

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, biosample inventories, paper records, artifacts, manuscripts, datasets and programming scripts, as well as other historically important articles. Access to research data will be made through a web portal with clear accessibility rules that protect the privacy of our subjects.
2. Enhance RERF's ability to perform research by integrating all data and biosample inventories. Tools for data visualization, data assembly, and analysis will simplify and standardize access, and facilitate research.
3. Provide an administrative framework to use and share RERF resources and facilitate contracts and grants.

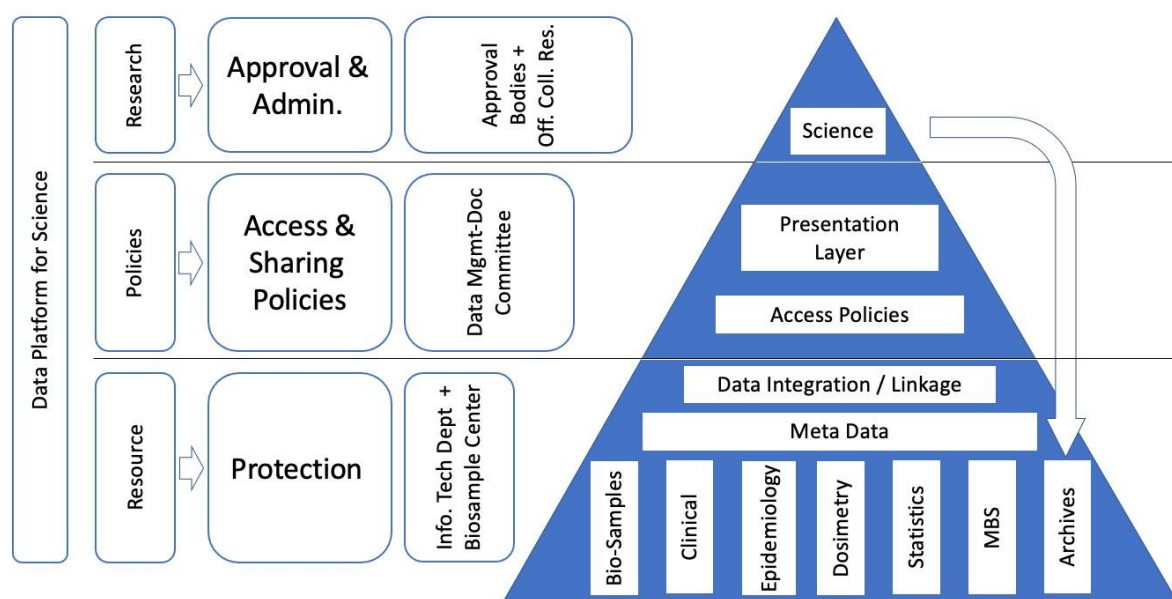


Figure 1. Functional structure of the Research Resource Center

Referring to Figure 1, the bottom layer of the pyramid is for protection. This will include a data management system for digital data as well as a Content Management System (CMS) to catalog and index all of the digital scans of RERF materials. 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 products (research proposals and manuscripts) based on those data, and

visualize/query the data in order 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 Collaborative” can facilitate the research by executing contracts and assisting with procedures to distribute data and biosamples (Mission #3). At the conclusion of a study, all the materials associated with the study (data, 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.

The RRC’s structure, including staffing, reporting lines, relationship to other departments, budget, etc. are still in the discussion phase. Overall leadership for this planning is under the authority of the “Preparatory Committee for the Establishment of the Research Resource Center”, which was established in December 2019. That steering committee currently includes five subcommittees, one of which is the “Subcommittee for drafting guidelines for establishment of the RRC”. It is this “Drafting” subcommittee that must make recommendations regarding the structure of the RRC.

While no decisions regarding the structure of the RRC have been made to date, it is envisioned that the RRC will have a number of permanent staff with sections to support IT development, archiving, and the Office of Collaborative research, which will be the initial contact point for external researchers wishing to access RERF data and biosamples. The RRC will be required to have collaborative relationships with all entities within RERF for access to inventories and to help in preserving/cataloging materials.

Despite the lack of formal structure, the ITD and the Associate Chief of Research have held bi-weekly meetings for the past two years to discuss technical issues, perform pilot projects for needed computer infrastructure, perform pilot projects for document scanning, have meetings with outside vendors regarding software infrastructure, etc. Please see the next sections for specific work completed and future work planned.

FY2020 Departmental Achievements

Despite the lack of formal structure or leadership, much has been accomplished in 2020, including a number of technical achievements, administrative achievements and pilot projects.

