

RDA Interest Group Charter Template
Version 0.91, 9/19/2017

Name of Proposed Interest Group: *Physical Samples and Collections in the Research Data Ecosystem*

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):*

Physical samples are a basic element for reference, study, and experimentation in research. Tests and analysis are conducted directly on samples, such as biological specimens, rock or mineral specimens, soil or sediment cores, plants and seeds, water quality samples, archeological artefacts, or DNA and human tissue samples, because they represent a wider population or a larger context. Other physical objects, such as maps or analog images are also direct objects of study, and, if digitized, may become a source of digital data. There is an urgent need for better integrating these physical objects into the digital research data ecosystem, both in a global and in an interdisciplinary context to support search, retrieval, analysis, reuse, preservation and scientific reproducibility. This group aims to facilitate cross-domain exchange and convergence on key issues related to the digital representation of physical samples and collections, including but not limited to use of globally unique and persistent identifiers for samples to support unambiguous citation and linking of information in distributed data systems and with publications, metadata standards for documenting samples and collections and for landing pages, access policies, and best practices for sample and collection catalog, including a broad range of issues from interoperability to persistence.

A growing community of stakeholders, comprising domain scientists, collection curators, information scientists, data managers, all working at the interface with computational science, are developing detailed practices and standards around identifiers, vocabularies, and software interfaces, which are necessary for wider community application. Publishers and funders represent additional stakeholders interested in best practices for sample citation and registration of sample metadata in online catalogs that are fundamental for reproducibility of sample-based data and future use of valuable collection specimens. Currently, these efforts are fragmented, as is the communication of technical solutions and organizational best practices. This IG will support cross-disciplinary and international dialog helping to build technical and social bridges among a broad range of stakeholders to align and coordinate ongoing efforts, strengthen solutions, and broaden their adoption.

At RDA Plenary 4 and Plenary 6 Bird of Feather sessions were held that already gathered an international and multi-disciplinary group of stakeholders. A preliminary case statement was reviewed by participants in the P6 BoF and informed the current version.

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

Best practices, standards, and infrastructure are needed to properly link physical samples and collections to digital data generated by their study or to features in the real world. Samples need to be cited with

45 globally unique, persistent, and resolvable identifiers in publications to ensure that they can be
46 unambiguously linked to online metadata profiles (landing pages) and to other data generated by other
47 studies of the same sample. Scientists want to search for data for a given sample across the entire
48 literature. This can now be achieved as sample PIDs can be included in publication DOIs or data DOIs as
49 related identifiers that can be harvested and searched through systems like SCHOLIX. Scientists also
50 want to find out where a given sample can be accessed to reproduce data or add new measurements to the
51 available knowledge about a sample. Both the approaches to, and maturity of technical and organizational
52 solutions and infrastructure differ across the many disciplines that work with physical samples. Diverse
53 and uncoordinated practices make it difficult to advance the adoption of best practices that link physical
54 samples to the digital research data ecosystem. Further, commercial software providers for museum and
55 collection catalogs and publishers are reluctant to implement best practices if they are different and
56 incompatible across domains.

57

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

61 RDA presents a multi-disciplinary and international community engaged in research data management
62 that presents a unique opportunity for the goals of this IG. The objectives of this IG are:

- 63 1. Identify commonalities and diversities across the stakeholders and establish prioritized action items
64 that are appropriate for Working Groups. Relevant issues are: unique sample identifiers; sample
65 documentation including vocabularies and taxonomies and alignment with international metadata
66 standards; sample registration and interoperability of digital online catalogs; policies for sample
67 citation in publications; and access to samples and sample metadata.
- 68 2. Identify and characterize existing systems and solutions relevant to linking physical samples with
69 digital research data; identify gaps and challenges.
- 70 3. Facilitate international cooperation to develop harmonized approaches and best practices for physical
71 object identification and digital curation; enable the facilitation of object and sample identification
72 infrastructure both at the national and international levels.
- 73 4. Build linkages between object repositories and museums, digital data repositories, scientific
74 publications, museum software providers, and science communities.

