Physical samples & RDA

Workshop on FAIR for Physical Objects
Utrecht, 1 March 2023

10.5281/zenodo.7679137

Google slides

Presented by Esther Plomp (e.plomp@tudelft.nl)

RDA IG Physical Samples

IG Physical Samples and Collections in the Research Data Ecosystem

- Need for **FAIR Samples**: reproducible, reusable samples and recognition for sample collection/curation
- Objectives:
  - Identify and characterize existing systems and solutions relevant to **linking physical samples with digital research data** and publications, identify gaps and challenges;
  - Identify commonalities and diversities across the **stakeholders with aim to converge**;
  - Facilitate **international cooperation** to develop harmonized approaches and best practices for physical object identification and digital curation.

RDA IG Physical Samples

Focus of this presentation:

- **IGSN**: Global Persistent Identifier for Material Samples
- Collaboration with ESIP on **sample citation** recommendations
- RDA & ESIP Physical Samples **Webinar Series** (2021)
- **23 Things** for Physical Samples

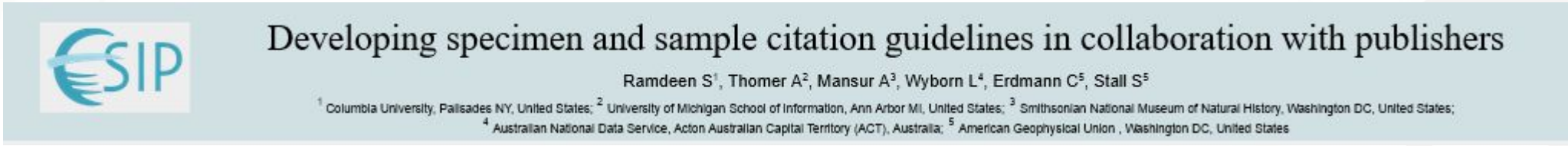
Other highlights:

- Biospecimens in the Life Sciences: a new proposed group (Rory McNeil, Sara El-Gebali, Rorie Edmunds)
- Sampling Nature Research Coordination Network
- iSamples

IGSN

- **I**nternational **G**eneric **S**ample **N**umber
  - IGSN 2040
- Partnership with DataCite (to register IGSNs via DataCite)
- Community of Practice (Archaeology)
  - Contact Rorie Edmunds

ESIP: Sample citation



Earth Science Information Partners (ESIP) work on **sample citation guidelines in the Physical Samples cluster**