- **Chief of ITD.** Dr. Ono was hired as the Chief of the Department of Information Technology. This position had been open since 2017. The new chief can work with the RRC to help in the initial design decisions.
- **New programmers.** Two new programmers hired at RERF. Two recent computer scientist graduates joined RERF in April. These are the first new hires in this department in more than 10 years, filling a needed gap in both knowledge and age distribution. They are expected to be contributors to the RRC.
- **MEXT Grant submitted.** A MEXT grant in collaboration with Hiroshima University’s

Research Institute for Radiation Biology and Medicine (RIRBM) was submitted. This collaborative application between RERF and RIRBM (possibly the first of its kind) is to establish a public-facing web-based archive system to give the public access to jointly held historical materials of interest to the public and social scientists. The design is to have federated databases (each institute can independently control which materials are provided) with a single interface for search and presentation. Knowledge gained from the design of this system can be leveraged for inclusion in the creation/design of the RRC. The opening date for the (proposed—award announced in April) web site is 2025, the 80th anniversary of the atomic bombings. PIs from RERF are Drs. Grant and Ono.

- **Initiation of the "Preparatory Committee for Establishment of the Research Resource Center" and the "Subcommittee for Drafting Guidelines for the Establishment of the RRC."**

RRC development efforts prior to the initiation of this committee were undertaken solely by members of the ITD under the direction of the Associate Chief of Research. These were preliminary efforts for a project that will require the support of RERF directors and all departments within the Institute. In order to make RRC development efforts official and garner the support of all departments, an ad hoc committee was established in December 2019 with Drs. Niwa and Ullrich as co-chairs, Drs. Grant and Kodama as co-vice-chairs, all department chiefs, and the chief of the Secretariat, plus two executive secretaries. This formalized committee should provide the RRC with sufficient authority to begin the process of implementing the RRC with support from departments, the Secretariat, and other administrative support. The "Drafting" subcommittee is responsible for determining the position of the RRC within the institution, including reporting lines, leadership, budget, personnel, and internal structure of the RRC. These decisions are essential to move the establishment of the RRC forward and begin operations.

- **Initiation of the "Digitization Subcommittee" (10 persons under Dr. Ono and Ms. Marumo/ITD) and the "Data Inventory Subcommittee" (13 persons under the direction of Dr. Sakata/Epidemiology).** These two committees have completed the first phase of inventorying all materials within RERF and identifying those that will need digitization. After numerous meetings and interviews with all departments, the Inventory Subcommittee identified and organized lists of 20 different types of materials (digital files, notebooks, maps, coding manuals, paper data cards, lab photos, etc.). This represents millions of paper sheets, digital records, filing cabinets full of folders, boxes, etc. These lists are currently being moved to a REDCap database (online database) to further categorize and assign priorities for preservation. The Digitization Subcommittee worked together with the Inventory Subcommittee to identify materials that require scanning.

Medical chart scanning pilot project completed. Via a contract with an external vendor, a pilot project was performed to scope the effort required to scan the paper medical charts stored in the Hiroshima Department of Clinical Studies. After a three-week trial period testing various work-flow methods, overhead vs. flatbed scanners, methods of dealing with various fasteners in varying states of decay, and handling odd-sized paper, a report was

written and delivered to RERF. The pilot project estimated that there were 10,000,000 sheets of paper contained in the medical charts. The vendors concluded that digitizing all of the materials in the medical charts would cost roughly \$5M dollars. A 25-member team could complete the job in 3 years. A smaller team could do it for the same cost, but it would take a longer period of time. It is estimated that Nagasaki materials would add roughly 40% to the total costs.

- **“Gen3” Data Commons Pilot Project ongoing.** Following a site visit from Dr. Robert Grossman from the University of Chicago, RERF decided to attempt a pilot project using the Gen3 system. Gen3 is an open-source software platform for large scale data storage and analysis. RERF had no expertise or available personnel when this pilot was discussed. Through contacts, Dr. Grant recruited a computer scientist working on his PhD in Epidemiology from the University of Texas Medical Center to perform an internship with RERF. The initial proposal was for the programmer to visit the University of Chicago for several weeks of training in the spring and then visit RERF through the summer to work full time on the Gen3 installation in Hiroshima. However, with the COVID pandemic, all site visits had to be canceled. To his credit, the computer scientist met with Dr. Grant for weekly teleconferences throughout the summer and into December 2020. Unfortunately, the pilot project has not been fully installed. This is mostly a confluence of bad circumstances rather than a fault with any individual. Compounding factors included the nascent state of the Gen3 system with incomplete documentation, the Gen3 engineering staff being overwhelmed by their day-to-day activities and unable to give sufficient time to the RERF installation, and a graduate student trying to perform this project remotely.
- **Bi-weekly Technical Team meetings.** Dr. Grant, ITD leadership, and members from Systems section and Library & Archives section 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. Significant time and research were devoted to technical needs and identifying groups/software platforms that are currently implementing our desired features. This group has focused on technical issues. The larger "Preparatory Committee for Establishment of Research Resource Center" will identify and execute many of the tasks that need performed but the ITD team will likely be tasked with the actual technical implementation work (with the help of external experts). These meetings continue outside the scope of the formal committees.

