

RESEARCH RESOURCE CENTER

The Research Resource Center (RRC) is envisioned to be a core component of RERF's infrastructure. The successful implementation and execution of the RRC is a necessary component to advance RERF's strategic plans.

The RRC's mission is 3-fold. The mission includes:

1. Protect, index, and integrate RERF's research assets. These include data, bio-sample inventories, paper records, artifacts, manuscripts, datasets and programming scripts, as well as other historically important articles. Access to research data will be made through various kinds of cloud environments with clear accessibility rules that protect the privacy of our subjects.
2. Enhance RERF's ability to perform research by integrating all data and bio-sample inventories. Tools for data visualization, data assembly, and analysis will simplify and standardize access, and facilitate research.
3. Provide an administrative framework to facilitate internal research as well as facilitate collaborative research projects, including contracts and external grants.

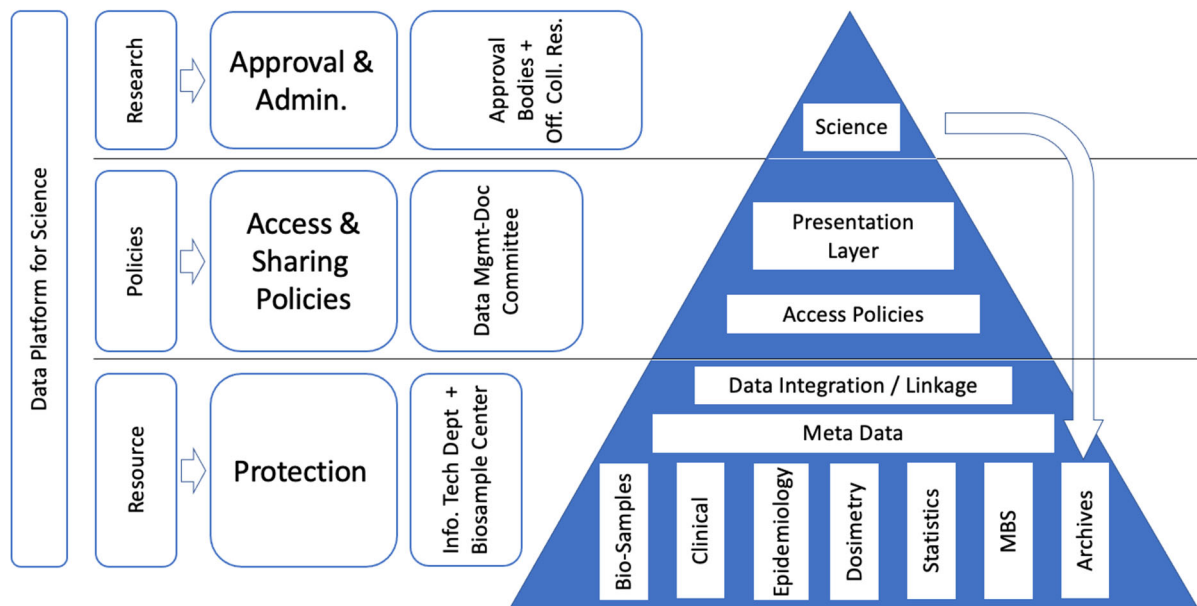


Figure 1. Functional structure of RERF's Research Resource Center

Referring to Figure 1, the bottom layer of the pyramid is for protection. This will include a framework for digital data storage, integration, and visualization. Beyond digital data, it will also include a Content Management System (CMS) to catalog and index all digital scans of RERF materials and inventories of historical artifacts. Copies of data and scans will be housed centrally (Mission #1). All materials will have information describing each entry (“metadata”) so that the entries can be located, be appropriately associated with related materials, and be searchable. The second layer is the “Policy and Access” layer. Using institutional rules and modern authentication methods, only those materials to which a person is authorized will be viewable. This layer will include the Presentation tools so that researchers (both internal and external) can peruse RERF data catalogs. Authorized users can also see research proposals and

manuscripts based on those data and visualize/query the data to design new research (Mission #2). The top layer is the Science Layer. Here, authorized research can be performed. If the research is collaborative, the “Office of Research Support” (ORS) can facilitate the research by executing contracts and assisting with procedures to distribute data and process bio-sample requests (Mission #3). At the conclusion of a study, all the materials associated with the study (data, assay results, analysis scripts, manuscripts, etc.) will be placed into the Archives with appropriate metadata and linkages so that those products can be reused by future researchers.