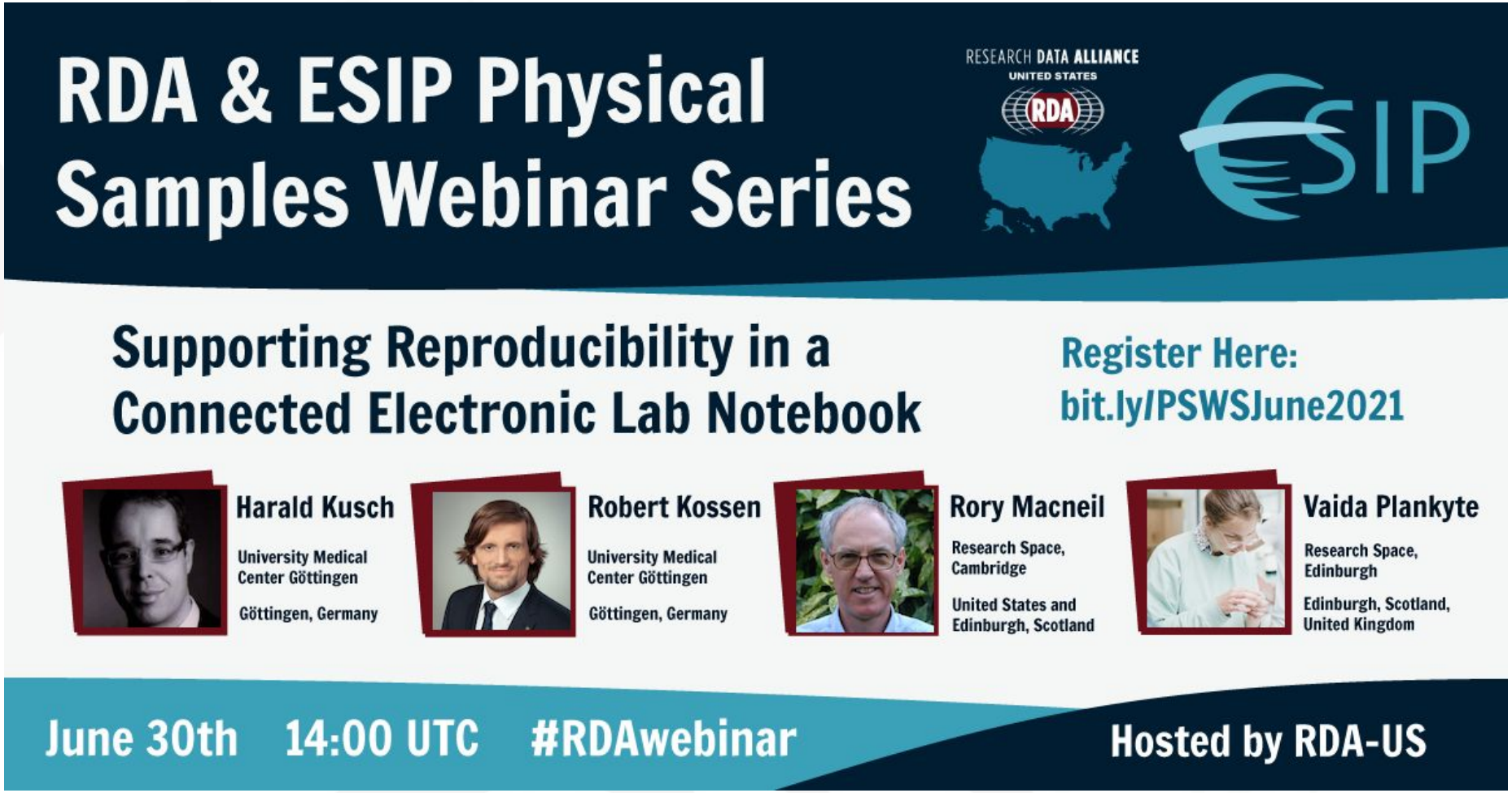
ESIP session on 'Cite your Samples!'

Session notes & recording

Contacts: Sarah Ramdeen & Val Stanley

See also the RDA Complex Citations Working Group

RDA/ESIP Webinar Series



[Supporting Reproducibility by Capturing Physical Sample Data and Metadata in a Connected Electronic Lab Notebook]