75

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

79 Communities that will be involved in this IG range from museum and collection curators, to research data
80 managers, researchers in domain sciences, information sciences, and computer sciences, to publishers and
81 funders. Various workshops happened over the last few years that brought together stakeholders primarily
82 from interested in the topic of Physical Samples in the Digital Research Ecosystem, including:

- 83 • *Linking Environmental Data and Samples*, CSIRO, Australia, May 2017
- 84 • *Physical Samples and Digital Collections*, iConference, China, March 2017
- 85 • *Physical Samples, Digital Collections*, ASIS&T Conference, Denmark, October 2016

86 In the previous two BoF sessions at RDA P4 and P6 the following communities were represented:

- 87 • Biodiversity
- 88 • Oceanography

- 89 • French science archive
- 90 • Australian meteorology - water quality sampling and provenance
- 91 • German Research Center, library
- 92 • European PID network
- 93 • Geological Society
- 94 • Kew Gardens
- 95 • CDL - neurobiology, Berkeley museum
- 96 • Agricultural research in Italy, soil samples
- 97 • Zoology and environmental science
- 98 • Provenance and workflows, biodiversity workflows
- 99 • Material science, Air Force
- 100 • Ethnography
- 101 • Natural History Museums
- 102 • National Repositories
- 103

104 We will facilitate workshops at ASIS&T, JCDL, SPNHC, and domain-specific conferences such as AGU
 105 for the Earth and Space Sciences to broaden participation and dissemination of outcomes.

106 We will work with the following organizations to engage relevant communities.

- 107 1. International Geo Sample Number IGSN e.V. (Kerstin Lehnert, Jens Klump;
 108 <http://www.igsn.org>) - *Global implementation organization for unique sample identifiers,*
 109 *members if 5 continents)*
- 110 2. Global Biodiversity Information Facility (Donald Holborn)
- 111 3. Taxonomic Data Working Group (John Wiczorek, <http://www.tdwg.org>)
- 112 4. DISSCO (Distributed System of Scientific Collections, <http://dissco.eu>)
- 113 5. AuScope (Lesley Wyborn, <http://www.auscope.org.au/>)
- 114 6. EPOS (Kirsten Elger, <https://www.epos-ip.org/>)
- 115 7. SPNHC (Society For The Preservation of Natural History Collections, <http://www.spnhc.org>)
- 116 8. DataCite (<https://www.datacite.org>)
- 117 9. CODATA Task Group on Coordinating Data Standards amongst Scientific Unions (Marshall Ma,
 118 <http://www.codata.org/task-groups/coordinating-data-standards>)
- 119 10. ESIP (Earth Science Information Partners) (Erin Robinson, <http://www.esipfed.org>)
- 120 11. Scientific Collections International (SciColl, <http://scicoll.org>)

121
 122 Related RDA groups

- 123 • TAB
- 124 • WG/IG Chairs
- 125 • [Biodiversity Data Integration IG](#)
- 126 • [Long tail of research data IG](#)
- 127 • [PID IG](#)
- 128 • [Research Data Provenance](#)
- 129 • [RDA / TDWG Metadata Standards for attribution of physical and digital collections stewardship](#)

130
 131

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

- 134 1. A report that synthesizes existing best practices for digital curation and sharing of physical
- 135 samples from disparate disciplines and institutions.
- 136 2. A journal special volume on sample and collection management in the research data ecosystem
- 137 (journal TBD).
- 138 3. Creation of RDA Working Groups to develop recommendations for best practices and standards
- 139 related to sample unique identifiers, sample metadata, and sample citation, such that they can be
- 140 linked with data and publications derived from them.
- 141 4. Joint sessions with other RDA groups such as Biodiversity Data Integration IG, Long Tail of
- 142 Research Data IG, PID IG, Research Data Provenance, and others as appropriate for knowledge
- 143 exchange, to align with emerging relevant standards, and to promote recommendations from the
- 144 IG.
- 145 5. Facilitation of collaborations that advance interoperability between collection catalogs, sample
- 146 registries, data repositories, and publications for improved data sharing across disparate
- 147 disciplines, through e.g., alignment of sample metadata with existing metadata standards.

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

- 152 • Primary mechanism for communication will be email with periodic (quarterly web conferencing)
- 153 meetings, plus sessions at plenary meetings.
- 154 • We will also leverage other meetings such as EGU, AGU, SciDataCon, and ESIP
- 155 • Knowledge gathering and capture will be via RDA IG web site. We may use other collaboration
- 156 tools as appropriate, e.g. wiki's, or tools such as GitHub, or Center for Open Science Open
- 157 Science Framework.

