OSU-MGR
Data Discoverability

IGSN’s QR Code Labelling
Online Holdings & Data Portal

Robert Hatfield, Cara Fritz, Maziet Cheseby, Rupert Minnett, Anthony Koppers, Bobbi Conard, Kevin Konrad, Joe Stoner & our team of hardy, refrigerated space labelling, workers and undergads!
Goals & Aims

(Issues & Challenges)

- To move away from a paper analog system to a digital (discoverable) collection
- Embrace the use of unique identifiers (e.g. IGSN)
- Easier access for the wider research community (extra-OSU)
- Make day to day workflow easier (controlled vocabulary, reduce errors, automation)

- 200+ cruises: legacy data, inconsistent formats, incomplete meta data
- Diverse collection: Rocks, marine & lake sediment cores, drill cores...
- Human (intuitive) and machine readable
- Using software familiar to most users of the repository

- Information hosted online and also on the end of a d-tube!
Goals & Aims

(Issues & Challenges)

- To move away from a paper analog system to a digital (discoverable) collection
- Embrace the use of unique identifiers (e.g. IGSN)
- Easier access for the wider research community (extra-OSU)
- Make day to day workflow easier (controlled vocabulary, reduce errors, automization)

- 200+ cruises: legacy data, inconsistent formats, incomplete meta data
- Diverse collection: Rocks, marine & lake sediment cores, drill cores...
- Human (intuitive) and machine readable
- Using software familiar to most users of the repository

- Information hosted online and also on the end of a d-tube!
Three Pronged Approach

- **IGSN**: Unique identifier for each core section and sample
- **QR codes**: Sample labels encoded with sample metadata and IGSN
- **Database**: Contains the metadata, IGSN, and QR code and accessible through an online portal

www.osu-mgr.org/collections/
IGSN

- Already exist - SESAR
- However, potentially greater flexibility by determining our own unique identifier
- Requirements
  - Be unique!
  - Human & machine readable
  - Can intuitively link parents (cores) and daughters (samples)
  - Can be assigned offline before, during, or after collection - e.g. at sea
  - Follows the IGSN syntax requirements ("A-Z", "0-9", ".", ",") and < 32 characters

- Parent IGSN: Utilize and exploit the existing Cruise-Core-Section naming scheme already used by many repositories - “IGSN Short Form”
- Daughter IGSN: Append the parent IGSN with a PI identifier and unique sample code - “IGSN Long Form”
**IGSN - Short Form**

- **Parent IGSN**: Utilize and exploit the existing Cruise-Core-Section naming scheme already used by many repositories.

### Cores

**OSU-EW0408-79JC-10W**

<table>
<thead>
<tr>
<th>(i)</th>
<th>(ii)</th>
<th>(iii)</th>
<th>(iv)</th>
<th>(v)</th>
<th>(vi)</th>
<th>(vii)</th>
<th>(viii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>“OSU”</td>
<td>“XX[X]”</td>
<td>“##”</td>
<td>“##”</td>
<td>“#”</td>
<td>“#”</td>
<td>“#”</td>
<td>“X”</td>
</tr>
</tbody>
</table>

- **(i)** “OSU”
  - Oregon State University allocating agent prefix
- **(ii)** “XX[X]”
  - Two to three letter code identifying vessel (EW = Maurice Ewing)
- **(iii)** “##”
  - Last two digits of the year the coring cruise started (04 = 2004)
- **(iv)** “##”
  - Two digit month the coring cruise started (08 = August)
- **(v)** “#”
  - One to three digit identifier of the core name or number (79 = core 79)
- **(vi)** “#”
  - One to three digits identifying dredge number (2 = second dredge)
- **(vii)** “X”
  - Core section (A = archive half; W = working half; R = whole round)
- **(viii)** “[X]”
  - Optional single character denoting a subsample (A = first subsample)

### Rocks

**OSU-RR1504-D2-4A**

<table>
<thead>
<tr>
<th>(i)</th>
<th>(ii)</th>
<th>(iii)</th>
<th>(iv)</th>
<th>(v)</th>
<th>(vi)</th>
<th>(vii)</th>
<th>(viii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>“OSU”</td>
<td>“XX[X]”</td>
<td>“##”</td>
<td>“##”</td>
<td>“#”</td>
<td>“#”</td>
<td>“”</td>
<td>“[X]”</td>
</tr>
</tbody>
</table>

