

1 **RDA Interest Group Charter Template**

2 *Version 0.9, 9/17/2017*

3  
4 **Name of Proposed Interest Group:** *Physical Samples and Collections in the Research Data Ecosystem*

5  
6 **Introduction** *(A brief articulation of what issues the IG will address, how this IG is aligned with the RDA*  
7 *mission, and how this IG would be a value-added contribution to the RDA community):*

8  
9 Physical samples are a basic element for reference, study, and experimentation in research. Tests and  
10 analysis are conducted directly on samples, such as biological specimens, rock or mineral specimens, soil  
11 or sediment cores, plants and seeds, water quality samples, archeological artefacts, or DNA and human  
12 tissue samples, because they represent a wider population or a larger context. Other physical objects, such  
13 as maps or analog images are also direct objects of study, and, if digitized, may become a source of digital  
14 data. There is an urgent need for better integrating these physical objects into the digital research data  
15 ecosystem, both in a global and in an interdisciplinary context to support search, retrieval, analysis, reuse,  
16 preservation and scientific reproducibility. This group aims to facilitate cross-domain exchange and  
17 convergence on key issues related to the digital representation of physical samples and collections,  
18 including but not limited to use of globally unique and persistent identifiers for samples to support  
19 unambiguous citation and linking of information in distributed data systems and with publications,  
20 metadata standards for documenting samples and collections and for landing pages, access policies, and  
21 best practices for sample and collection catalog, including a broad range of issues from interoperability to  
22 persistence.

23  
24 A growing community of stakeholders, comprising domain scientists, collection curators, information  
25 scientists, data managers, all working at the interface with computational science, are developing detailed  
26 practices and standards around identifiers, vocabularies, and software interfaces, which are necessary for  
27 wider community application. Publishers and funders represent additional stakeholders interested in best  
28 practices for sample citation and registration of sample metadata in online catalogs that are fundamental  
29 for reproducibility of sample-based data and future use of valuable collection specimens. Currently, these  
30 efforts are fragmented, as is the communication of technical solutions and organizational best practices.  
31 This IG will support cross-disciplinary and international dialog helping to build technical and social  
32 bridges among a broad range of stakeholders to align and coordinate ongoing efforts, strengthen solutions,  
33 and broaden their adoption.

34  
35 At RDA Plenary 4 and Plenary 6 Bird of Feather sessions were held that already gathered an international  
36 and multi-disciplinary group of stakeholders. A preliminary case statement was reviewed by participants  
37 in the P6 BoF and informed the current version.

38  
39  
40 **User scenario(s) or use case(s) the IG wishes to address** *(what triggered the desire for this IG in the first*  
41 *place):*

42  
43 Best practices, standards, and infrastructure are needed to properly link physical samples and collections  
44 to digital data generated by their study or to features in the real world. Samples need to be cited with

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

56  
57 **Objectives** (*A specific set of focus areas for discussion, including use cases that pointed to the need for*  
58 *the IG in the first place. Articulate how this group is different from other current activities inside or*  
59 *outside of RDA.*):

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

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

74  
75 **Participation** (*Address which communities will be involved, what skills or knowledge should they have,*  
76 *and how will you engage these communities. Also address how this group proposes to coordinate its*  
77 *activity with relevant related groups.*):

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

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

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

- 86 • Biodiversity
- 87 • Oceanography
- 88 • French science archive

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

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

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

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

121 Related RDA groups

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

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

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

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

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

156  
 157 **Timeline** (Describe draft milestones and goals for the first 12 months):

- 158
- 159 September 2017 - P10 session: BoF session, presentation of Case Statement
- 160 December 2017 - AGU meeting and progress report
- 161 March 2018 - P11 session, evaluate progress, revisit workplan
- 162 September 2018 - P12

163  
 164  
 165 **Potential Group Members** (Include proposed chairs/initial leadership and all members who have  
 166 expressed interest): *Bold indicates co-chairs*

