Preliminary evaluations using GeoTIFF and DICOM datasets were implemented in early 2019; correspondence between PSA-WG and ELIXIR led to a meeting at PIDapalooza and project meetings scheduled at Plenary-13; multiple internet meetings have led to ESIP joining PSA-WG and discussions are now scheduled at Plenary-13 to develop collaborative project roadmaps. Additionally, pilot projects in science and technology have commenced before Plenary-13 (see Section 5) and will deploy the same ORCID/URN/Zenodo-DOI annotation identifier scheme already developed in PSA-WG projects for Plenary-12—forming the first tier of adopters.

### 3.3 Input to Standards Activities

Preliminary PSA-WG partners The Bodleian Libraries, CERN Repositories Section and Data Futures LBG have been developing discussions with the standards activities of multiple communities including IIF and W3C's WADM since 2017. The Bodleian Libraries was instrumental in establishment of IIF and Data Futures was a founder member. CERN is a founding OpenAIRE partner. Separately, Data Futures established some of the first workflows employing IIF-Mirador to be deployed in high volume, and has already worked with both community groups as well as WADM to resolve roadblocks and contribute to new functionality. Significantly, the Chair of IIF's 3D Community Group has joined PSA-WG membership and has commenced a pilot annotation project using the WG's pilot infrastructure. The Chair of the Universal Viewer Community (UV) Group has joined PSA-WG, and preliminary discussions about display of existing annotations using UV and development of a roadmap for creation and maintenance of annotations in UV will continue after Plenary-13.

Early drafts of PSA-WG's Technical Report will be provided to the IIF and WADM editors and also circulated to the Mirador and UV developers to precipitate consultation before publication. Separately, existing collaboration between PSA-WG and ORCID and Zenodo will continue, leading to development of preservation of annotation documentation for those platforms. Finally, it is expected that consultation with another tier of persistent identifier authorities and repository developers will lead to support for annotation work and data types mirroring what has already been achieved with ORCID and Zenodo through implementation of extensions to those technologies.

---

<sup>21</sup> <http://projectmirador.org/>

### 3.4 Plan for Adoption

Two phases of adoption of PSA-WG recommendations by the community are envisaged. It has been demonstrated that collaboration with mainstream identifier and repository organizations has already led to delivery of one robust identifier scheme for preservation of annotation in the humanities and science and technology communities. PSA-WG will commence by applying that scheme to historic research investment redelivered by the preliminary partners and then apply it to a new research activity. Following evaluation of these solutions the WG will support an already-identified first tier of early adopter research projects. Publication of use cases from these activities, together with the guidelines identified in WP1 and WP7 of the work-plan in Section 4, will provide an effective blueprint for preservation of annotation-based research data in both redelivery and new projects. Funding applications are also being developed for a second tier of projects by the PSA-WG implementation partners (see Section 5). It is envisaged that these two tiers of adoption will create sufficient critical mass to support an international conference on preservation of scientific annotation in the humanities and science and technology communities during the 12 months following the completion of the WG work-plan.

In a second phase of adoption the WG will work with its collaboration partners in the Earth sciences, bio-technology and medical imaging communities to establish a tier of adoption projects building on the pilot annotation use-cases in those communities. These projects cannot be identified at the outset, but there are clear strategic differences compared with the phase one adoption plan set out above. In the humanities and science and technology, research activities are more corpus-specific in comparison with the selected phase two communities, where relatively homogeneous data supports a multiplicity of research activities. For example data at a given level of detail might be produced by different sensors: tomography can represent a cross section through a human body using X-rays or ultrasound. This data can support a wide range of expert interpretation but all medical data is about the human body. PSA-WG collaboration partners such as ESIP and ELIXIR and DICOM users have broad constituencies and will enable the WG to develop annotation preservation solutions more generally than the corpus-specific workflows already encountered in the phase one communities.

