Sampling Nature: Samples with Consequences Workshop

NYU | New York
6-7 December 2023

Background

The Sampling Nature Research Coordination Network (RCN) is creating connections across the natural history value chain that samples nature at field sites, labs, repositories, and museums through thematic workshops and virtual events to exploit the potential of an accessible, integrated corpus of material sample data. For more details on the project and planned deliverables, see the website: https://samplingnature.github.io/.

The Sampling Nature Research Coordination Network (RCN) is a project supported by the U.S. National Science Foundation (NSF 2129268) to establish a network of stakeholders in the research community that enhances the Natural History Value Chain for Sustainability Science. The project brings together an interdisciplinary group of investigators from multiple NSF directorates, federal agencies, Indigenous groups, museums, publishers, and infrastructure. The Sampling Nature RCN aims to build a collaborative research community around an accessible

and integrated corpus of material samples and associated data, communicating the importance of material samples in science, society, and sustainability.

The workshop will focus, in particular, on how samples have multiple consequences depending on context and need and how these can change over time. Samples also have afterlives - a sample that did not seem so important comes to hold a central place in an analysis previously not considered. This raises questions about finding and curating samples, samples that the research community may not know about, sample chains (for instance, who and what are the changing parties that have an interest in a certain sample), and the policies that may or may not be in place to govern the samples storage, access, and reuse. This workshop will bring together 4 different contexts for considering different kinds of sampling activities in New York City.

Further questions we will consider include:
When do certain samples become obsolete? Who makes decisions about obsolescence?

Workshop Goals

1. Understand how samples are collected in different contexts like a city (who and why do they collect samples, i.e., climate change, storms, population density - what urban brings forth)
    a. We will discuss use cases presented by people sampling in NYC. Sampling different things for different purposes
    b. In these use cases, we want examples of projects that intend to sample (1) bio, (2) geo, (3) anthro artifacts
    c. Each use case will be reviewed through the lens of metadata, policy, long-tail, and education/training
2. Do we see corresponding patterns across these? What happens if there are new uses for those samples? Does the policy, metadata, and curation of the sample permit/facilitate reuse?
3. Collecting responses to the community - What kind of samples do you wish you had access to? How can samples be repurposed for multiple uses?
4. Digitization of collection - could iSamples in a box be added? Could the metadata be harvested? Are there policy limitations?

Use Cases

1. NYC Botanic Garden
2. City Archeologist - Amanda Sutphin
3. Biosampling for Public Health - Braden Tierney

Expected Outcomes

1. To understand ethical, social, and legal issues for samples and how it differs from place to place and how material sample policies might differ in comparison to policies for digital data
2. To identify stakeholders that are responsible for samples and/or sampling, to document their policies with respect to samples and sampling (e.g., funding, academic, journals, government) and determine how those policies might affect reuse.
3. Recommendations for policy and infrastructure solutions? - LocalContext, How are legal metadata tracked (what is legal metadata?)
4. Present synthesis of these recommendations and review/finalize policy white paper
5. Grow the Sampling Nature community

Expectations for Participants

- Out-of-town participants should arrive on Tuesday, December 5, and be prepared to start at 9 AM on December 6.
- All workshop participants are expected to be present 9-5:30 on Wednesday, December 6, and Thursday, December 7
- Please prepare **one slide representing a specimen or collection** example and a social, ethical or legal issue. Please add them to this slide deck by Tuesday, December 5.
- Participants are expected to comply with the meeting code of conduct.

Agenda

Workshop Material Folder | Miro Board

Tuesday, 5th December (Arrival Day)

- Pre-Meeting workshop:
    - GEOME Steering Committee (plus guests) 2-4pm EST (hybrid - zoom)
- Optional Social Hour and dinner on your own

Wednesday 6th December Kimmel Center - Room 903

8:45   Registration

9:00   Welcome
- Introductions (Jane)
- Sampling Nature RCN Introduction (Kerstin) - Slides
- Workshop Overview (Erin)

9:30   NY Botanical Garden (Alex and NYBG team) - [Slides](#)

10:30   Coffee break

11:00   Participant Intro & Share a sample (choose a social, legal, or ethical issue that sits around the sample) - [Slides](#)

12:00   Lunch - (Out in Greenwich Village) - Top Thai Greenwich, Ramen Takumi, Chow house

13:30   NYC Archaeological Repository (Amanda Sutphin) - [Slides](#)

14:30   Breakout Group 1: Sample Journeys

15:15   Drop stuff and leave for field Trip (by 15:30)

16:00   Field Trip: Tour of NYC Archaeological Repository (114 W 47th St) led by Jessica Striebel MacLean https://archaeology.cityofnewyork.us/

18:00   Workshop Group Dinner (At Washington Square Hotel in Lounge)

Thursday 7th December | Dean's Conference Room 122 (4 Washington Square N)

8:45   Registration

9:00   Welcome back and recap from Day 1 - What stuck with you

9:30   NYC Use Cases
● Braden Tierney (Weill Cornell Medicine NYC)

10:30   Coffee Break

10:45   Showing of "[E Kore Au E Ngaro | The Connection Remains](#)," a film by the Whakatōhea Māori Trust Board and Local Contexts

~~11:15   Breakout Groups 2: Defining and/or clarifying ethical, legal, and social issues~~

~~12:00   Report back~~

4

12:30 Lunch (Out in Greenwich Village) - Ramen Takumi (Japanese), Uncle Ted's (Chinese): 38 E 8th St

  13:30 - Indigenous Metadata Bundles
- Report back on GEOME pre-meeting - Rename
- LocalContext - Status of Indigenous Metadata Bundle, where does it go?

