# Outreach and Resource Collection SOP for the Open-i project

## Overview

The process of collecting information about biomedical resources at participating universities is broken down into 4 stages:

1. Institutional Inventory (Internal Landscape development)
2. Collection Plan
3. Outreach
4. Data collection and entry

The role of Resource Mining covers the functions of Resource Navigation (four stages above) and Resource Data Curation (use of the eagle-i SWEET to ensure data quality and consistency). This document is designed primarily to address the four stages above.

**Notes about the process**

As will be evident in the following recommended procedures, being highly familiar with your institution’s research environment is a critical predictor of success. Also highly desirable is the knowledge of a broad range of biomedical research resources and the understanding of the value of resource description to different communities. Finally, the understanding and appreciation of data organization is very desirable for a potential resource miner.

Boilerplate Job description for sites hiring resource miners:

***{University}*** *is seeking a highly motivated, enthusiastic, proactive and interactive individual with a background in biomedical research to be part of a nation-wide effort to discover and inventory biomedical research resources and make them visible to researchers around the world. The open-i project aims to utilize the existing eagle-i resource discovery system (www.eagle-i.net) to share resource information at institutions in the CTSA network (*[*www.ctsaweb.org/*](http://www.ctsaweb.org/)*). Harvard Medical School is guiding the central administration and project management. The successful candidate will be based at {university name} and will report to the {university} lead PI ({name of PI}). The candidate will also report to the lead resource navigator and lead curator of the project, both of whom are based at Harvard Medical School. Two training sessions for this position will be held in Boston after initiation of the project.*

*This individual will establish connections with researchers and will collect information about {university}’s biomedical resources such as animal models, unique research equipment, core facilities, educational resources, and research opportunities. Information about these resources may be stored in PI-managed laboratories, institutional core facilities, shared resource centers, or in central offices.*

*The successful candidate will must have superior interpersonal and communication skills, experience with community engagement, and be comfortable with data curation and management. S/he will be the first point of contact for the project team and the public face of the project at {university} responsible for promotion of the benefits of participation. S/he will also be responsible for managing the lifecycle of rich and diverse data sets--from resource identification and entry, through to annotation and publication. To maximize the data’s consistency and usefulness to the community, the data must be managed according to existing curation guidelines.*

*REQUIRED:*

* *MS or PhD in a biomedical science field*
* *Broad knowledge of medical/scientific terminology and technology*
* *Familiarity with the researchers and research landscape of {university}*
* *Technical and organizational capabilities, including excellent MS Office skills*
* *Proficiency synthesizing and presenting data*
* *Strong verbal, written, and interpersonal skills*
* *The ability to work both independently and flexibly as part of a team*

*PREFERRED:*

* *Wet-lab research experience*
* *Background in information science*
* *Experience annotating, managing, or integrating large data sets*

### I) Inventory (Internal landscape)

#### What is it?

An Internal landscape is a document that outlines your specific institutional organization and is populated with information about specific research labs and core facilities at different divisions, departments, centers, and schools. It serves as both a higher-level inventory of resource types at your institution and as a roadmap for resource information collection. Your institution may have a medical school with an affiliated hospital, or a dental school or school of public health. Each of these would make up higher-level tiers of your institution's internal hierarchical landscape (figure 1).

***Figure 1: Sample hierarchical organization of an academic research institution***

**University**

**Medical School**

***Department of Pathology***

Core Facility 1

Core Facility 2

Research Lab 1

Research Lab 2

Research Lab 3

***Department of Immunology***

Core Facility

Research Lab

***Department of Neurology***

Core Facility

Research Lab

***Department of Medicine***

Division of Genetics

Core Facility

Research Lab

Division of Cardiovascular research

Core Facility

Research Lab

**School of Graduate Studies**

***Department of Human Genetics***

Core Facility

Research Lab

***Department of Microbiology***

Research Lab

**School of arts and sciences**

***Department of Biology***

Research lab

***Department of Chemistry***

Research lab

***Department of Cellular and Molecular Biology***

Core facility

Research lab

**Center for Cancer Research**

Core Facility

Core Facility

Research lab

Research lab

Any specific division, department, or center may, of course, have multiple core facilities and numerous different research labs within them. Additionally, a specific facility or research lab may belong to multiple departments, programs, or centers. If so, they should be listed under each organization to which they have an affiliation.