---

## 4. Work Plan

### 4.1 Activities

The PSA-WG work plan is organized into work-packages (WPs) representing three activities:

- outreach to insure that the existence of data preservation vulnerabilities arising from incorporating stand-off annotation are comprehended by the research community; subsequently, effective communication of the Technical Report of PSA-WG as well as its use-case outputs will be central to success of the WG—WP1 and WP7
- evaluating and selecting techniques and technologies to plug the gaps in current data preservation planning relating to annotation, and in so doing overcome roadblocks to creating workable end-to-end solutions for long-term preservation of annotation—WP2, WP3, WP4 and WP6

- consultation with multiple research communities already employing a range of workflows for creating and maintaining annotation-based research data, in order to create effective preservation solutions which are tailored for those communities and encapsulate them in use-cases—such pilot project activities with PSA-WG partners will occur in WP4 and WP5

These activities will require that PSA-WG produce multiple documents:

- annotation Guideline communicating preservation vulnerabilities, plus an Annotation Primer describing techniques and effective use of annotation—initially posted as a PDF document on the RDA website and subsequently conveyed via summary presentations and posters at key conferences
- a Technical Report will be produced and posted as a PDF document on the RDA website after evaluating preservation of annotation using the *core implementation scheme* (see below), and implementing pilot use-cases together with evaluation/reassessment
- Use-Case publications will serve two purposes—firstly in preliminary form as evaluations supporting the PSA Technical Report, and subsequently as Recommendations tailored for specific communities

PSA-WG’s technical evaluation and selection activities are expected eventually to address *multiple identifier implementation schemes*, reflecting varying PID adoption and annotation instruments in use by different communities. However, the WG must initially be able to demonstrate concrete preservation solutions within specific fields, in order to articulate vulnerabilities and demonstrate credible remedies to the wider community. Accordingly the WG will first evaluate a ‘*core*’ *identifier scheme* using already-accessible technical components. Discussions among the initial PSA-WG membership (see Section 6.) including CERN and ORCID between P12 and P13, have explored developing extensions of existing services to enable robust preservation of stand-off annotation to be available at the time of launch of the WG. Specifically, these include the creation of an Annotation WorkType by ORCID and the support of an Annotation Collection data resource type by Zenodo, which developments have now been confirmed to PTPP by those organizations. Together with identification of annotated digital resources using identifiers including URN, this provides PSA-WG with necessary and sufficient components for its core identifier scheme. The development and testing of these mechanisms in partnership with CERN and ORCID early in its timeline will enable PSA-WG to turn its focus rapidly towards development of pilot projects, leading to use cases and consultation with other stakeholders. Proceeding beyond the core identifier scheme through such consultation is addressed in Section 3.

A potential follow-on activity of PSA-WG which is under consideration is convergence of different instrument-dependent representations for annotation (including variants of OADM) which are currently in use, and which lead to lack of interoperability: annotations created using one instrument cannot be discovered, viewed or maintained using other instruments. Discussion with other stakeholders has commenced, with proposed development of a further Recommendation document to achieve this, but it is not addressed further here.

## 4.2 Work-Packages

In summary, PSA-WG work-packages will include the following communication and technical activities:

WP1: produce ‘Annotation Preservation Vulnerabilities’ guideline

WP2: implement and evaluate ORCID/URN/Zenodo core identifier scheme using ‘Twinger’ corpus<sup>22</sup> (recently-commenced research project)

WP3 redeliver pre-WADM humanities research projects (Bodleian Libraries and Heidelberg University) using core identifier scheme; produce use-cases

WP4: plan and implement pilot preservation projects, leading to production of use-cases in partnership with ESIP and European Bioinformatics Institute (EMBL-EBI), as well as either developing further an existing University College Hospital, London, DICOM pilot (which was commenced after P12) or establishing another medical imaging annotation project

WP5: consultation with California Digital Library, IIF 3D User Community, Research Activity Identifier, Research Object Consortium to extend the core identifier implementation scheme to additional research activity and digital resource PIDs and potentially to CoolURIs<sup>23</sup>