- 2:00 Breakout conversations
    - Metadata - Interoperability spec; walk through loan
    - Tutorial Materials
        - Run through of how to register for a LocalContext notice
        - LocalContext Project: https://localcontextshub.org/institutions/projects/actions/99/3f341ef3-31dc-41ac-ab3b-d41481626660/
    -
    -

15:00 Coffee Break

  15:30 Report Back
  16:30 [Workshop Synthesis and Next Steps] - Now what?
    - Cyberinfrastructure needs (recommendations for iSamples)
        - Emu - need a champion to implement existing framework

17:30 Workshop Ends

Post-workshop happy hour

Participant List

Name	Organization
Jane Anderson	NYU
Pier Luigi Buttigieg	Alfred Wegener Institute / IOC-UNESCO IODE & GOOS / Omic BON
Eric Crandell	Penn State
Neil Davies	UC Berkeley
John Deck	UC Berkeley
Sarah Kansa	Open Context
Eric Kansa	Open Context
Leigh Lieberman	Open Context
Kerstin Lehnert	Columbia
Alex McAlvay	New York Botanical Garden
Chris Meyer	Smithsonian
Matthew Pace	New York Botanical Garden
Lindsay Powers	USGS
Natalie Raia	Arizona
Erin Robinson	Metadata Game Changers
Amanda Sutphin	NYC Archaeological Repository
Luca Telesca	LDEO
Braden Tierney	Weill Cornell Medicine NYC
Dave Vieglais	Kansas
Maheata White Davies	Richard B. Gump South Pacific Research Station, UC Berkeley
Saebyul Choe	LDEO
Sean Cao	LDEO

Code of Conduct

The Sampling Nature RCN Project is dedicated to providing a welcoming and supportive environment for all people, regardless of background or identity. As such, we do not tolerate disrespectful behavior to our community members or that excludes, intimidates, or causes discomfort to others. We do not tolerate discrimination or harassment based on characteristics that include, but are not limited to: gender identity and expression, sexual orientation, disability, physical appearance, body size, citizenship, nationality, ethnic or social origin, pregnancy, familial status, veteran status,

genetic information, religion or belief (or lack thereof), membership of a national minority, property, age, education, socio-economic status, technical choices, and experience level.

To report a violation, email workshop organizers: Jane Anderson (ja77@nyu.edu), or Erin Robinson (erin@metadatagamechangers.com) or the project PI, Kerstin Lehnert (lehnert@ldeo.columbia.edu)

Venue & Logistics

Location: New York University ([Map of NYU](#) with buildings on campus)
- *Wednesday Workshop, [Kimmel Center](#) Rm 903*
- *Wednesday Field Trip, NYC Archaeological Repository ([114 W 47th St](#))*
- *Wednesday Dinner, Washington Square Hotel*
- *Thursday Workshop, Dean's Conference Room, [4 Washington Square N.](#)*

NOTE: we will meet in the lobby at 8:30AM on Wed and Thurs morning to walk over together

Zoom: Remote participation will be available both days

Join Zoom Meeting
ID: 84764620229
Passcode: 646035

View conference details
keyboard_arrow_up
Hide conference details
(US) +1 689-278-1000

Passcode: 646035

Miro: https://miro.com/app/board/uXjVNJVBE0I=/?share_link_id=791672457239

Hotel: Washington Square Hotel (https://washingtonsquarehotel.com/).
Do not book your own room! Meeting participants' hotel rooms will be booked by Sara Robinson at Columbia. Please let us know your arrival and departure dates.

Flights: There are many options to choose from.
- JFK
- LGA
- EWR

Travel Reimbursement Instructions

For those of you who will be seeking reimbursements here are instructions:

1. Send me PDFs of all receipts, please do not send them as one single PDF - Send each one as a separate document. If PDFs are too cumbersome, photos of your receipts are okay.
   a. Try to send everything in a single email if possible, but I understand if it's not possible.
2. Attached is an expense sheet for you to list your expenses. It is really just a simple list, and it is helpful so I don't miss anything.
3. It is best if the receipts have the last 4 digits of the credit card on them.
4. WE can reimburse per diem for meals - or you can submit receipts for meals, it is up to you. But if you use receipts, then I need the itemized receipt for the food.
5. Allowable costs include car/taxi, mileage, meals, airfare, lodging if you paid your own - if you have anything else let me know and I'll see if we can include it.

I think that about sums it up. Let me know if you have any questions.

Hope you all enjoy the meeting. (Workshop?)

All the best,
Sara

--
Sara Robinson
Geoinformatics/MGG
LDEO of Columbia University
61 Rt 9W, Palisades, NY 10964
srobin@ldeo.columbia.edu

Notes

12/06/23

Introduction
Jane and Kerstin
- Manhattan also known historically by Manahatta - tribal land is divided into 3 groups
- Lenape tribes and reconnecting to the land (land acknowledgement)

- Introductions of the group
    - Jane Anderson - NYU, Open Context
    - Kerstin Lehnert - IEDA2 and Astromaterials at LDEO
    - Erin Robinson - Metadata game changers - focus on sampling in field station
    - Natalie Raia - iSamples
    - Lindsay Powers - USGS
    - Kirin - NYBG
    - Rae - NYBG
    - Alex McAlvay- NYBG
    - Matthew Pace- NYBG
    - Peter Buttiegieg
    - Eric Crandell
    - Kenneth Walker - NYBG
    - Sean Cao - LDEO
    - Saebyul Choe - LEDO
    - Leigh - Open Context
    - Sarah Kansa - Open Context
    - Neil Davies - Co-PI, iSamples - U.C. Berkeley
    - Maheata Davies - Gump Field Station (working on permits)
    - Chris Meyers - Smithsonian
    - John Deck - UC Berkeley, BioCode, GEOME, Local Context
    -

[RCN Background and Work](#)