158
 159 **Timeline** (Describe draft milestones and goals for the first 12 months):

- 160
 161 September 2017 - P10 session: BoF session, presentation of Case Statement
 162 December 2017 - AGU meeting and progress report
 163 March 2018 - P11 session, evaluate progress, revisit workplan
 164 September 2018 - P12

165
 166
 167 **Potential Group Members** (Include proposed chairs/initial leadership and all members who have
 168 expressed interest): *Bold indicates co-chairs*

169

FIRST NAME	LAST NAME	EMAIL
Kerstin	Lehnert (co-chair)	lehnert@ldeo.columbia.edu
Lesley	Wyborn (co-chair)	lesley.wyborn@anu.edu.au
Jens	Klump (co-chair)	Jens.Klump@csiro.au
Simon	Cox (co-chair)	Simon.Cox@csiro.au
Helen	Glaves	hmg@bgs.ac.uk
Rowena	Davis	rowenaidavis@email.arizona.edu

Markus	Stocker	mstocker@marum.de
Lindsay	Powers	lpowers@usgs.gov
Christopher	Lenhardt	clenhardt@renci.org
Denise	Hills	dhills@gsa.state.al.us
Dirk	Fleischer	dfleischer@kms.uni-kiel.de
Kirsten	Elger	kelger@gfz-potsdam.de
Wim	Hugo	wim@saeon.ac.za
Colleen	Strawhacker	colleen.strawhacker@colorado.edu
Sarah	Ramdeen	ramdeen@email.unc.edu
John	Wieczorek	tuco@berkeley.edu
Leslie	Hsu	lhsu@usgs.gov
Donald	Hobern	dhobern@gbif.org
Nicky	Nicholson	n.nicolson@kew.org
Unmil	Karadkar	unmil@ischool.utexas.edu
Ashlee	Dere	adere@unomaha.edu
Nicholas	Car	Nicholas.Car@ga.gov.au
Anusuriya	Devaraju	anusuriya.devaraju@csiro.au
Sean	Toczko	sean.jamstec@gmail.com
Lynne	Yarmey	yarmel@rpi.edu
Dawn	Wright	DWright@esri.com kelger@gfz-potsdam.de
Marshall	Ma	max@uidaho.edu

170

171

172

The following are additional potential participants who attended the previous BoF sessions at P4 and P6:

FIRST NAME	LAST NAME	EMAIL	Institution
Aaron	ADDISON		Washington University in St Louis
Arturo	ARIÑO PLANA	artarip@unav.es	University of Navarra
Toshihiro	ASHINO	ashino@acm.org	Toyo University
Sven	BINGERT	sven.bingert@gwdg.de	GWDC
Daphne	DUIN	daphne.duin@naturalis.nl	Naturalis Biodiversity Center
Ian	FORE		National Council Institute (NIH)
Kazu	FUKUDA		JAMSTEC
Bryon	FOSTER	Bryon.Foster@us.af.mil	USAF/AFRL
Margaret	FOTLAND	m.l.fotland@admin.uio.no	University of Oslo
Jason	JACKSON		Indiana University; Mathers Museum of World Cultures
John	KRATZ	John.Kratz@ucop.edu	California Digital Library
Giovanni	L'ABATE	giovanni.labate@entecra.it	Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CRA-ABP)

			Research centre for agrobiolgy and pedology
Bertram	LUDAESCHER	ludaesch@gmail.com	University of Illinois, Urbana-Champaign
Paolo	MISSIER	Paolo.Missier@ncl.ac.uk	Newcastle University
Magalie	MOYSAN	magalie.moysan@univ-paris-diderot.fr	Université Paris Diderot
Fiona	MURPHY	fionalm27@gmail.com	Research Consultant
Ritsuko	NAKAJIMA	rnakajim@jst.go.jp	Japan Science and Technology Agency
Nicky	NICOLSON	n.nicolson@kew.org	Royal Botanic Gardens, Kew
Kunihao	NIWA		Research Organization of Information and Systems (Japan)
Yoshinori	OCHIAI		JST
Carole	PALMER		iSchool Washington
Paul	SHEAHAN	sheahanpaul@hotmail.com	Sheahan
Paola	TAROCCO	ptarocco@regione.emilia-romagna.it	Geological, Seismic and Soil Survey, Emilia-Romagna Region (Italy)
Anne	THESSSEN	annethessen@gmail.com	The Data Detektiv
Nicholas	WEBER	nmweber@uw.edu	University of Washington
Matt	WOODBURN		Natural History Museum London
Themis	ZAMANI	sakka@grnet.gr	GRNET
Carlo	ZWÖLF	carlo-maria.zwolf@obspm.fr	Observatoire de Paris