While not occurring in 2020, development efforts for the RRC started in 2017. Here is a brief list of other activities/achievements that occurred prior to 2020.

1. Visited data and archiving specialists from around the world. As the ITD staff do not have existing expertise for data integration or archiving, it is necessary to contact experts from various institutes involved in large-scale projects to understand options and development strategies. Face-to-face meetings were held with the following experts/institutes:
 - Dave Thompson/Digital Curator, Wellcome Library

-
- Stephen Chanock/ US NCI, Director of Cancer Epidemiology and Genetics
 - Anthony Philippakis/Broad Institute, Chief Data Officer
 - James Cuff/Harvard University, Assistant Dean for Research Computing
 - Brian White/Harvard Univ., Manager of Research Computing IT Operations
 - John Quackenbush/Dana-Farber, Biostatistics and Computational Biology
 - Meir Stampfer/Harvard SPH, PI Nurses Health Study
 - James Lacey/California Teachers Study, Director Health Analytics
 - Tom Murphy, UMich. ICPSR, Director of Computer and Network Svcs.
 - Robert Grossman/UChicago, Chief Research Informatics Officer, Division of Biological Sciences
 - Nobumichi Ariga/Japan National Museum of Nature and Science, Department of Science and Engineering
 - Shigeo Sugimoto/University of Tsukuba, Library, Information and Media Science

2. Site visit by Dr. James Cuff (Harvard), May 2017

Dr. James Cuff, Assistant Dean for Research Computing, Harvard University came to RERF for a 3-day visit to evaluate RERF's research needs and current infrastructure. Cuff worked with ITD and RERF leadership to outline a "White Paper" in an effort to scope the problem and break it down into logical steps.

3. Site visit by Dr. Robert Grossman (University of Chicago), January 2019

As a follow-up to a visit to the University of Chicago where discussions of RERF's needs and whether the Gen3 system may be a good solution for RERF's data integration needs, Dr. Robert Grossman, Chief Research Informatics Officer, Division of Biological Sciences visited RERF for two days. He gave a talk ("Building a Data Center with Gen3") and met with visiting archivists, our ITD staff, and RERF senior management. At the end of the meeting, it was decided that a pilot program using the Gen3 system should be implemented at RERF.

4. Site visit by Dr. Christy Henshaw (Wellcome Trust), March 2019

Dr. Henshaw, Digital Production Manager, delivered a talk at RERF entitled "Digital Engagement at Wellcome Collection" in which she discussed the strategy of the Wellcome's online presence as well as many of the technical details in their operations. Dr. Henshaw also met with our archivists and other local archivists invited to attend her lecture. Many of the tools that Wellcome uses are open source and we discussed possible collaborative projects that may assist us in our RRC development.

5. Purchased servers to support the RRC

After the pilot project was agreed upon in the January meeting, discussions of technical requirements began. The Gen3 team recommended a minimum of two

high-spec servers. RERF decided to acquire 3 servers (two for Gen3 and one to support the web server with search and archives). In summer, the Supply Section commenced an open bidding process. The computers were purchased and delivered in September 2019.

6. Site Visit by Robert Grossman and two engineers

In November 2019, Dr. Grossman and two engineers from the University of Chicago came to install the pilot version of the Gen3 Data Commons on the newly acquired servers. Dr. Grossman stayed two days and the engineers stayed for the full week. Our top systems engineer worked together with the engineers for the entire week to gain an understanding of the installation tools used and the network setup required. At the end of the week, an initial installation had been completed. This system also required RERF to adopt a modern authentication system using a 3rd-party authentication system with certificates (Microsoft Azure infrastructure).

7. Created two secure computing rooms with video surveillance and biometric entry for collaborative human genome research

RERF is introducing human genomic data for which we do not have sufficient in-house computing power. To perform this type of research, access to supercomputers is needed (our initial target is the Tohoku Medical Megabank supercomputer). To use this computer, an approved physical space and computer infrastructure was required. Using construction vendors and specifications, two rooms have now been built at RERF for this specialized function.

8. Staff members attended conferences or lectures on Open Library Services, Open Data Initiatives within Japan, Conference on Open Repositories, lectures on Professional development

9. Completion of White Paper