- Driver: FAIR samples and to increase multidisciplinary use of samples that are consistent with sample documentation, sample identification, access policies, and sample citation
- We need a common voice across the sample community
- iSamples (internet of samples) is the software/tech side of the Sampling Nature RCN for FAIR samples
- Goals:
    - Establish interdisciplinary community
    - Clarify incentives and overcome barriers
    - Formalize policies and standards for sample gathering, best practices
    - Raise awareness for general public on importance
- Deliverables:
    - Metadata standards
    - White paper on policy recommendations
    - Training materials for metadata use and dissemination of policies
    - Communication through social media campaigns on importance on material samples
    - New collaborations within the community
- 4 workshops in total, this one is the 3rd one
    - 1st - At Moorea focus on water, rivers, coasts, and oceans
    - 2nd - At Arizona focus (loosely) on food and health, what sample metadata mean to each domain
    - 3rd - At NYU - this one!
    - 4th - At Smithsonian (future)
- Cross-Community Coordination
    - Input to ESIP Physical Samples Cluster
    - Sampling Nature Webinars

- ○ Presentations at RDA Plenaries
- ○ Coordinated effort with GBIF, GSC, IPDS etc.
- ● Looking Forward
  - ○ AGU participation
  - ○ Policy on Curated and Un-curated collections
  - ○ Adoption of CDIF (Cross-Domain Interdisciplinary Framework)

Use Case Presentation - [NYBG](#)
Alex McAlvay from NYBG
- ● Ethnobotanist
- ● 8 million + dried plants specimens that were collected before IRBs with history including social and ethical background/issues
- ● Matthew:
  - ○ One thing for researchers to have conversation on CARE and FAIR samples, but admin do not understand and need to be educated on it
  - ○ What are ethnobotanical and economic botany collections?
    - ■ Specimens and artifacts derived from plant products
    - ■ Obtained from field, markets, industrial sources
    - ■ Has collections in jars, dried, baskets etc.
    - ■ These specimens have metadata on common names, local/traditional uses etc.
    - ■ Need more work needed on labels and keeping records tidy
    - ■ Historically, biocultural collections went through many editions:
      - ● NYBG Museum of Economic Plants (by Rusby)
      - ● Drug Museum (by Rusby)
      - ● Columbia College of Pharmacy (10k specimens)
      - ● IEB Economic Botany/Ethnobotany (Mexico, South America, etc)
  - ○ NSF grant awarded for biocultural specimens and curating (physical and digital preservation)
  - ○ How do you deal with samples and field notebooks? Such as moving them to different buildings, broken containers, bad labels, digitizing etc.
  - ○ What do you do with throwing away collections/deaccessioning?
    - ■ In 2013, Economic Botany collection was moved to a storage unit to bring in a new collection
  - ○ Since Jan 2023 - a lot of work was done by the team
  - ○ Q - What does acquiring collections look like?
  - ○ Interesting to see the layering of different decisions and infrastructures
- ● Kenneth - project manager of NYBG
  - ○ Currently have IEB teaching collections from medicinal research
  - ○ Ethical issue even when the example was purchased locally;
  - ○ Giving digital sovereignty over the collections and continually conversing and doing collections management (using blur tool to remove and how to communicate the wishes and the rights
  - ○ Economic botany standard and re-curating with citizen science activities
  - ○ Working with community gardeners etc. to look into repatriation of specimens
  - ○ Working with ethnic groups on how immigrants and people use species of plants historically and culturally
  - ○ Harding Lab at the Institute of Economic Botany houses different collections and programs
  - ○ Contemporary collections incorporate collections from NIH, National Cancer Institute etc.

- - - ■ Samples have NIH codes to be able to trace the chemical components from the plants
      - ■ Some programs are documentation on ethnomedical practices by communities
        - ● Ex. Medicinal herb costus speciocus for diabetes use in communities
    - ○ Goal is to identify collections and acknowledge the rights by making them FAIR
      - ■ Samples collected with prior consent and documentation
      - ■ How do groups get digital sovereignty for the sample metadata?
      - ■ Working with curators to find what communities want publicly accessible and choosing what gets published or not
        - ● Issue is in sharing samples and emphasizing the community needs to other collaborators and institutes who do not follow these guidelines
    - ○ Collections and Rice Diversity
      - ● Philippine biocultural collection on rice varieties

- ● Alex
  - ○ Afterlives of collections beyond identification of plants
    - ■ Different chemical and biological research and work can now be done beyond identification such as pathogens, genomes etc.
    - ■ Do communities want to see this data/information on historic plants and what are the consequences and things to consider?
  - ○ Project working with Shinnecock Nation (Long Island - near Hamptons)
    - ■ Dr. Kelsey Leonard and team
    - ■ Vegetation and habit mapping
    - ■ Ethnobotanical interviews with elders and other knowledgeable experts on uses, historic context of plant samples
    - ■ Burning off and cultural fires on impacts on native plants and land use/preservation
      - ● Used for tick, mosquito control, berry growth, game and habitat etc.
      - ● How vegetation communities change pre and after burning
      - ● Burn interviews - found a person who was conducting a burn and now he is going to lead a workshop on burns youth
      - ● How to move forward with ethical voucher statementsGoals: create a community guidebook, storymap, interpretive signage, and vegetation map
  - ○ Challenges and opportunities
    - ■ Goal: Keep knowledge and specimens together to maintain cultural context
    - ■ Who are the end users and where do we store the samples? How do you maintain how other institutions store and curate collections?
    - ■ Allow communities to curate samples in community herbariums (if this is feasible)
    - ■ Collections are always in the making and on-going curation/re-curation
  - ○ Ethical
    - ■ Challenge with voucher metadata migrating to GBIF - make clear to communities what you can and can't do. NYBG communicates that the metadata may not have a label.
    - ■ If there are issues don't publish until they can work out.
  - ○ Legal
    - ■ Lack of legal standing - currently extra-legal
    - ■ Policies
  - ○ Social