Personnel and functional units to develop/manage the RRC have been tentatively set by the Research Resource Center Operating Committee. This ad hoc committee was initiated in November 2021 and had its first meeting in December 2021, and the second meeting was held in May 2022. It is led by Dr. Ono as chair and Dr. Tanabe as co-chair and includes all department chiefs. The previous chair Dr. Grant retired in December 2022. This committee supersedes a previous ad hoc committee entitled the “Preparatory Committee for the Establishment of the Research Resource Center”, which was established in December 2019. See Figure 2 illustrating the functional units and reporting lines of the RRC.

Three primary development needs

To realize the RRC with the functionality outlined in Figure 1, there are three broad development areas that need to be completed. These three areas are a “data framework”, a “content management system”, and the Office of Research Support (ORS). The functionality and the development areas have been carefully thought through by RERF “Technical Team” that started biweekly meetings in 2017. The team had included senior members of the ITD and Dr. Grant. Currently, this function has given to Research Resource Section.

Data Management Framework. The “data management framework” will be the underlying computer environment for RERF data. It must be able to store, annotate, and integrate data. Ideally, it would be flexible so that new data can be easily added to the system using FAIR principles. Ideally, annotation would be possible by the researcher who created the data (rather than forcing computer-savvy administrators to perform this work). The system would support both English and Japanese, could be installed on local servers or in a cloud environment, and would allow remote researchers to run analyses in situ so that the data don’t “leave” the institute. There would need to be a hierarchical and highly secure authentication system so only authorized personnel could access the data. It would also allow “data visualization applications” to interact with the underlying data so that it could be visualized and queried for data exploration.

A pilot for such a system (The Gen3 Data Commons, developed by the University of Chicago) was run for over a year in 2020-2021. Unfortunately, the pilot was not successful for a variety of reasons discussed later in this chapter. RERF is now searching for a suitable data management framework. Currently, we are planning the evaluation of the function and capabilities by “Proof of Concept” of Seven Bridges as a candidate that is a group company

headquartered in the USA.

A pilot project with data visualization software was underway in last fiscal year, as a result, in the Epidemiology department, Dr. Sugiyama has provided extremely practical use-case for the visualization using actual datasets in the institute. Therefore, in this fiscal year, we have added 4 licenses of Tableau to the Epidemiology department. More time and budget is needed to make use of it in the entire institute.

Content Management System. The Content Management System (CMS) is a system for storing all paper/text-based assets within RERF. A flexible and strong security system that allows access to RERF materials while strictly adhering to RERF's authorized access policies is required.

For more general materials (i.e. lists of publications, etc.), RERF currently has a number of databases for viewing various lists (Research Protocols, Technical Reports, etc.) but they are often separate databases (i.e. multiple sites need to be interrogated), and there is no full-text search ability anywhere within RERF. These issues (no centralized access, no full-text search) are major shortcomings in RERF's current computer environment. RERF needs a CMS that can house scans of paper records within RERF, all RERF pdf files, lab notebooks, questionnaires, etc. The system needs to be bilingual, intuitive, and allow individuals to upload locally held assets for preservation and re-use.

Office of Research Support. Performing research at RERF is unnecessarily slowed by administrative bottlenecks. All work is performed on paper and many approval forms overlap or conflict with other approval forms. Further, there is no "presence" on RERF's external home page for outside researchers to peruse RERF holdings or initiate conversations to explore collaborative research projects. The Office of Research Support (ORS) would try to solve each of these problems. We envision a thorough review of RERF's application forms and redesigning them to avoid overlap. Further, the entire process should be moved to an online system. We are planning the development and introduction of an electronic workflow system that covered whole decision-making in the institute, and all procedure in the institute will be digitized and visualized because of the introduction.