RDA/ESIP Webinar Series

RRIDs: A Way to Track Samples Through the Scientific Literature

RDA/ESIP Webinar Series

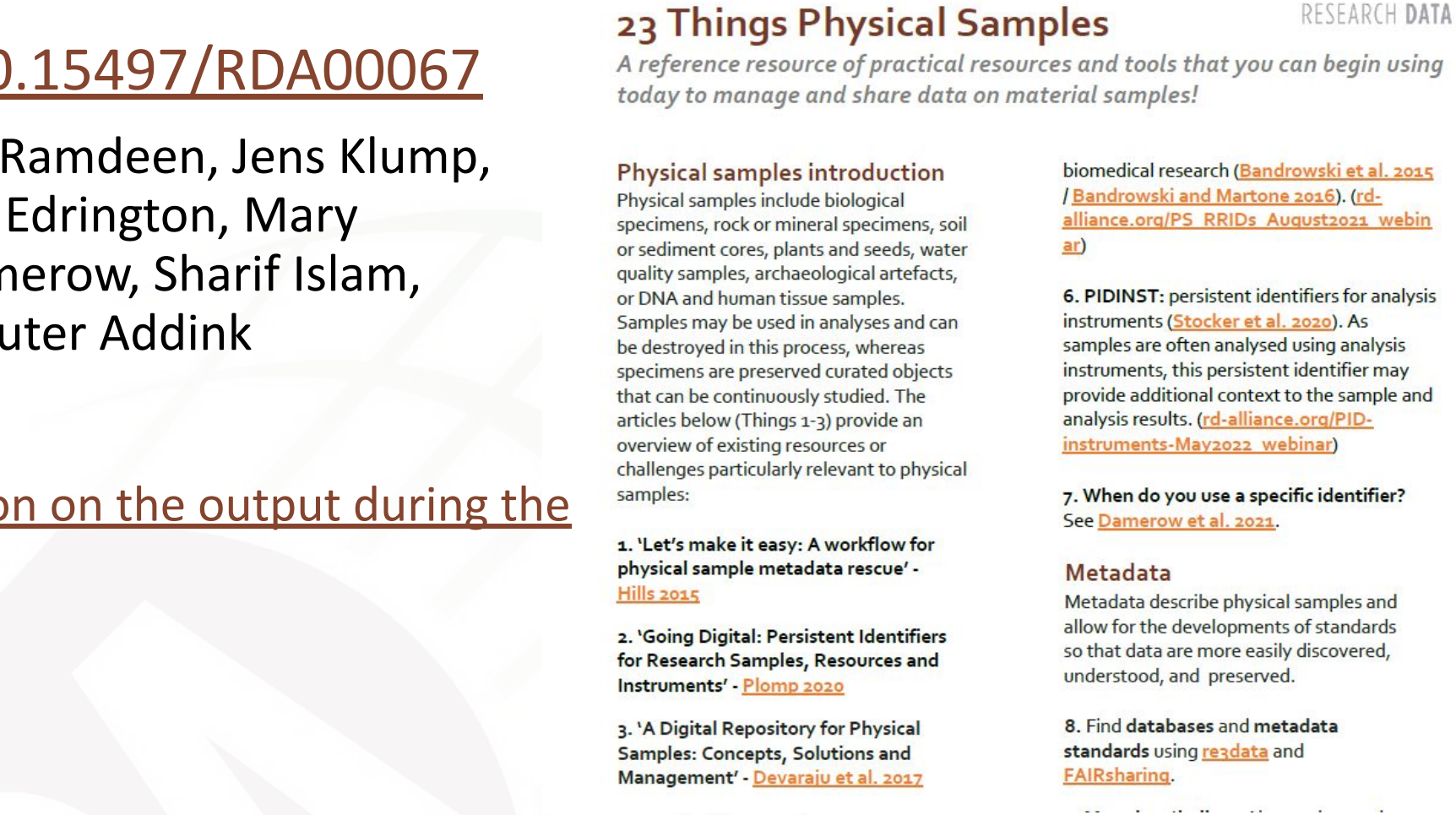
Supporting Interdisciplinary Sample Data Discovery, Integration, and Reuse

23 Things Physical Samples

https://doi.org/10.15497/RDA00067

- Esther Plomp, Sarah Ramdeen, Jens Klump, Lesley Wyborn, Clint Edrington, Mary Donaldson, Joan Damerow, Sharif Islam, Natasha Simons, Wouter Addink

- Video of a presentation on the output during the RDA plenary



23 Things Physical Samples

A reference resource of practical resources and tools that you can begin using today to manage and share data on material samples!

Physical samples introduction

Physical samples include biological specimens, rock or mineral specimens, soil or sediment cores, plants and seeds, water quality samples, archaeological artefacts, or DNA and human tissue samples. Samples may be used in analyses and can be destroyed in this process, whereas specimens are preserved curated objects that can be continuously studied. The articles below (Things 1-3) provide an overview of existing resources or challenges particularly relevant to physical samples:

1. 'Let's make it easy: A workflow for physical sample metadata rescue' - Hills 2015

2. 'Going Digital: Persistent Identifiers for Research Samples, Resources and Instruments' - Plomp 2020

3. 'A Digital Repository for Physical Samples: Concepts, Solutions and Management' - Devaraju et al. 2017

biomedical research (Bandrowski et al. 2015 / Bandrowski and Martone 2016). (rd-alliance.org/PS_RRIDs_August2021_webinar)

6. PIDINST: persistent identifiers for analysis instruments (Stocker et al. 2020). As samples are often analysed using analysis instruments, this persistent identifier may provide additional context to the sample and analysis results. (rd-alliance.org/PID-instruments-May2022_webinar)

7. When do you use a specific identifier? See Damerow et al. 2021.

Metadata

Metadata describe physical samples and allow for the developments of standards so that data are more easily discovered, understood, and preserved.