167

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

Lindsay	Powers	lpowers@usgs.gov
Christopher	Lenhardt	<a href="mailto:clenhardt@renci.org">clenhardt@renci.org</a>
Denise	Hills	<a href="mailto:dhills@gsa.state.al.us">dhills@gsa.state.al.us</a>
Dirk	Fleischer	dfleischer@kms.uni-kiel.de
Kirsten	Elger	<a href="mailto:kelger@gfz-potsdam.de">kelger@gfz-potsdam.de</a>
Wim	Hugo	wim@saeon.ac.za
Colleen	Strawhacker	colleen.strawhacker@colorado.edu
Sarah	Ramdeen	<a href="mailto:ramdeen@email.unc.edu">ramdeen@email.unc.edu</a>
John	Wieczorek	tuco@berkeley.edu
Donald	Hobern	dhobern@gbif.org
Nicky	Nicholson	<a href="mailto:n.nicolson@kew.org">n.nicolson@kew.org</a>
Unmil	Karadkar	unmil@ischool.utexas.edu
Ashlee	Dere	adere@unomaha.edu
Nicholas	Car	Nicholas.Car@ga.gov.au
Anusuriya	Devaraju	<a href="mailto:anusuriya.devaraju@csiro.au">anusuriya.devaraju@csiro.au</a>
Sean	Toczko	<a href="mailto:sean.jamstec@gmail.com">sean.jamstec@gmail.com</a>

168  
169  
170

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	<a href="mailto:artarip@unav.es">artarip@unav.es</a>	University of Navarra
Toshihiro	ASHINO	<a href="mailto:ashino@acm.org">ashino@acm.org</a>	Toyo University
Sven	BINGERT	<a href="mailto:sven.bingert@gwdg.de">sven.bingert@gwdg.de</a>	GWGDG
Daphne	DUIN	<a href="mailto:daphne.duin@naturalis.nl">daphne.duin@naturalis.nl</a>	Naturalis Biodiversity Center
Ian	FORE		National Council Institute (NIH)
Kazu	FUKUDA		JAMSTEC
Bryon	FOSTER	<a href="mailto:Bryon.Foster@us.af.mil">Bryon.Foster@us.af.mil</a>	USAF/AFRL
Margaret	FOTLAND	<a href="mailto:m.l.fotland@admin.uio.no">m.l.fotland@admin.uio.no</a>	University of Oslo
Jason	JACKSON		Indiana University; Mathers Museum of World Cultures
John	KRATZ	<a href="mailto:John.Kratz@ucop.edu">John.Kratz@ucop.edu</a>	California Digital Library
Giovanni	L'ABATE	<a href="mailto:giovanni.labate@entecra.it">giovanni.labate@entecra.it</a>	Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CRA-ABP) Research centre for agrobiolgy and pedology
Bertram	LUDAESCHER	<a href="mailto:ludaesch@gmail.com">ludaesch@gmail.com</a>	University of Illinois, Urbana-Champaign
Paolo	MISSIER	<a href="mailto:Paolo.Missier@ncl.ac.uk">Paolo.Missier@ncl.ac.uk</a>	Newcastle University
Magalie	MOYSAN	<a href="mailto:magalie.moysan@univ-paris-diderot.fr">magalie.moysan@univ-paris-diderot.fr</a>	Université Paris Diderot
Fiona	MURPHY	<a href="mailto:fionalm27@gmail.com">fionalm27@gmail.com</a>	Research Consultant

Ritsuko	NAKAJIMA	<a href="mailto:rnakajim@jst.go.jp">rnakajim@jst.go.jp</a>	Japan Science and Technology Agency
Nicky	NICOLSON	<a href="mailto:n.nicolson@kew.org">n.nicolson@kew.org</a>	Royal Botanic Gardens, Kew
Kunihao	NIWA		Research Organization of Information and Systems (Japan)
Yoshinori	OCHIAI		JST
Carole	PALMER		iSchool Washington
Paul	SHEAHAN	<a href="mailto:sheahanpaul@hotmail.com">sheahanpaul@hotmail.com</a>	Sheahan
Paola	TAROCCO	<a href="mailto:ptarocco@regione.emilia-romagna.it">ptarocco@regione.emilia-romagna.it</a>	Geological, Seismic and Soil Survey. Emilia-Romagna Region (Italy)
Anne	THESEN	<a href="mailto:annethessen@gmail.com">annethessen@gmail.com</a>	The Data Detektiv
Nicholas	WEBER	<a href="mailto:nmweber@uw.edu">nmweber@uw.edu</a>	University of Washington
Matt	WOODBURN		Natural History Museum London
Themis	ZAMANI	<a href="mailto:sakka@grnet.gr">sakka@grnet.gr</a>	GRNET
Carlo	ZWÖLF	<a href="mailto:carlo-maria.zwolf@obspm.fr">carlo-maria.zwolf@obspm.fr</a>	Observatoire de Paris