To find specific core facilities, a resource miner can start by scouring local departmental and laboratory websites, and follow-up by communications with core facility directors and institutional managers of core facilities. When continued efforts no longer expose new facilities, we can be confident that we have identified the vast majority, if not all, of the core facilities at our sites. A similar approach can then be taken for mouse models with the exception that after scouring websites, we will focus on local animal facilities and a literature analysis to discover researchers producing or working with murine models of human disease. Information acquired from this phase will be used to produce a table of all facilities, laboratories, and people to contact in our next phase.

For each group of research labs and core facilities (e.g. a department) we create a detailed analysis of the labs and facilities within them. These should include the eagle-i resource types that the lab is likely to offer along with detailed contact information about the facility (table 1).

Table 1: Sample analysis of core facilities at the Border Biomedical Research Facility at the University of Texas El Paso.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Name of facility** | **type** | **resource\_types** | **Contact \_name** | **contact\_title** | **contact\_email** | **contact\_ phone** | **url** |
| ANALYTICAL CYTOLOGY CORE FACILITY | imaging | training, microscope access | Marian Viveros | Core Coordinator | mviveros@utep.edu | (915) 747-5125 | http://science.utep.edu/bbrc/General/analytical-cytology |
| Biomolecule Analysis Core Facility | proteomics | equipment access, training | IGOR C ALMEIDA, PH.D. | Director | icalmeida@utep.edu | (915) 747-6086 | http://science.utep.edu/bbrc/General/biomolecule-characterization |
| Cell Culture and High Throughput Screening (HTS) Core Facility | screening | equipment access, training, screening service | Renato J. Aguilera, Ph.D. | Director | raguilera@utep.edu | (915) 747-6852 | http://science.utep.edu/bbrc/General/cell-hts |
| DNA Analysis core facility | molecular biology | sequencing, microarray, qPCR | Dr. Kyle L. Johnson | Director | kljohnson@utep.edu | (915) 747-6889 | http://science.utep.edu/bbrc/General/dna-analysis |
| Statistical Consulting Laboratory (core facility) | informatics | consultation, training | Dr. Peter Moschopoulos | Director | peter@utep.edu | (915) 747-6764 | http://science.utep.edu/bbrc/General/statistical-consulting-laboratory |
| Bioinformatics Computing Core Facility | informatics | consultation, training | Dr. Ming-Ying Leung | Director | mleung@utep.edu | (915) 747-6836 | http://www.bioinformatics.utep.edu/BCL/ |

Research labs should be detailed similarly to the core facilities (table 2).

Table 2: Sample analysis of research labs at the Division of Genomics at Children's Hospital Boston (Harvard Medical School)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Lab Principal Investigator** | **Area of research** | **resource types** | **email** | **phone** | **url** |
| Beggs, Alan, PhD | Genetics of neuromuscular disease | animal models, antibodies, plasmids | beggs@enders.tch.harvard.edu | 617-919-2169 | http://www.childrenshospital.org/cfapps/research/data\_admin/Site122/mainpageS122P0.html |
| Engle, Elizabeth, MD | Congenital eye movement disorders | animal models, antibodies | elizabeth.engle@childrens.harvard.edu | 617-919-4030 | http://www.childrenshospital.org/cfapps/research/data\_admin/Site163/mainpageS163P0.html |
| Gussoni, Emanuela, PhD | Muscle stem cells | biospecimens, antibodies | gussoni@enders.tch.harvard.edu | 617-919-2152 | http://www.childrenshospital.org/cfapps/research/data\_admin/Site238/mainpageS238P0.html |
| Hirschhorn, Joel, MD, PhD | Obesity genetics, Complex trait genetics | data sets, algorithms | joel.hirschhorn@childrens.harvard.edu | 617-919-2129 | http://www.childrenshospital.org/cfapps/research/data\_admin/Site282/mainpageS282P0.html |
| Holm, Ingrid, MD, MPH | Endocrine genetics | human health studies | ingrid.holm@childrens.harvard.edu | 617-919-2338 | http://www.childrenshospital.org/cfapps/research/data\_admin/Site283/mainpageS283P0.html |
| Kunkel, Louis, PhD | Muscular dystrophy | animal models, antibodies, plasmids | kunkel@enders.tch.harvard.edu | 617-355-6729 | http://www.childrenshospital.org/cfapps/research/data\_admin/Site220/mainpageS220P0.html |

Information describing the larger hierarchical organization can generally be obtained from institutional, school, and departmental websites. Information about the specific resource types that are likely to be used in a specific lab often requires a quick PubMed search and perusal of the lab's recent published works.