- **(i)** “OSU”
  - Oregon State University allocating agent prefix
- **(ii)** “XX[X]”
  - Two to three letter code identifying vessel (RR = Roger Revelle)
- **(iii)** “##”
  - Last two digits of the year the dredging cruise started (15 = 2015)
- **(iv)** “##”
  - Two digit month the dredging cruise started (04 = April)
- **(v)** “#”
  - One digit identifying the sample type (D = Dredge; G = Grap; R = ROV)
- **(vi)** “#”
  - One to three digits identifying dredge number (2 = second dredge)
- **(vii)** “#”
  - One to three digits denoting sample number (4 = fourth sample in D2)
- **(viii)** “[X]”
  - Optional single character denoting a subsample (A = first subsample)

**Note 1**: Multicores are slightly different for (vi) because the multicore tube number follows the core type. The Multicore syntax for (vi) is “XX#” (MC3 = multicore tube number 3).
IGSN - Long Form

- Daughter IGSN: Append short form with a PI identifier and unique sample number

  **OSU-EW0408-79JC-10W.SM101**

  (i) **Parent IGSN**  IGSN short code of parent core section
  (ii) “XX”   Two character PI identifier code
  (iii) “###”   Up to three digits that are unique to the PI code within a core section

- PI Identifier: Constructed from the PI’s initials
  - e.g. SM = Sarah Morgan
  - If Steve Marshall requested samples we use first two digits of surname “MA”
    - If not available then we use alternative surname combinations e.g. “MR”, “MS”, “MH”, “MA”, “ML” then first name combinations “ST”, “SE”, “SV”, failing that a number, “S1”, “S2” ... “S9”, “S0”, then a random available assignment.

- Up to three digits: Designate unique samples within the core
  - E.g. if Sarah Morgan takes 101 samples from EW0408-79JC-10W these would be called SM1, SM2 ... SM10, SM11 ... SM101 etc.
  - After 999 we replace the first digit with an alpha character e.g. A1, A2 ... ZZ9 (+2574 identifiers)
  - After 3573 we replace the second digit with an alpha character e.g. AA1 ... AA9, AB1, AB2 ... ZZ9 (9657 unique combos)
OSU IGSN FORMATT

OSU-EW0408-79JC-10W
OSU-EW0408-79JC-10W.SM101

Characteristics

- Unique at the core section and individual sample level
- Intuitively readable
  - Core section (short form) differentiated from individual sample (long form)
- Follows IGSN syntax requirements
  - Longest short IGSN - 22 characters
  - Average short IGSN - 18 characters
  - Long IGSN plus 4-6 characters
    - Long - 28, Av. - 24

Additional Advantages

- Can be seamlessly assigned “offline” at sea during normal curation process
- Short IGSNs can be used (and historically have been used) as identifiers to track cores in the repository
  - Database
  - QR codes
QR Codes
OSU-MGR label applied to d-tube, end cap, and core liner

- Cruise-Core-Section Name
- Core interval (cm)
- Data Link (left QR)
- Metadata (right QR)
- Space for alternate core name

EW0408-79JC-10W
Core Interval: 1279 - 1429 cm

Alternate name (if applicable) (with space over multiple lines)
Meta Data Link (static information)

- IGSN: OSU-EW0408-79JC-10W
- Lat: 59.5357
- Long: -141.7609
- Water Depth: 158m
- Total Sections in Core: 12
- Total Length: 1724cm
- Section Length: 150cm
- Top: 1279cm
- Bottom: 1429cm
- PI: Alan Mix
Data Link (non-static information)

- [Link](http://osu-mgr.org/OSU-EW0408-79JC-10W)