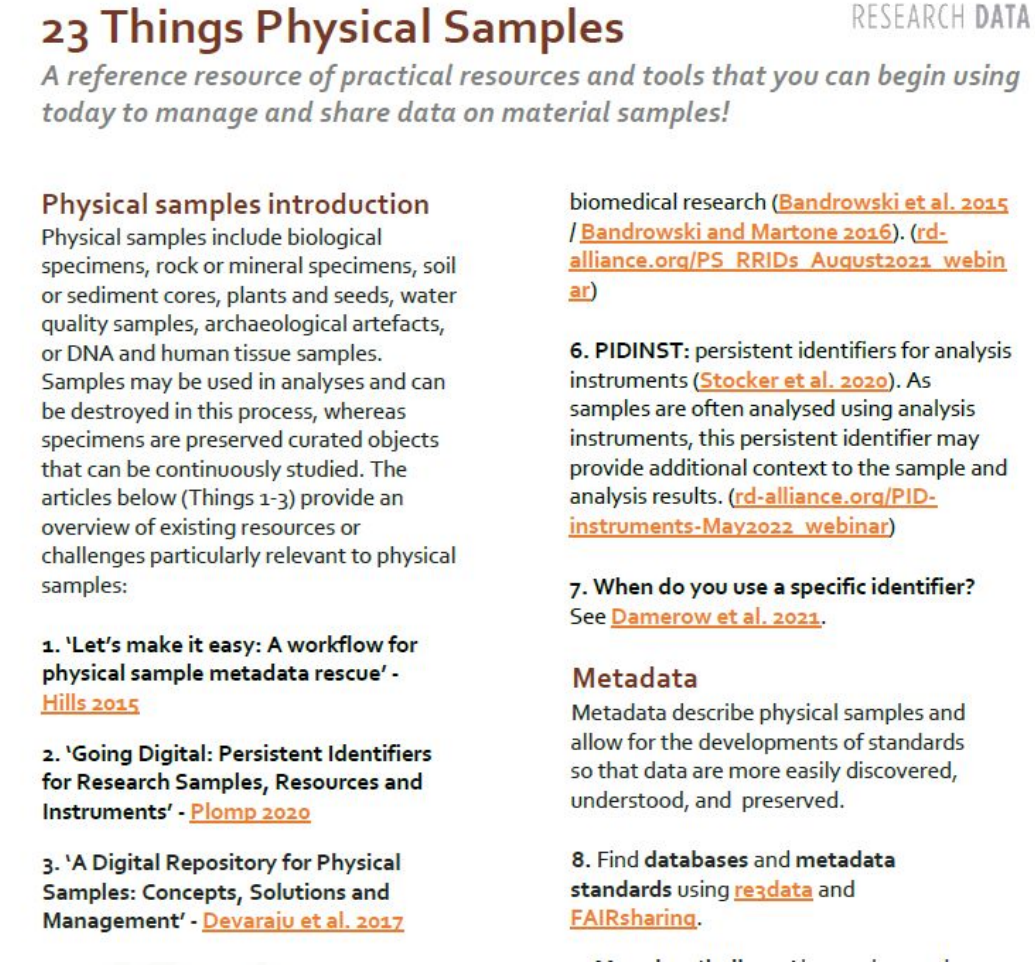
8. Find **databases** and **metadata standards** using re3data and FAIRsharing.

23 Things Physical Samples

https://doi.org/10.15497/RDA00067

3 pager on:

- Physical Samples introduction
- Persistent identifiers
- Metadata
- Citing Samples
- Data Licensing and Ownership
- Physical data tools
- Repositories
- Communities of Practice
- 23. Join the Research Data Alliance!



23 Things Physical Samples

A reference resource of practical resources and tools that you can begin using today to manage and share data on material samples!

Physical samples introduction

Physical samples include biological specimens, rock or mineral specimens, soil or sediment cores, plants and seeds, water quality samples, archaeological artefacts, or DNA and human tissue samples. Samples may be used in analyses and can be destroyed in this process, whereas specimens are preserved curated objects that can be continuously studied. The articles below (Things 1-3) provide an overview of existing resources or challenges particularly relevant to physical samples:

1. 'Let's make it easy: A workflow for physical sample metadata rescue' - Hills 2015

2. 'Going Digital: Persistent Identifiers for Research Samples, Resources and Instruments' - Plomp 2020

3. 'A Digital Repository for Physical Samples: Concepts, Solutions and Management' - Devaraju et al. 2017

biomedical research (Bandrowski et al. 2015 / Bandrowski and Martone 2016). (rd-alliance.org/PS_RRIDs_August2021_webinar)

6. PIDINST: persistent identifiers for analysis instruments (Stocker et al. 2020). As samples are often analysed using analysis instruments, this persistent identifier may provide additional context to the sample and analysis results. (rd-alliance.org/PID-instruments-May2022_webinar)

7. When do you use a specific identifier? See Damerow et al. 2021.

Metadata

Metadata describe physical samples and allow for the developments of standards so that data are more easily discovered, understood, and preserved.