RESEARCH RESOURCE CENTER

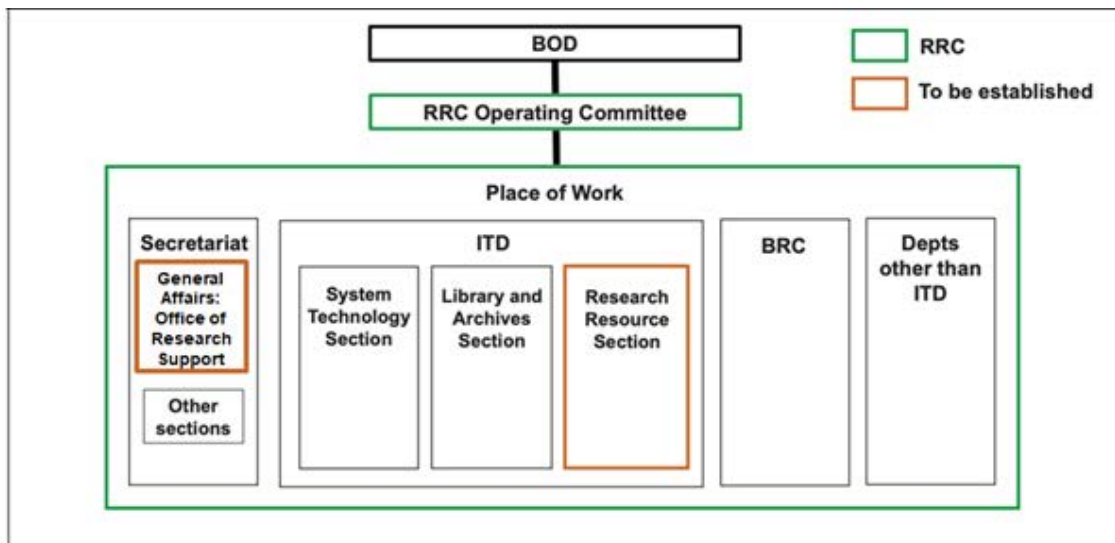


Figure 2. Structure of the RRC. This is the first of a two-step process where sections are created within existing units at RERF. The second step will be to establish the RRC as a “stand alone” center with its own personnel and physical location.

FY2022 Departmental Achievements

The primary accomplishment of the Research Resource Center in 2022 was the re-design of its administrative structure and digitizing of microfilm and the planning digitization of negative film of chromosome.

Initially, plans for the RRC were led by the “Preparatory Committee for the Establishment of the Research Resource Center”, established in December 2019, with Drs. Niwa and Ullrich as co-chairs. This ad hoc committee was established to inform and garner the support of all departments. Within the ad hoc committee, a “Drafting Guidelines” subcommittee was responsible for determining the position of the RRC within the institution, including reporting lines, leadership, budget, personnel, and internal structure of the RRC. Its 5th meeting was held in 2021, by which time several projects involving inventory and scanning preparations had been completed. The ad hoc committee was disbanded in November 2021 with the establishment of a new ad hoc committee “Research Resource Center Operating Committee” (RRCOC), chaired by Dr. Eric Grant with Dr. Tanabe as the vice-chair. The first of the RRCOC was held in December 2021; Since then, two meetings have been held. At the end of this year, Dr. Grant has retired, and Dr. Ono assumed the chairmanship of the committee. With Dr. Grant retirement, some of last year’s plans have been changed.

The RRC was initially envisioned as a stand-alone “Center” within RERF with its own physical presence and personnel. This is still the long-run plan. However, to speed development, a “two-step” process has been embraced to initially establish several new sections within existing units at RERF and then later bring the groups together as the RRC. As described in the “Overview” section, there are three primary development areas needed to make the RRC functional. They include 1) The installation and operation of a data framework for data integration, data tagging, data visualization, and dataset assembly. 2) A Content Management System that will centralize all written materials within RERF and provide full-text search (bi-lingual). 3) An Office of Research Support that will be tasked with streamlining all application/approval procedures and moving these procedures to an online system, as well as establishing an online presence for the RRC. To accomplish these tasks, the RRCOC has approved the creation (or repurposing) of three sections. They are:

- **Research Resource Section.** This new section (the “RRS”) will be a technology-based section located within the department of Information Technology and will be led by the Department Chief, Dr. Ono. The Section will be tasked overseeing the installation of a “data management framework” that will provide a flexible system for centralizing RERF data, including tagging the data, visualizing the data, and assembling new datasets for research. The section will start by hiring a senior contract programmer to lead the RRC programming efforts. In FY2023, a new programmer will be hired into the Section. The Section will further be bolstered with two (existing) programmers that will be concurrently assigned from ITD’s Systems Technology Section to assist with data integration and development tasks.
- **Library and Archives Section.** This section already exists within RERF but will be given the additional assignment of identifying, uploading, and cataloging manuscripts, research protocols, technical reports, etc. into the new Content Management System (CMS). They will also work with outside vendors to customize the CMS for RERF’s needs. The Section reports directly to Dr. Ono.