- - - ■ Collection built for research and not for the community; Institutional - doesn't matter if community is interested in same pre-conceived
      - ■ Linguistically diverse and not speaking same language
      - ■ Ethnobiology has a colonial past (current bioprospecting, biopiracy)
      - ■ Is co-curation enough
      - ■ Community has certain wishes
  - ● Questions?
    - ○ Do you use EMU? Yes, since the mid-90s. NMNH now applying in rights field - share practice w Te Papa expand and demonstrate
    - ○ Pier- identify value chain details and common features of where applications converge or collide - esp in relevance to ABNJ
    - ○ Concept of community: what does that mean and how do you envision the interactions?
      - ■ Resources to manage collections, working with other groups
      - ■ People need to retire! Changing the framework of how people view the collections
      - ■ There is a consortium of ethnobotany collections where they meet to discuss the cultural and voucher specimens for a paper
      - ■ Direct outreach to community gardens and why ethnobotany samples can be important to them that is different than from the research perspective
      - ■ Expanding accessibility -language and communication
      - ■ Historically plant use and data collection is based on "how is it beneficial to me" mentality, as part of colonization
    - ○ How do you work with other groups and agencies for seeds?
      - ■ Collab for world vegetable center to help identify what crops they have in their collection
      - ■ USDA in Geneva, NY for cannabis
      - ■ Rice is not viable but there is still more to learn such as sequencing -stress genes, DSI issues
        - ● Advertising and communication to multiple academic groups and also to the non academic community
        - ● Enabling databases to be machine readable and searchable
        - ● Peter: Federate diverse datasets where portal is accessible to everyone; Changing the behavior of the internet
    - ○ Data sovereignty: is this an institutional policy? How do you stop people from using sensitive data?
      - ■ Currently extralegal
      - ■ Need a system; possibly a notice that warns people/user on what they can or cannot do with data
    - ○ Question from Jane on ethical rights and NYBG workflow
      - ■ Community decides - they have the opportunity to make a choice
      - ■ NYBG chooses what to publish and what not to publish online in systems such as GFIC, iDigBio
      - ■ Beware of FAIR and CARE -washing, talk and no action, lip service aspects
      - ■ Thinking about resources: best they can do is put notices about availability of data but it is harder to reach out to all communities
      - ■ Communities are also always being made and changing in urban settings. What does sovereignty mean in changing environments and who/group is in charge of the decision making?
      - ■ Documenting harm or benefit exposes lessons learned

- ○ How do other groups do outreach? Come back to the ways group outreach do.

Food for Thought
- ● Benefit sharing: who gets the credit and rights to plants that are used in medical (large pharma and biomedical companies)
  - ○ I.e. Madagascar periwinkle used for breast cancer chemo
- ● Ethical, legal social gets bundled up together, what do the 3 words mean and how can it be used for each sample type?

Transfer of collections? What happens to data when sample is transferred; used specify - to get data into EMu

How does USGS define what a Native American collection? USGS didn't have a policy for first 150 years. - Legal

[Participant Samples Slides](#)
- ● Erin Robinson - Arizona State Tree Ring Lab samples
- ● Lindsay Powers (USGS) - collections have been collected before federal or indigenous lands were created
  - ○ Federal government says the collection should be curated by those who own or are in charge of these lands
  - ○ USGS pushing back stating samples should be kept together, for example fossil collection sent to Smithsonian (legal obligation that supersedes the fed gov).
    - ■ Offering samples to Smithsonian first (national center) prior to other repositories
    - ■ Using SPECIFY to get the information and data into EMU
  - ○ Data is published and accessible
  - ○ Questions: what is a Native American collection?
    - ■ USGS is still working through collections management and policies to hold people accountable
    - ■ Why does the Fed not mandate or have regulations?
    - ■ "Empire building"
      - ● Academic publishing - not sharing data b/c of open science get more funding from gov if they are building a collection
    - ■ What is the ethical issues in collection at USGS
      - ● Not sure what is available at USGS and collections are distributed across different offices/scientists/labs etc. Care and understanding on what they have is the biggest ethical issues
    - ■ Do interagency offices discuss collections management?
      - ● Focus right now is on biological specimens and Nagoya protocols rather than domestics collections.
      - ● Chris: Language in CHIPS act is currently a big topic that needs to be flushed out. NSF is leading discussion to broader collections and disciplines
- ● Neil Davies - geoomic standards consortium and Island sampling day
  - ○ Island Sampling started from the top of the reef to ocean - sediment and soil samples including coral reefs
  - ○ Effort to bring field experience and understanding on how samples are collected or data people

- ○ Legal issue: this is access to biodiversity, it falls into Nagoya protocol.
    - ■ Application for authorization for collection is now needed form gov. Previously just a notification that a sample will be collected was sufficent. Does Neil's group get grandfathered in or do they need to get authorization and declaration saying it's okay to collect data?
    - ■ Gov = environmental and fishier organizations in France/French Polynesia
  - ○ Peter: as sovereign states find data is important for them, they want more control. This impacts data sharing due to compliance if data is published already. Tricky situation for collectors and researchers on legacy collections/samples vs. new ones. Limited infrastructure
- ● Eric Crandell - Parachute science, "losing Nemo"
  - ○ Sampling throughout the South Pacific as a grad student
  - ○ Created DIPnet - moderate impact of parachute science, taking and collecting samples without sharing the information back to the community
  - ○ Helped develop GEOME to make data available to students and researches available to developing nations
  - ○ Currently getting metadata in papers into GEOME to make them more FAIR. Including processed data and not just raw data, so researchers who can't get high computing can still access data
  - ○ Questions:
    - ■ Erin: How did you figure out what stats are important and significant?
      - ● Using geode and communicating a way to get importance and use to researchers etc.
- ● Sarah Kansa - Open Context - mapping weaving tools at Poggio Civitate
  - ○ Legacy collections and current excavations
  - ○ How to develop best practices and digitize and curate sample and metadata
  - ○ Found spatial patterns once they gave samples geolocation
  - ○ Weaving tools plots - spindle spools, roceti (?) used for larger looms
    - ■ More elite areas there are more spindles and common sites more looms spools
  - ○ Social and ethical: working with new excavations vs. legacy and getting samples descriptions and metadata when samples are collected rather than later on by other researchers (cleaning up after yourself)
  - ○ Question:
    - ■ Where do the collections live?
      - ● Dependson the dig and countries, but this specific collection is underground, held by the city and local government.
      - ● Data: countries like Greece are more concerned with documentation and data, places like Turkey are not. Depends on who is in charge and what they decide to do
    - ■ Is there a culling/assessment tools to figure out better workflow on prioritization of samples and/or data (common practices)
      - ● Depends on the samples, location, etc. Looking through field notebooks but process is not standardized
      - ● USGS is also facing similar problems: workflows and practices on orphaned collections to see how much metadata is needed to make samples valuable to other
      - ● Sample management plans needed for collections
      - ● Userside decision on what is available or not
- ● Saebyul Choe