8. Find **databases** and **metadata standards** using re3data and FAIRsharing.

Thank you!

Chairs RDA group: Lesley Wyborn, Jens Klump, Kerstin Lehnert

Get in touch with Lesley if you'd like to be involved!

ESIP: Sarah Ramdeen & Val Stanley

DataCite: Rorie Edmunds

& all community members

23 Things Physical Samples

23 Things Physical Samples

A reference resource of practical resources and tools that you can begin using today to manage and share data on material samples!

Physical samples introduction

Physical samples include biological specimens, rock or mineral specimens, soil or sediment cores, plants and seeds, water quality samples, archaeological artefacts, or DNA and human tissue samples. Samples may be used in analyses and can be destroyed in this process, whereas specimens are preserved curated objects that can be continuously studied. The articles below (Things 1-3) provide an overview of existing resources or challenges particularly relevant to physical samples:

1. 'Let's make it easy: A workflow for physical sample metadata rescue' - Hills 2015

2. 'Going Digital: Persistent Identifiers for Research Samples, Resources and Instruments' - Plomp 2020

3. 'A Digital Repository for Physical Samples: Concepts, Solutions and Management' - Devaraju et al. 2017

biomedical research (Bandrowski et al. 2015 / Bandrowski and Martone 2016). (rd-alliance.org/PS_RRIDs_August2021_webinar)

6. PIDINST: persistent identifiers for analysis instruments (Stocker et al. 2020). As samples are often analysed using analysis instruments, this persistent identifier may provide additional context to the sample and analysis results. (rd-alliance.org/PID-instruments-May2022_webinar)

7. When do you use a specific identifier? See Damerow et al. 2021.

Metadata

Metadata describe physical samples and allow for the developments of standards so that data are more easily discovered, understood, and preserved.

8. Find **databases** and **metadata standards** using re3data and FAIRsharing.

10.5281/zenodo.7679137

Process

- The idea was proposed during the 16th RDA Plenary meeting by a member (9 Dec 2020)
- Draft presented during the 17th RDA Plenary (20 April 2021)
  - Set up as a Google doc with comment permissions
- Two discussion sessions and asynchronous contributions
  - 14 June 19:00 UTC 2021
  - 15 June 8:00 UTC 2021
- Draft available for community review until 2nd of December 2021
- Feedback from community review was implemented early 2022
- Output is now finalised! https://doi.org/10.15497/RDA00067

Introduction

These articles provide an **overview of the existing resources or challenges** that are particularly relevant to physical samples:

- ‘Let’s make it easy: A workflow for physical sample metadata rescue’. Hills 2015: https://doi.org/10.1016/j.grj.2015.02.007
- ‘Going Digital: Persistent Identifiers for Research Samples, Resources and Instruments’. Plomp 2020: http://doi.org/10.5334/dsj-2020-046
- ‘A Digital Repository for Physical Samples: Concepts, Solutions and Management’. Devaraju et al. 2017: https://doi.org/10.1007/978-3-319-67008-9_7

Physical samples introduction

Physical samples include biological specimens, rock or mineral specimens, soil or sediment cores, plants and seeds, water quality samples, archaeological artefacts, or DNA and human tissue samples. Samples may be used in analyses and can be destroyed in this process, whereas specimens are preserved curated objects that can be continuously studied. The articles below (Things 1-3) provide an overview of existing resources or challenges particularly relevant to physical samples:

1. ‘Let’s make it easy: A workflow for physical sample metadata rescue’ - Hills 2015