Once the internal landscape is complete for your institution, you will have a bird's eye view of the potential resources that can be entered into eagle-i, their locations, and their primary contact information.

### II) Collection Plan

#### What is it?

The collection plan is a prioritized list of labs and core facilities to visit based on your specific institutional needs. For example, in the first two years of the eagle-i project, parts of the ontology were developed before others. Since it was highly desirable to have the latest ontology prior to our data entry, it was important to mirror resource type focus with the development of the ontology for that specific resource type.

1. Core Facilities
   * sequencing
   * genomics
   * imaging
2. Research labs - Antibodies
3. Research labs - Plasmids
4. Research labs - Organisms
5. Core Facilities
   * Biospecimens

Note that the example order of collection reflects the specific needs during the first year of data collection for the project. This prioritization should reflect the specific institutional motivation and focus. For example, a group may prioritize collection based on field of research, disease focus, translational level, or impact level of labs. If there exists a specific prioritization focus that is not reflected in the internal landscape development, the fields collected should be adjusted to ensure that the landscape produced communicates the critical information for triage. For example, if the desire is to prioritize collection based on the level of translation research the labs performs (or the core facility supports) then it would be essential to indicate that type of translational research the lab performs (T1, T2, T3, T4) when developing the institutional landscape.

For open-i, the deliverable schedule for resources will determine the order of approach to researchers. The order of resource delivery for open-i is depicted in table 3.

Table 3: Resource types ordered by delivery date.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Delivery dates** | **7/23/12** | **8/23/12** | **9/23/12** | **10/23/12** | **12/23/12** | **1/23/13** | **2/23/13** | **3/23/13** | **4/23/13** | **7/23/13** |
| Cores | 25% | 50% | 75% | 100% |  |  |  |  |  |  |
| Algorithms |  |  |  | 25% | 50% | 75% |  | 100% |  |  |
| Model Organisms |  |  |  |  | 25% |  | 50% |  | 75% | 100% |

The final part of putting together a collection plan is the timing of lab visitations. By understanding the number of labs that need to be visited and the time frame for getting information into the eagle-i repository, you should be able to develop a realistic plan for communicating with, visiting, and collecting information from your institution’s researchers.

### III) Outreach

#### What is it?

All of the communications that are used to connect with your target labs, explain the value of publishing open linked data in a semantic framework, and obtain consent to collect their resource information for inclusion in the eagle-i repository. These will naturally be adjustable to accommodate proven communication strategies at specific universities, but samples are included below to provide a starting point.

Recommended communication plans:

**Top down**

1. Engage institutional leaders first (School Deans, Center Directors, Department Chairmen, Division Chiefs, etc., as appropriate for your institution), by leveraging the influence of the institution’s open-i principal investigator
2. Meet with leaders to discuss the advantages of having their institutional resources listed as part of the open-i project.
3. Provide them with boilerplate language to disseminate to the researchers and core facilities in their area.
4. Follow up with an introductory communication referencing the leader’s communication. Schedule a meeting with individual labs.

**Bottom up**

1. Convene meetings of known proactive early adopters and information sharers.
2. Facilitate discussions of information sharing for the purpose of reducing redundancy in resource development and testing.
3. Leverage working group to reach out to the algorithm development community.
4. Follow up with an introductory communication referencing the leader’s communication. Schedule meetings with individual labs.

Initial broad outreach to each organization should be coordinated with press releases to local research publications, flier distribution, and announcements at symposia and local events as appropriate for each institution.

### IV) Data collection and entry

The Resource Miners at each university will collect all data for the open-i project. After establishing consent from researchers, the open-i team will perform preliminary entry of data from all target labs. Data entry training documents will be used to familiarize resource miners with the existing eagle-i search and data entry tools. The Resource Miners will communicate with researchers when needed to ensure accuracy and thoroughness of information. The final step is the curation of data according to the eagle-i repository data curation guidelines. All data will be published as open linked data.