- ○
- Alex and Matthew - NYBG on data sovereignty and sensitive information on plants
  - ○ Historic documentation from collector about plant sample and cultural uses etc.
    - ■ What do with this document in regards to validity of the sample and context
    - ■ Questions on who to contact and if these contacts are current
      - ● Led to sending mail and eventually got to lawyer who questioned where in public institutional domain of NYGB states legal, ethical, and social commitment
  - ○ Questions:
    - ■ Is there an ethical review process?
      - ● NYGB still does not until more recently. Currently working with admin to create and develop one.
      - ● Hurting the public by limiting information and not being transparent about the data.
- Leigh Lieberman - Open Context on fieldnotes and other records
  - ○ Art and Archeology at Princeton working on the archives in the department
  - ○ Howard Crosby Butler collection - who worked in the middle east, specifically Syria
  - ○ Howard's work needs to be digitized and curated
  - ○ This work was important because it impacted Wilson's policies for the middle east
  - ○ How can we use legacy collections as use cases for FAIR data; how do you deal with sensitive data; how can we use these materials in the humanities?
  - ○ Questions: what kind of expertise is needed to decipher old texts?
    - ■ Use tools like transcribus to get handwritten text into digital text
    - ■ Collective Access
- Natalie Raia - University of Arizona on helicopter research in New Caledonia
  - ○ Working with 200 samples on metamorphic rocks for geochemistry
  - ○ New Caledonia is a French territory with political contest - really important nickel mining
    - ■ French mining companies, even Tesla is now involved
    - ■ Issue in formal permitting infrastructure and oversight by government (not local indigenous groups)
  - ○ How do researchers work and engage with local communities
  - ○ Emphasis needed on community curated resources - who to contact prior to collection, what cultural nuances need to be learned and considered, understanding historic context of people and groups - geology and colonial history
  - ○ There is no CBD for geology
  - ○ Issue in sampling in the same site all the time - offending locals but also getting samples there already is a abundance of
    - ■ Difference in really needing samples again vs. metadata was just not carefully recorded at the time of initial sampling

**RCN to provide different communities on how to ethically sample - like identify and establish who the leaders are and work on getting permission. Tribal liaisons

**Issues in funding agencies to make sure data management plans are included and implemented

**Data is not private - needs to be made public

**Develop partnerships with communities and groups

Presentation: NYC Archaeological Repository (Amanda Stuphin)
Overview
Starting slide: Image of World Trade Center
- Sample of ship from dig is with the underwater archeology group at Texas A&M

Landmarks and Preservation Commission
- Public is afraid we are losing the historic sections of NYC (i.e. Penn Station)
- Repository started in 2014 - collection of archeological artifacts that mostly that were done within the past 40 years and older collections (that may or may not be recorded well)

Reasons for Archeology (and Relevance)
- Environmental reviews (esp in the city): archeologists look at history, field testing, curation plan, and community access to the knowledge/information
  - Have guidelines and way to implement as a regulatory facility and repository
  - Applicant pays for the archeological work
  - List of people/contract archeologists who have qualifications are available, however the NYC Archeological group cannot recommend one specific person/group.
  - Information, reports, metadata is available through repository and NYC gov website, but some information can be held and others are internal (such as reviews on sensitive information)
  - Work is credited to the contractors, however the data is not theirs and they cannot withhold information
  - Question: what does consultation look like?
    - Depends on the resource. Ex. burial grounds (many in the city). There was no law protecting these sites until this year.
      - Unmarked burial law - burial committee decides what should happen on native burials, but limited to certain groups… impact is still in assessment
    - Ask descendants and community to work with them

Sampling in Archeology
- What does sampling mean?
- Process of sampling → Site sampling, specific locality sampling, specimen sampling (what is kept and what is not), lab experts sampling, and finally curation sampling - screens out samples through each layer/workflow

Archeology NYC
- Indigenous people and sites
  - Staten Island - Munsee (consulting with Delaware nation and other indigenous groups)
- Historic Sites
  - Southern Hanover Sq - historically New Amsterdam, Dutch settlement
  - Jugs from Stadt Huys project - lower Manhattan have dutch design made in NYC/local
  - Sugar mold and punch bowl - sugar bowl started conversation of history of sugar in NYC and social, ethical, and political issues (slavery)
  - African Burial Ground: 400+ burials, stressors for physical labor, etc.
    - Shroud pins - pin shroud on burial (most common artifact)
    - If you want to visit - 290 Broadway
  - Question: what happens to biological collections for ex. carrot flowers found in one of the burials?
    - Can apply to study them
  - Thomas Commeraw - freed slave who started a pottery business

Types of Samples and What They Tell Us
- Chert gunflint
- Animal bones - how people used animals and what animals existed