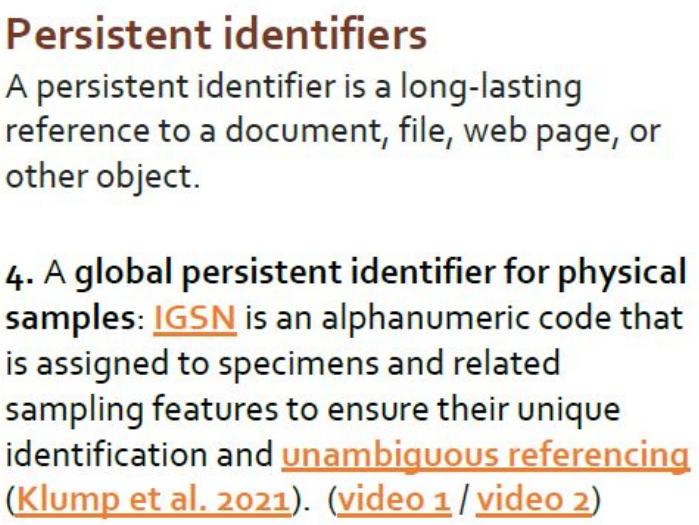
2. ‘Going Digital: Persistent Identifiers for Research Samples, Resources and Instruments’ - Plomp 2020

3. ‘A Digital Repository for Physical Samples: Concepts, Solutions and Management’ - Devaraju et al. 2017

Persistent identifiers

A persistent identifier is a long-lasting reference to a document, file, web page, or other object.

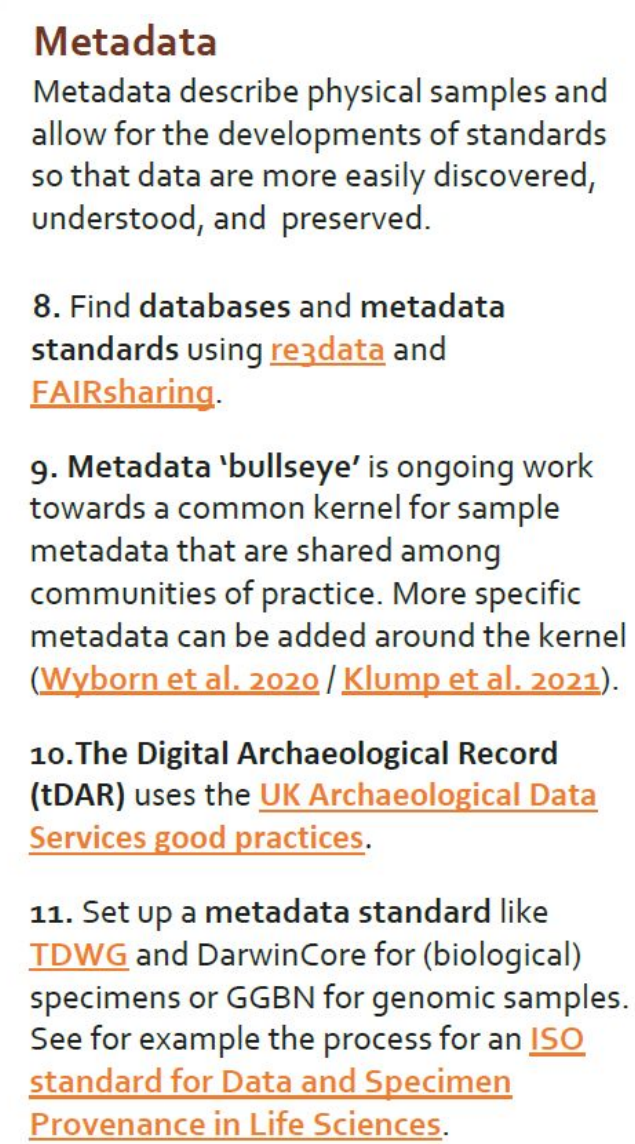
- A global persistent identifier for physical samples: **IGSN** is an alphanumeric code that is assigned to specimens and related sampling features to ensure their unique identification and unambiguous referencing. Klump et al. 2021: http://doi.org/10.5334/dsj-2021-033
- The Research Resource Identifiers (**RRIDs**) supports persistent and unique identifiers for referencing a research resource (such as antibodies, organisms) in biomedical research https://scicrunch.org/resources
- **PIDINST:** persistent identifiers for analysis instruments https://doi.org/10.5334/dsj-2020-018
- When do you use a specific identifier? See Damerow et al. 2021: http://doi.org/10.5334/dsj-2021-011



Metadata

Metadata describes physical samples and allows the developments of standards so that data is more easily discovered, understood, and preserved.