WP6: production of PSA-WG Technical Report on Preserving Scientific Annotation

WP7: launch Technical Report and Recommendations based on use-cases

### 4.3 Timeline



### 4.4 Milestones

It is hoped that endorsement of PSA-WG by TAB will be forthcoming before the beginning of June 2019. Accordingly, a four-month window is available for completion of the WG’s first Guideline before Plenary-14, delivered as an RDA website document and for preparation of presentation and poster

<sup>22</sup> <https://indico.hasdai.org/event/26/>

<sup>23</sup> <https://www.w3.org/Provider/Style/URI>

materials. PTPP will schedule a mid-term review of PSA-WG activities at Plenary-15, which will also provide an opportunity to update the wider community. All key activities identified in this work-plan will be complete in time for Plenary-16 with the exception of communication activities, which the WG anticipates extending beyond WP7. The following summarizes PSA-WG milestones:

M1: publish Annotation Primer

M2: release vulnerability Guideline to coincide with RDA Plenary-14

M3: complete new annotation and redelivery preservation (Twinger/Basel, Curiosities/Bodleian Libraries and Hachiman/Heidelberg University corpora) using core identifier scheme

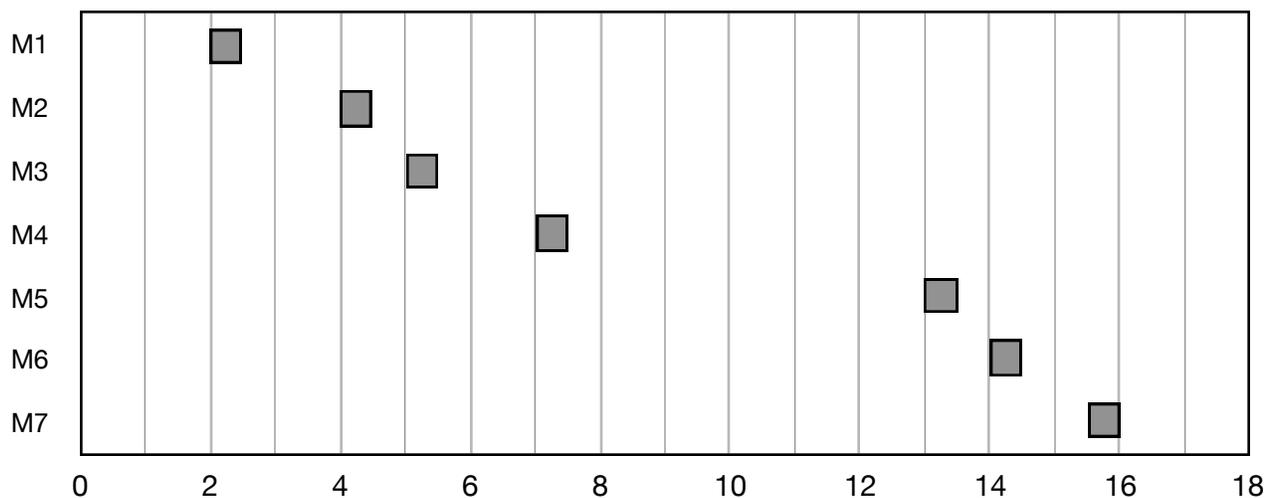
M4: commence use-case implementations with ESIP, EBL & DICOM communities

M5: release preliminary use-case reports for use in Technical Report and development of community-specific Recommendations for preservation of annotation

M6: complete identifier scheme consultation, extending initial PSA-WG core identifier scheme

M7: release Technical Report and Recommendations to coincide with RDA Plenary-16

PSA-WG milestones, by month



## 5. Initial Membership

Initial membership of PSA-WG dates from discussions at RDA Plenary-11, between PTPP-IG chairs, Bodleian Libraries, CERN, and Data Futures. Between P-11 and P-12 four legacy humanities research projects from Heidelberg University, plus a Bodleian Libraries project, all with complex annotation,

were redelivered using IIF image services and Invenio, together with conversion of annotations into OADM in a collaboration between CERN and Data Futures. A ‘Preserving Scientific Annotation’ Birds-of-a-Feather meeting took place at P-12, at which this work was presented by Data Futures, Bodleian and CERN, and the following members attended and expressed interest in joining PSA-WG:

Adachi, Sumiko	Japan Science and Technology Agency
Downs, Robert R.	CIESIN, Columbia University
Garcia, Leyla	ELIXIR HUB
Hienola, Anca	Finnish Meteorological Institute
Jejkal, Thomas	Karlsruhe Institute of Technology (KIT)
Jenkyns, Reyna	Ocean Networks Canada
Juty, Nick	University of Manchester and Identifiers lead for ELIXIR-UK
Jeremy Kenyon	University of Idaho
Lambert, Simon	UKRI-STFC
Li , Shih-Chieh Llya	CEO Xtrea.io
Martin, Jose	KAUST
Myers, Natalie	Research Librarian Notre Dame University
Morrison, Monica	Stellenbosch University
Stockhause , Martina	DHRZ
Weber, Tobias	Leibniz-Rechenzentrums, LRZ

Since Plenary-12 a number of additional RDA members have expressed interest in joining the WG:

Ó Carragáin, Eoghan	University College, Cork Library and Research Object community
Decker, Eric	Research Navigator, Europa Institute, Basel
Lamberty, Tom	Publisher, Merve Verlag
Narock, Tom	Department of Mathematics, Notre Dame of Maryland University, ESIP
Weale, Sara	The National Library of Wales, and Chair UV Community Group
Stozka, Rainer	Karlsruhe Institute of Technology (KIT)

Several new proposals for PSA-WG preservation of annotation pilot projects are currently in progress, including with CERN’s Digital Memory Project; Earth Science Information Partners (ESIP) via Jet

Propulsion Laboratory, Caltech; with ELIXIR via The European Bioinformatics Institute and with ORCID. Additionally, projects with The Bodleian Libraries, Maison de l'Orient et de la Méditerranée Jean Pouilloux, Lyon, Mnemoscene and with Notre Dame's Reilly Center for Science and Technology have commenced before Plenary-13. As a result, a new tier of pilot implementation partners is also in place to develop use-cases which, together with the WG chairs are represented by the following RDA members:

Cornwell, Peter (co-chair)	University of Westminster and Director, Data Futures LBG
Haak , Laure	Executive Director, ORCID
Jefferies, Neil	Head of Innovation, The Bodleian Libraries, Oxford University
Juty, Nick	Identifiers lead for ELIXIR-UK interoperability platform
Gonzalez, Jose	Section Leader, Digital Repositories, CERN
Meyers, Natalie (co-chair)	Head of Digital Scholarship, Notre Dame University
McGibbney, Lewis	JPL and Chair, ESIP Semantic Technology Committee
Morandiere, Bruno	Head of Digital Infrastructure for Overseas Laboratories, CNRS
Serif, Ina (co-chair)	Department of History, University of Basel
Silverton, Edward	Director, Mnemoscene Ltd and Chair, IIF 3D Community Group

---

## 6. Copyright Notice, License and Disclaimers

Copyright © 2019 PTTP editors and contributors. Published by the RDA Preservation Techniques, Tools and Policies Interest Group (PTTP-IG) under the CC-BY license; see disclaimer.

### 6.1 License

Specifications published by PTTP-IG are made available using the Creative Commons Attribution Required (CC-BY) license.

Please note that this license forbids the assertion, implied or explicit, that PTTP-IG, RDA or any of its members endorses or is any way associated with uses of the specifications or implementations thereof.

### 6.2 Disclaimers

Specifications published by PTTP-IG are made available with the following disclaimer of liability:

This document is provided “as-is”, and copyright holders make no representations or warranties, express or implied, including, but not limited to warranties of merchantability, fitness for a particular purpose, non-infringement, or title; that the contents of this document are applicable for any purpose,

nor that the implementation of such contents will not infringe any third-party patents, copyrights, trademarks or other rights.

Copyright holders will not be liable for any direct, indirect, special or consequential damages arising out of any use of PHTTP documents or the performance or implementation of the contents thereof.

This disclaimer is based on that employed by W3C specifications.