- Shells - oysters in NYC was super popular and important

Current Initiatives
- NAGPRA Compliance - indigenous material
- How to deal with biocultural collections such as body remains where information on where they came from is limited or poorly recorded
- Making city collections publicly accessible
    - Catalog record for public collections - some have images, others do not
    - Use QI as their sample management tool
- Artifact catalog terminology for controlled vocabularies and best practice
    - Working with experts and then public before publishing
    - Will this catalog machine be readable? Currency is a pdf format and yes, eventually
    - Networking Archeological Data Communities (NADAC) working group to introduce FAIR data/metadata
        - Working on project: 7 Hanover Sq for legacy collection of images, documents, etc.
        - Establish data protocol
        - Clean/Complete data (ex. Info from punch cards from 70s etc)
        - Incorporate archive
        - Make Data FAIR
            - What does interoperability mean? How do you make sure that the information is translated moving forward between different groups and people
            - Part of this project is to use as training tool for next urban archeologists and getting information standardized and distributed to anyone who is working on similar projects or within this realm
            - Important to emphasize workflows will change and we need to adapt
        - Creating spatial data and map of where samples were found
- Making data FAIR
    - How do you share your implementation expectations with other parties?
        - Starting to begin figuring out implementation and understanding issues of what is the best practice
        - In historical archaeology the idea of FAIR is limited and still in the beginning stages

Book: Buried Beneath the City - CHECK IT OUT!

Mid-Point Synthesis
Conversations on what is coming up in our mind/reflections
- **Greater implementation of LocalContext**
- **Where are points to make most effective change of expectation within infra to have larger flow out into larger contexts**
- **Culture change and changing culture that values the correct practices**
- **Sample not taken - expose why you didn't sample (advocacy, activism and research)**
- **Does it produce harm/benefit and people who are impacted decide**
- **Frustrated that this conversation has been going for a long time**
- **Technical capacity to link the context around a sample, but if we do it what happens - who decides**
- **Role of institutions and professional societies and rules for collections being accessioned and how to ensure that collections don't end up abandoned**

- **Role of collections as tech advances - how does it change the value proposition**
- **Optional will be skipped**
- **Everything coming into the Smithsonian get's transacted**
- **Being mindful about how local communities see us**

12/07/23

Recap from Day 1

SBC: It was cool to see how other repositories/archive centers work with their samples (artifacts!) and more funding is needed to help curate and also bring the collections to the public. Seems like many bio based collections and programs have more focus on working with local communities/the general public than geoscience does

LAL: More money and more time would solve so many issues! (thinking of the work of the NYC Repository); meeting groups where they are, because total compliance often seems like an far off dream; need for small proofs of concept to advocate for the resources necessary to do this work

NR: I found the blurring tool from NYBG interesting as a first implementable step for some data systems; Three major realms of this work have surfaced for me- institutional policy (including supporting guidance from societies, organizations like ESIP), structural/functional within data systems but also funding and publishing processes, community education/communication; importance of consortia like NADAC in introducing best practices to new communities, and facilitating coordination and uptake of common practices (e.g., use of PIDs, common data standards)

ER: I liked seeing places other than museums that store samples; I thought it was interesting to see the connections between the NYBG and the NYC Archeologists

SK: Thinking about outcomes, especially around training materials… Who is the audience, the users of the samples/specimens/collections, what do they need, etc.… Lindsay pointed out that we need to get the attention of the funding agencies/panels. How are we reaching them and with what? Who else? Prof development but also public outreach. Where else can we make an impact?

LP: Policy helps as long as it is supported with resources. Lots of opportunities for knowledge/experience sharing. So important to engage with stakeholders where they are and how they need to be engaged with.

KL: yesterday gave us so many great stories to bring physical samples and collection into the spotlight
- Rich knowledge exchange is really important across domains;
- common challenges need an ongoing dialog - how will we continue after the RCN grant ends?
- Need to focus on products of the RCN - we have gained a lot of insights, need to turn them into resources for the community to support sample & collection management

SSC: Value chains are important! How do you balance the different social/ethical/legal perspectives of different groups on not only local/individual examples (like NYBG's work with indigenous people in the NYC-area) but also cross-domains (this workshop)? Especially when some of these value chains boil down to money/resources.

EC: I enjoyed hearing about so many other use cases for samples and their metadata and broadening my perspective. I am looking forward to discussing further how Local Contexts can be implemented by researchers in the field, and documented in metadata.

PLB: Given the existing work by Open Context for packaging (meta)data about samples in JSON-LD, I'd like to create an IOC-UNESCO ODIS pattern for samples using schema.org semantics with more domain-level semantics nested within as described in Section 3.3.3 of this recent WorldFAIR deliverable. I would like to cross-link this co-development with the CDIF Core patterns (very new, so some wrangling to do) and see if we can cross-link this to the Cultural Heritage Work Package in WF.

- https://github.com/Cross-Domain-Interoperability-Framework/cdif-core/issues/4
- https://github.com/iodepo/odis-arch/issues/376

CPM- legacy vs newly born and the changing terms of use and conditions pertaining to sample reuse or data sovereignty, legal, ethical and social - legal seemed to trump everything - concept of institutional risk vs reputational impact, the sample not taken, authority to enforce policy and practice - define and inputs/outputs and allow practitioner flexibility to use existing tools - just hit the interoperable targets, implementation challenges, old dogs new tricks, who brokers rights and for what class / category of material samples and their derived data or future use and data - concern about compliance "enforcement" - who is responsible for understanding terms as values might be changing (still too idealistic)

JBD- We saw a range of issues surrounding various rights of samples. We need to think about ways to clearly identify rights in a machine operable manner.   We need to coalesce on a metadata profile for samples so folks can use whatever identifier scheme they want.   We need a format for distributing metadata that includes digitally signing the actual data to track provenance.