**IGSN:** OSU-EW0408-79JC-10W

<table>
<thead>
<tr>
<th>Parent Core:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material: Marine Sediment Core</td>
</tr>
<tr>
<td>Core Name: EW0408-79JC</td>
</tr>
<tr>
<td>Core Type: Jumbo Piston Core</td>
</tr>
<tr>
<td>Core Length: 1724 cm</td>
</tr>
<tr>
<td>Total Sections: 12</td>
</tr>
<tr>
<td>Latitude: 59.5332°</td>
</tr>
<tr>
<td>Longitude: -141.7609°</td>
</tr>
<tr>
<td>Water Depth: 158 m</td>
</tr>
<tr>
<td>Research Vessel: Maurice Ewing</td>
</tr>
<tr>
<td>Date of Collection: 2004-09-15</td>
</tr>
<tr>
<td>Cruise PI: Alan Mix</td>
</tr>
<tr>
<td>Cruise PI Institution: Oregon State University</td>
</tr>
<tr>
<td>Cruise PI Email: <a href="mailto:amix@coas.oregonstate.edu">amix@coas.oregonstate.edu</a></td>
</tr>
<tr>
<td>Alt. Cruise Name:</td>
</tr>
<tr>
<td>Alt. Core Name:</td>
</tr>
<tr>
<td>Location: OSU Marine Geology Repository (99.1.3.2)</td>
</tr>
<tr>
<td>Location Contact: <a href="mailto:corelab@coas.oregonstate.edu">corelab@coas.oregonstate.edu</a></td>
</tr>
</tbody>
</table>

**Section Info:**

- Section: 10
- Section Half: Working
- Section Interval: 1279 - 1429 cm
- Notes: |

**Data Available:**

- Coring Data Sheet: [Link](http://osu-mgr.org/OSU-EW0408-79JC-10W/CoringDatasheet)
- Core Description: [Link](http://osu-mgr.org/OSU-EW0408-79JC-10W/CoreDescription)
- Line Scan Image: [Link](http://osu-mgr.org/OSU-EW0408-79JC-10W/LineScan)
- MST Data: [Link](http://osu-mgr.org/OSU-EW0408-79JC-10W/MSTData)
- XRF Data: |
- CT Scan Data: [Link](http://www.ngdc.noaa.gov/peoamples/cruise.jsp?cr=EW0408&inst=OSU&shp=4&source=Jeweling)
- Link to Publications: [Link](http://osu-mgr.org/OSU-EW0408-79JC-10W/Publications)
IGSN: OSU-EW0408-79JC-10W

Parent Core:
- Material: Marine Sediment Core
- Core Name: EW0408-79JC
- Core Type: Jumbo Piston Core
- Core Length: 1724 cm
- Total Sections: 12
- Latitude: 59.5357°
- Longitude: -141.7609°
- Water Depth: 158 m
- Research Vessel: Maurice Ewing
- Date of Collection: 2004-09-15
- Cruise PI: Alan Mix
- Cruise PI Institution: Oregon State University
- Cruise PI Email: amix@coas.oregonstate.edu
- Alt. Cruise Name:
- Alt. Core Name:
- Location: OSU Marine Geology Repository: [93.1.3.2]
- Location Contact: correlab@coas.oregonstate.edu

Section Info:
- Section: 10
- Section Half: Working
- Section Interval: 1279 - 1429 cm
- Notes:

Data Available:
- Coring Data Sheet: http://osu-mgr.org/OSU-EW0408-79JC-10W/CoringDataSheet
- Core Description: http://osu-mgr.org/OSU-EW0408-79JC-10W/CoreDescription
- Line Scan Image: http://osu-mgr.org/OSU-EW0408-79JC-10W/LineScanImage
- MST Data: http://osu-mgr.org/OSU-EW0408-79JC-10W/MSTData
- XRF Data:
- CT Scan Data:
- Link to NGDC: http://www.ngdc.noaa.gov/geoamples/cruise.jsp?cruise=EW0408&inst=OSU&site=Maurice%20Ewing

Additional Info
- Material: e.g. Core, Dredge, ROV, Grab sample
- Core Type
- Research Vessel
- Date of Collection
- PI contact info
- (Alternate name/info - optional)
- Location in repository (Rack 93, Row, 1, Space 3, Position 2)

http://osu-mgr.org/OSU-EW0408-79JC-10W
Architecture: Server Organization

Data Available:
- Core Description
- Coring Data Sheet
- Line Scan Images
- MST Data
- CT Scan Data
- Link to NGDC
- Link to Publications

Data Accessibility:
- URL - IGSN - Data Type
- Multitude of data types
  - .pdf, .jpeg, .tiff, .txt, .bmp etc

Core Description - .ai and .pdf of samples taken
Coring Data Sheet - .pdf of original coring data sheet
CT-Scans - .dicom and .jpeg/.tiff files if available
IGSN Sheets - IGSN Output Sheet
Line Scan Images - .bmp/jpeg/tiff files if available
Miscellaneous - anything else/other notes
MST Data - MST calibration, raw, and output (calibrated) files
NDGC Upload - data file uploaded to IMLGS
Publications - pdfs of publications resulting from the cores
Samples Taken - list of samples taken from the cores
XRF Data - text file detailing measurements made
EW0408 Meta Data - excel file containing section meta data
Database: Metadata & Samples Distributed