- Find databases and metadata standards using re3data and FAIRsharing: https://www.re3data.org/ / https://fairsharing.org/
- Metadata 'bullseye': Working towards a common kernel for sample metadata that are shared among a community of practice. See Wyborn et al. 2020: http://doi.org/10.5281/zenodo.4694740
- The Digital Archaeological Record (tDAR) uses the UK Archaeological Data Services good practices
- Set up a metadata standard. See for example the process for an ISO standard for Data and Specimen Provenance in Life Sciences: https://doi.org/10.5281/zenodo.5093125



Metadata

Metadata describe physical samples and allow for the developments of standards so that data are more easily discovered, understood, and preserved.

8. Find **databases** and **metadata standards** using re3data and FAIRsharing.

9. **Metadata 'bullseye'** is ongoing work towards a common kernel for sample metadata that are shared among communities of practice. More specific metadata can be added around the kernel (Wyborn et al. 2020 / Klump et al. 2021).

10. **The Digital Archaeological Record (tDAR)** uses the UK Archaeological Data Services good practices.

11. Set up a **metadata standard** like TDWG and DarwinCore for (biological) specimens or GGBN for genomic samples. See for example the process for an ISO standard for Data and Specimen Provenance in Life Sciences.

Citing Samples

- State how others should cite the data to make it easy for people to refer to your resources, see RRIDs and DataCite for examples.
- Work is ongoing on guidelines specifically for specimen and sample citation.

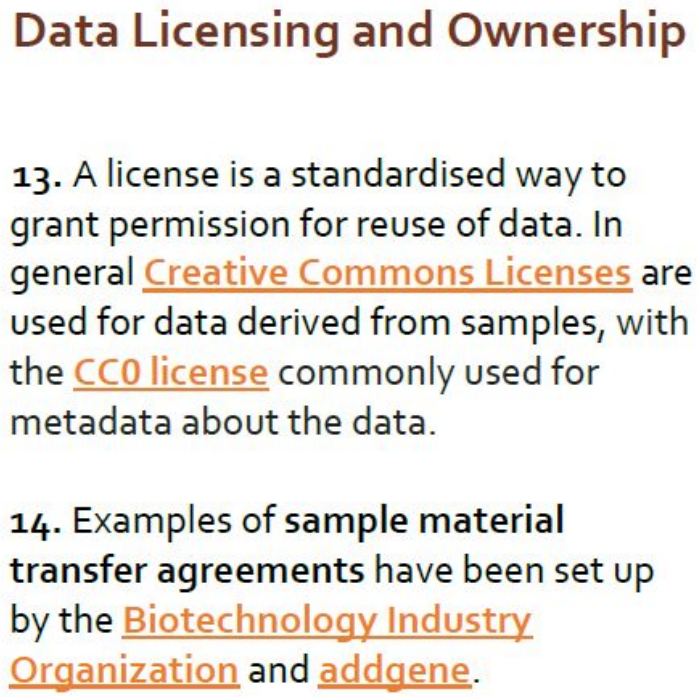
Citing Samples

There are existing initiatives that are helping to promote the scholarship of data by encouraging and enabling data citation, assigning identifiers to data, creating links between documents and data, and helping users properly attribute credit to data producers.

12. State how others should cite the data to make it easy for people to refer to your resources, see RRIDs and DataCite **for examples**. Work is ongoing on guidelines specifically for specimen and sample citation.

Data Licensing and Ownership

- In general **Creative Commons Licenses** are used for data derived from samples, with CC0 license commonly used for metadata about the data https://creativecommons.org/licenses
- Examples of **sample material transfer agreements** have been set up by the Biotechnology Industry Organization and addgene: https://help.addgene.org/hc/en-us/sections/201196819-Material-Transfer-Agreements-MTAs- / https://www.wipo.int/tk/en/databases/contracts/texts/bio.html


Data Licensing and Ownership

13. A license is a standardised way to grant permission for reuse of data. In general Creative Commons Licenses are used for data derived from samples, with the CC0 license commonly used for metadata about the data.

14. Examples of **sample material transfer agreements** have been set up by the Biotechnology Industry Organization and addgene.