ND - specimen, sample, artifact? How about a "curaton" (an item being curated)

Braden Tierney - Use Case Urban Swabbing
Four sampling (swabbing) sites, workflow on swabbing from field to lab
Shotgun sequencing - taking a snapshot of the DNA (microbs) present at the site during sampling
MetaSUB - Started in NYC → Goals: geospatial ecosystem maps in cities, anti-micorbial etc.
Global city sampling day at the end of June! Send out kits around the world and get samples back
Uses Kobo for metadata collection and data storage at GeoSeq
NGS on sewage (FL, WI, and NYC) - building dataset of 1000 targeted sequencing; mutation tracking
- Sampling was done manually, FL → U of Miami (hourly, daily, and weekly sampling)

Issues to think about related to privacy of people in the community
Prescription and microbes correlation
Correlation between microbes and zip codes (communities) from sewage
Journey to Mars! → figuring out how human physiology changes overtime in space
2FP Carbon Initiative → microbes for carbon capture in oceans; building a living database of microbes experiment
- Nagoya compliant - questions on ownership
- Cual-id - Globally Unique, Correctable, and Human-Friendly Sample Identifiers for Comparative Omics Studies
- UTEX - collection of algae and other microbes
- CitSci - citizen science organization

Wants standard log files. Could work with him on identifiers

Movie Time! (Short Film by Jane)
Making films help visualize importance of research and data, also the historical, cultural, and legal implications to communities
Are there other communities making labels that can be linked to their group and also maintain access to their data?
Need emphasis and push to funders - need for transparency

Workshop Goals

20

Topics to Divide and Discuss

Develop interoperability specifications

Create notices on how to register

iPlaces

Indigenous metadata bundles - Jane

- Based on workshop held in May: looking to two specific questions → data repositories for CARE principles, identifying metadata fields for inclusion at Open Context
- Bundle has 5 categories of topics that were found most relevant (governance, provenance, etc)
- RIGHTS Module → set of notices that were developed for institutions

Policies

Uncurated Collections

Interoperability Group

Training Materials Group - GET LIBRARIANS INVOLVED (and educators?)!

- Placement/Crosslisting for Materials
    - AGU data leadership pages > they already have some resources developed, but not many people know about or are using them (example here)
    - What is the reach for our own RCN website?
    - Professional societies like AGU and ICOM
        - Workshops at Conferences
        - Webinars pushed through the societies and main points about it summarized if people don't want recording
- Audiences
    - Museums/Institutions (future, higher level)
    - Researchers but with a focus on early career (/uncurated collections)
    - Data managers
    - Collections managers
        - Introduction to metadata standards
        - Introductions to persistent identifiers (ORCIDs, IGSNs, ARKs, RORs?)
            - What each is good for, how to get them, etc.
            - Getty ULAN is doing work similar to a ROR in cultural heritage LOD circles
            - Visual map of how systems are related in each discipline
    - Admin, but through interventions with the staff working on the collections
    - Publishers
        - Guidance for publication samples
        - Guidance for adoption of notices and labels
        - Pointing out successful use cases
            - Molecular Ecology cites GEOME as an example for data availability statements in their author guidelines
- Is there space for a longer run training/working group program?
    - Building a curriculum for instance: https://alexandriaarchive.org/nadac/
        - And, regarding tracking the outputs/future lives of participants in these programs, this white paper
- Training materials for the SPECIFIC open repositories
    - Videos are useful, especially for technical explanations
    - handbooks for internal use but those would be kind of a drag to read

- **Roadmapping** (group activities for engagement)
  https://sites.haa.pitt.edu/sustainabilityroadmap/
  - Here are the questions you need to ask yourselves, here's the case study of how it worked in instance 1, 2, 3; now go do the work; document your answers to that set of questions, make that as open and accessible as possible
  - Make it PUBLIC > accountability, general dissemination of information
  - **An output: Stories from RCN workshop visits or other examples with key aspects/things that have come up with pull quotes or videos from people**
- Wikis
- FAQs
- Forums (Fora?)

What outcomes do we want?
Collection of stories

GEOME - building trainers teaching others (Carpentries model)

Specifications working group / Breakout notes

Challenges: Open Context has a very abstract / generic global schema. Sites, objects (artifacts, ecofacts, specimens), archaeological contexts are all modeled in the same way in Open Context, but they will have variable properties. Also, data come from different sources (contributed by field researchers), so many attributes that may be useful to express in a common standard may be missing.

objects (artifacts, ecofacts, specimens) alignments with https://schema.org/Product
NOTE: All schema.org/Product attributes that DO NOT map to an Open Context equivalent are

NOTE: How would search engines interpret this? We don't want objects/artifacts aggregated in some sort of ecommerce setting.

Example animal bone specimen record from Open Context (in JSON-LD):
opencontext.org/subjects/c7680e87-a5c5-4624-9b73-3f29de9fdf29.json

schema.org/Product	Open Context
additionalProperty	Any "key" within an object of the oc-gen:has-obs list
aggregateRating	None
asin	None
audience	None

award	None
brand	None
category	category
color	If present, it would be in a key of an object within the oc-gen:has-obs list
countryOfAssembly	None
countryOfLastProcessing	None
countryOfOrigin	Second object in the oc-gen:has-contexts list
depth	If present, it would be in a key of an object within the oc-gen:has-obs list
funding	None
gtin	None
gtin12	None
gtin13	None
gtin14	None
gtin8	None
hasAdultConsideration	None
hasEnergyConsumptionDetails	None
hasMeasurement	If present, it would be in a key of an object within the oc-gen:has-obs list
hasMerchantReturnPolicy	None
height	If present, it would be in a key of an object within the oc-gen:has-obs list
inProductGroupWithID	None??
isAccessoryOrSparePartFor	None
isConsumableFor	None
isFamilyFriendly	None
isRelatedTo	If present, it would be in a key of an object within the oc-gen:has-obs list
isSimilarTo	If present, it would be in a key of an object within the oc-gen:has-obs list