- User friendly Excel based database
- Controlled vocabulary
- Drop down selection boxes constrain available options
- Makes for more efficient searching and less errors
- VBA coded to provide automatic error checking and automatically writes URLs for ‘available information’

<table>
<thead>
<tr>
<th>ISGN</th>
<th>IQA</th>
<th>CRUISE</th>
<th>(Cruise Name or Program Name)</th>
<th>RV Name</th>
<th>Deployment or Date</th>
<th>Material</th>
<th>Recovery Method</th>
<th>Core Collected</th>
<th>Collected</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Water Depth</th>
<th>Area</th>
<th>(Place Name)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8700</td>
<td>OSU</td>
<td>EW509B</td>
<td>79: NC 70W</td>
<td>Maurice Ewing</td>
<td>2004-09-15</td>
<td>15:45</td>
<td>58.5345</td>
<td>141.760</td>
<td>158 M3</td>
<td>Malapaya Shell/1n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8710</td>
<td>OSU</td>
<td>EW509B</td>
<td>79: NC 64W</td>
<td>Maurice Ewing</td>
<td>2004-09-15</td>
<td>15:45</td>
<td>58.5345</td>
<td>141.760</td>
<td>158 M3</td>
<td>Malapaya Shell/1n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8720</td>
<td>OSU</td>
<td>EW509B</td>
<td>79: NC 59W</td>
<td>Maurice Ewing</td>
<td>2004-09-15</td>
<td>15:45</td>
<td>58.5345</td>
<td>141.760</td>
<td>158 M3</td>
<td>Malapaya Shell/1n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8730</td>
<td>OSU</td>
<td>EW509B</td>
<td>79: NC 54W</td>
<td>Maurice Ewing</td>
<td>2004-09-15</td>
<td>15:45</td>
<td>58.5345</td>
<td>141.760</td>
<td>158 M3</td>
<td>Malapaya Shell/1n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8740</td>
<td>OSU</td>
<td>EW509B</td>
<td>79: NC 49W</td>
<td>Maurice Ewing</td>
<td>2004-09-15</td>
<td>15:45</td>
<td>58.5345</td>
<td>141.760</td>
<td>158 M3</td>
<td>Malapaya Shell/1n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8750</td>
<td>OSU</td>
<td>EW509B</td>
<td>79: NC 44W</td>
<td>Maurice Ewing</td>
<td>2004-09-15</td>
<td>15:45</td>
<td>58.5345</td>
<td>141.760</td>
<td>158 M3</td>
<td>Malapaya Shell/1n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8760</td>
<td>OSU</td>
<td>EW509B</td>
<td>79: NC 39W</td>
<td>Maurice Ewing</td>
<td>2004-09-15</td>
<td>15:45</td>
<td>58.5345</td>
<td>141.760</td>
<td>158 M3</td>
<td>Malapaya Shell/1n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Case Study: KM1609
Cruise of the Kilo Moana to Tuvalu, September 2016

- Toughbook
- Zebra Labeling software & printer
- Excel Spreadsheet

- Offline mirror of file structure that can be copied over to our main server on return from sea

- 349 samples recovered
- 439 IGSNs assigned offline at sea
  - Individual rocks (e.g. OSU-KM1609-D4-11A)
  - Personal Subsamples (e.g. OSU-KM1609-D4-11A.AK1)
Website Integration

- Information is geocoded using Google Fusion Tables and visualized on a Google map on our website
- http://www.osu-mgr.org/collections

- Data discoverable through online map interface
- Searchable by core type
- Ability to access:
  - NDGC/IMLGS data
  - Coring data sheets
  - Core descriptions
  - Images
  - MST data
  - Publications
  - XRF data
  - Etc....

- Website provides main external access to the collection
Progress/Summary
~24,000 IGSNs assigned & >31,000 labels printed!

- 10425 short IGSNs assigned
- 52% of entire collection labeled
- Cruises completed back to 1980
- Targeting Fall 2017 for digitization of entire legacy collection

- 13568 long IGSNs assigned
- Long IGSNs (internally) assigned to every sample since 2014
  - Not yet routinely served to PIs
- Aim to routinely provide IGSNs for every sample distributed from the OSU-MGR