Physical data tools

- Some **Electronic Lab Notebooks** have sample management modules RSpace / eLabNext / Labguru / Chemotion / Labfolder
- Examples of other **software tools** include the R-package baRcodeR to generate barcode labels for physical samples (Wu et al. 2020: https://doi.org/10.1111/2041-210X.13405) and Menoci which allows tracking of samples (menoci.io, 10.1186/s12859-020-03928-1), and FAIMS for offline data collection in the field.

Physical data tools

15. Some **Electronic Lab Notebooks** have sample management modules (RSpace / eLabNext / Labguru / Chemotion / Labfolder) (rd-alliance.org/physical-sample-webinar1-June2021)

16. Examples of other **software tools** include the **R-package baRcodeR** to generate barcode labels for physical samples (Wu et al. 2020), Menoci which allows tracking of samples (Surh et al. 2020), and FAIMS for offline data collection in the field.

Repositories

Existing infrastructures for physical samples include:

- iSamples
- The System for Earth Sample Registration (SESAR)
- Index to Marine and Lacustrine Geological Samples (IMLGS)
- Biological Research Samples Stock Centre List by Mary Donaldson https://doi.org/10.5281/zenodo.4923489
- Or set up a new repository, for example, a Biobank or a directory such as the BBMRI-ERIC Directory of Biobanks. You can follow recommended practices for Open Archival Information Systems or get inspired by ‘Managing Natural Science Collections: A guide to Strategy, Planning and Resourcing’ by Huxley et al. 2020

Repositories

Existing infrastructures for physical samples include:

17. Interdisciplinary Infrastructure for samples such as iSamples (Davies et al. 2021).

18. The System for Earth Sample Registration (SESAR) is a community platform that helps make Earth, Environmental, and Planetary Science samples more discoverable, accessible, and reusable (geosamples.org).

19. Index to Marine and Lacustrine Geological Samples (IMLGS) is a database that helps scientists and others discover, learn about, and obtain geologic materials from participating repositories.

20. Biological Research Samples Stock Centre List, which lists stock centres that might be suitable for the deposit of biological samples (Donaldson 2021).

21. Or set up a **new repository**, for example, a Biobank or a **directory** such as the BBMRI-ERIC Directory of Biobanks. You can follow recommended practices for **Open Archival Information Systems** or get inspired by **‘Managing Natural Science Collections: A guide to Strategy, Planning and Resourcing’** (Huxley et al. 2020).

Communities of Practice

Several communities are already involved in conversations about metadata and persistent identifiers for physical samples with the larger community of researchers, technologists, funders, publishers, and others to develop solutions and share best practices.

a. Earth Science Information Partners (ESIP)

b. Sampling Nature

c. Environmental Systems Science (ESS-DIVE)

d. Distributed System of Scientific Collections (DiSSCo)

e. the Humanities Commons

f. Consortium of European Taxonomic Facilities (CETAF)

g. The RDA group on Physical Samples and Collections

h. Australian Research Data Commons

10.5281/zenodo.7679137

Communities of Practice

Several communities are already involved in conversations about metadata and persistent identifiers for physical samples, working on solutions and sharing best practices:

22

a. Earth Science Information Partners (ESIP) has a Physical Samples Cluster and holds an annual conference with sessions that address physical samples in research.

b. Sampling Nature is a research coordination network organising workshops and working groups.

c. Environmental Systems Science (ESS-DIVE) stores and publicly distributes data from observational, experimental, and modelling research.

d. Distributed System of Scientific Collections (DiSSCo) provides research infrastructure and a platform for discussions for **Natural Science** Collections and digital specimens (Hardisty et al. 2021).

e. The Humanities Commons have several groups that work on **primary sources**.

f. Consortium of European Taxonomic Facilities (CETAF) is Europe's network of **biological and geological collections**.

g. The RDA group on Physical Samples and Collections facilitates cross-domain discussions around key issues related to the digital representation of physical samples and collections (rd-alliance.org/ps-interdisciplinarysampledata-October-webinar).

h. Australian Research Data Commons (ARDC) Information Management for Physical Samples Community of Practice will bring together practitioners in the Australian research sector to foster best practice in the identification, citation and management of physical samples throughout the research lifecycle.

Join the Research Data Alliance!

Belong to an international community who builds social and technical bridges to enable data sharing. It's free to join by visiting the website, then subscribe to the Physical Samples and Collections in the Research Data Ecosystem Interest Group

Compiled by the IG: 23 Things Physical Samples