isVariantOf	If present, it would be in a key of an object within the oc-gen:has-obs list
itemCondition	If present, it would be in a key of an object within the oc-gen:has-obs list
keywords	dc-terms:subject ?
logo	None
manufacturer	None
mpn	None
negativeNotes	None
nsn	None
offers	None
pattern	If present, it would be in a key of an object within the oc-gen:has-obs list
positiveNotes	None
productID	dc-terms:identifier
productionDate	In the "when" object of features / feature. Will typically be described as an interval time-range
purchaseDate	Totally not appropriate!
releaseDate	Is this like a publication date? If so dc-terms:issued
review	Maybe a peer review status (can be inherited from a project)
size	If present, it would be in a key of an object within the oc-gen:has-obs list
sku	None
slogan	None
weight	If present, it would be in a key of an object within the oc-gen:has-obs list
width	If present, it would be in a key of an object within the oc-gen:has-obs list
additionalType	If present, it would be in a key of an object within the oc-gen:has-obs list We use the CIDOC-CRM "Has Type" property

	here to indicate types according to some controlled vocabulary
alternateName	None
description	If present, it would be in a key of an object within the oc-gen:has-obs list
disambiguatingDescription	None
identifier	dc-terms:identifier
image	If present, it would be in a key of an object within the oc-gen:has-obs list
mainEntityOfPage	id (The URL/URI of the record in Open Context)
name	dc-terms:title
potentialAction	None
sameAs	If present, would be owl:sameAs
subjectOf	id (The URL/URI of the record in Open Context)
url	id (The URL/URI of the record in Open Context)

OpenContext example for

specimen/sample:
https://opencontext.org/subjects/ca1580d0-490e-4f53-bb90-d12f4933d1d5.json

Arch site:
https://opencontext.org/subjects/5d6b6454-017a-43c1-9f15-6dfe36c3558f.json

"when" describes when the thing was made and used, its lifetime

See the Datacite advice:
https://support.datacite.org/docs/local-contexts-notices-and-labels

TK labels / BC labels:

- https://scd.landcareresearch.co.nz/Specimen/CHR%20365035

Status:

Right now, looking at examples like this one in the Hub, the customised labels do not have a URL that functions as a PID. Once this is in place, we have a linked open data solution that can be used to declare the label to modern data-on-the-web solutions. Below, we're using placeholders for now

It's likely that the labels will have different domains - some are best as values of provenance, others licenses, others constraints. A careful mapping of labels to the right property should be done.

Guidance: when embedding a BC or TK label:
- In the keyword property of your specification, add a defined term that includes the PID/URL of the instantiated TK or BC label [https://localcontexts.org/labels/biocultural-labels/] [https://localcontexts.org/labels/traditional-knowledge-labels/]
- https://localcontexts.org/support/taxonomy/

- In addition, add a additionalProperty or equivalent for each of the labels, using the appropriate URL to identify and qualify the property. The range of the property would be the URL/PID of the instantiated label

Keyword example

```
{
  "@context": {
    "@vocab": "https://schema.org/",
  },
  "@type": "Map",
  "@id": "https://example.org/id/XYZ",
  "name": "Name or title of the document",
  "description": "Description of the map to aid in searching",
  "url": "https://www.sample-data-repository.org/creativework/map.pdf",
  "identifier": {
    "@id": "https://doi.org/10.5066/F7VX0DMQ",
    "@type": "PropertyValue",
    "propertyID": "https://registry.identifiers.org/registry/doi",
    "value": "doi:10.5066/F7VX0DMQ",
    "url": "https://doi.org/10.5066/F7VX0DMQ"
  },
  "keywords": [
```

```
        {
            "@id": "http://myPID.org/bc-provenance-customisation1",
            "@type": "DefinedTerm",
            "inDefinedTermSet": "http://myPID.org/bc-community1-instance.json",
            "termCode": "BC P Com1",
            "name": "Biocultural Provenance Com1"
        },
        {
            "@id": "http://myPID.org/tk-seasonal-customisation1",
            "@type": "DefinedTerm",
            "inDefinedTermSet": "http://myPID.org/tk-community1-instance.json",
            "termCode": "TK S Com1",
            "name": "Traditional Knowledge Seasonal Com1"
        }
    ],

additionalProperty [

    {
        "@type": "PropertyValue",
        "propertyID": "http://pids.localcontextshub.org/LC_0000023",
        "name": "Biocultural label"
        "value": {
          "@id": "http://myPID.org/bc-provenance-customisation1",
          "@type": "DefinedTerm",
          "inDefinedTermSet": "http://myPID.org/bc-community1-instance.json",
          "termCode": "BC P Com1",
          "name": "Biocultural Provenance Com1"
        },
        }




        {
        "@id": "http://myPID.org/tk-seasonal-customisation1",
        "@type": "DefinedTerm",
        "inDefinedTermSet": "http://myPID.org/tk-community1-instance.json",
        "termCode": "TK S Com1",
        "name": "Traditional Knowledge Seasonal Com1"
    }



]
}
```

27

It is vital that the date of the creation of the metadata record into which the labels are embedded is included in the record - if the label definitions change, that date is how you'd walk back to the right version that applied to the thing the record is about at the time

Permits are often not included in metadata , they are buried in a paper or a document, we need to include in the recommendations athat data without a property pointing to a permit or permission are to be either excluded or flagged as of concern

Guidance: when embedding a BC or TK notice:

- The notice should be typed as a schema:DigitalDocument or equivalent
- The Typed Notice should have a URI/DOI at the instance level
- The typed notice should be a value of a property such as schema:publshingPrinciples, this can be at the Organizational or Project level too.

Concern about the use of "Product"-- even if it's listed as "free", there's a concern about these being cultural heritage / artifacts– risky to associate them with the term

Could Schema make a subtype that's more appropriate?

View all the Labels that Whakatohea Labels in the Hub and the projects -
https://localcontextshub.org/communities/view/22/

Article with TK Notice