- **Office of Research Support.** Previously, this section had established within RERF's Secretariat. However, for the reason described below, this unit will be incorporated into the Research Resource Section. The first reason is to unify the chain of command. The second reason is that the role of this unit has been reconsidered in terms of the institute's strategic plan. The role of the unit is not only assistance or representation in various administrative procedure but also facilitate research activity like an URA (University Research Administrator) that is a human resource to strengthen research capabilities in the institute. For example, they should have management and utilization of intellectual property or responding to diversification of various external research funding sources. In addition, we are planning to digitize decision-making processes; these processes include not only administrative procedure (such as submission of various application) but also whole decision-making flow regarding research in the institute. The execution of this plan and the rolls requires extremely close collaboration with the technical unit and must operate under a unified chain of command.

In addition to the RRCOC setting up the future structure and personnel requirements of the RRC, several projects have been underway.

- **Five-year timeline for establishing the RRC.** We have worked to create a timeline for technical developments within the RRC. See Figure 1. Highlights of the timeline include several infrastructure improvements at RERF required for technical advancement (e.g. strengthen external communication line), and the introduction of an electronic workflow system to allow the automation/streamlining of a number of processes that are currently handled in a manual, paper-based manner within RERF. Other highlights include the timelines for establishing new sections to handle the tasks discussed above. The RRC should be in full operation by the end of FY2025 with an external on-line presence established by the end of FY2024.
- **Scanning Center Practical Project Initiated.** The Scanning Center constructed last fiscal year is now in the practical application phase. After logging in to the scanning machine, all scans are sent through an Optical Character Recognition feature and the newly created pdf file is centrally located under the creator's name. In this fiscal year, an electronic flow was established to allow paper-based forms submitted to ITD to be sent electronically using the Scanning Center. This streamlines submissions, especially those from distant Nagasaki institute. Also, this construction is a prototype for the electronic workflow system that will be introduced in the near future.
- **Content Management System Pilot.** RERF identified Open Source software for content management and then located the top programming firm in Japan specializing in the software. We then contracted that firm to develop a pilot system for RERF's content management. This is progressing but we have learned that RERF's design specifications are quite challenging. They include: A hierarchical access system that is compatible with RERF's single sign-on system, fully multi-lingual, complex relational links between various materials, flexible, and customizable. Currently, the requirements definition of the system was largely completed, and we are creating the specification document for the procurement of the system.

- **Data Visualization Pilot.** RERF has purchased 3 license of “Tableau” data visualization software in last fiscal year. This software is a leader in the field of data visualization and can be used as a ‘front end’ to explore RERF data. As with any complex software, there is a significant learning curve. The software looks promising but success will require personnel be dedicated to learning and developing the package for use at RERF. The purchased licenses were made available to the Epidemiology Department on a trial basis. As a result, in the Epidemiology department, Dr. Sugiyama has provided extremely practical use-case for the visualization using actual data-sets in the institute. Therefore, we have added 4 licenses of Tableau to the Epidemiology department. The additional licenses were distributed to the Tumor & Tissue registry Office and Analysis Office in the Epidemiology Department.
- **Digitization of deteriorating microfilms.** Several hundred microfilms located in the Master File Section and the Clinical Studies Section were deteriorating due to poor storage conditions over many decades. These films were identified and all have been digitized by an external vendor. Recovery was generally excellent, with only 4-5 reels being deteriorated to the point that they could not be digitized using automated machinery. From FY2020 to FY2021, about one million digitized files were created and were stored in the independent electronic storage.
- **“Seven Bridges” Data Management Framework evaluating is ongoing.** To identify a suitable framework for data storage and management tool, Dr. Grant and I visited IT vendors in the U.S. last June. Currently, we are planning a Proof of Concept to pilot one of a candidate system using real data in the institute. We will use single cell RNA sequence data from mice in the institute as the primary data for the PoC.
- **The retirement of Dr. Grant.** It is very unfortunate, in last December, Dr. Grant has retired. Dr. Grant, ITD leadership, and members from Systems section and Library & Archives section had initiated bi-weekly meetings in 2019 to scope the RRC into different tasks and begin planning for implementation. These discussions range from high-level conceptual discussions to specific features. The impact of losing a key person in the execution of these plan is extremely significant for us. However, Dr. Grant will contribute to the institute as a part-time researcher and will continue to work under his supervision to carry out the mission of the RRC.