[Host, Dr. Maryellen Giger] Welcome, everyone, to the MIDRC Town Hall. This is the first Town Hall that we’re hosting. I am Maryellen Giger, from the University of Chicago and AAPM, and I will give an intro.

Here’s our agenda for today. We’re going through a short overview. Then we will look at why one should contribute data to MIDRC, then we’ll go through the MIDRC data pathways (the how-to), and then we’ll have a question & answer period.

The COVID-19 pandemic presents an urgent and critical public health crisis
Essential biomedical research and development is needed to urgently address:
(i) surveillance and early detection of COVID-19 resurgence via monitoring of imaging and other clinical data
(ii) detection, triaging, and differential diagnosis of COVID-19 patients
(iii) prognosis, including prediction and monitoring of response, for use in patient management.
In response to this need, representatives of the RSNA, ACR, and AAPM with NIBIB have jointly developed the Medical Imaging and Data Resource Center (MIDRC) for rapid and flexible collection, AI research, and dissemination of imaging and associated data, to be administered and hosted through the University of Chicago.
So COVID-19 pandemic presents and urgent and clinical public health crisis. And this has given us a need to urgently address surveillance, early detection, triaging and differential diagnosis of COVID-19 patients, as well as prognosis, including prediction and monitoring of response, for use in patient management. Thus, in response to this need, representative of the RSNA, ACR and AAPM with NIBIB have jointly developed the Medical Imaging and Data Resource Center, or MIDRC (pronounced - mid-rik). And this is for rapid and flexible collection, AI research, and dissemination of imaging and associated data, which is then hosted through the University of Chicago.

**Thoracic imaging in the COVID-19 Pandemic**

While thoracic imaging, including chest radiography (CXR) and computed tomography (CT), are being re-examined for their role in patient management, the limitations for improved interpretation are partially due to the qualitative interpretation of the images, and thus we aim to develop artificial intelligence (AI) methods to aid in the interrogation of medical images from COVID-19 patients, *eventually including cardiac, brain, and other images.*

While thoracic imaging, including chest radiography and computed tomography, have been re-examined for their role in patient management of the COVID patient, we also will be aiming to look at areas beyond that. Currently our datasets include chest x-rays and CT, but in the future we aim to develop AI to aid in the interrogation of medical images from COVID-19 patients, including cardiac, brain and other images.

**Rapid Response to COVID-19 Pandemic**

*University of Chicago NIBIB Contract PI: Maryellen Giger*

- **American Association of Physicists in Medicine** (AAPM) PIs:
  - Maryellen Giger (University of Chicago & AAPM Data Science Committee Chair)
  - Paul Kinahan (University of Washington & AAPM Research Committee Chair)
- **Radiological Society of North America** (RSNA) PIs:
  - Curtis Langlotz (Stanford University & RSNA Board Liaison for IT & Annual Meeting)
  - Adam Flanders (Thomas Jefferson University & Member RSNA CDE Committee)
- **American College of Radiology** (ACR) PIs:
  - Etta Pisano (ACR Chief Research Officer & Harvard University)
  - Michael Tilkin (ACR Chief Information Officer)
- Gen3 PI: Robert Grossman
So what is MIDRC? In our rapid response to the COVID-19 pandemic, it was established at the end of August in 2020. It is co-led by the three associations, AAPM, RSNA and ACR, with representative and leaders from each of them, shown here, and you will be hearing from each of these today.

ACR, RSNA and AAPM have successfully worked together in the past. So it was logical for all of them to come together to help create MIDRC. And also we have to give kudos to NIBIB, for being creative and inspiring these three organizations to lead this effort. We acknowledge our advisors, many of whom you will recognize, Drs. Sullivan and Schnall. And also, when you’re communicating with MIDRC, you most likely will be talking with Katie Pizer, our lead Project Manager.

So MIDRC goals is to progress from data to deployment, hypothesis and discovery, and accelerate the creation and transfer of knowledge for clinical management of COVID-19. A major purpose of MIDRC is to create an open discovery data repository common, of high quality, diverse data and connectivity to clinical data. We also are funded to look at advances and development in machine intelligence.
This is the rough structure of MIDRC. Funding comes from NIBIB, it goes through the University of Chicago to fund Gen3, which you’ll hear about later, as well as the three organizations and various other institutions working with us. There’s two major scientific components. One is creation of the infrastructure, the data repository. And this is done through 5 technology development projects, which are advised through 3 data science subcommittees and advisory committees, as well as 12 collaborative research projects.

MIDRC spans the medical imaging community, with the 3 associations. If someone is in medical imaging, chances are they belong to at least one of the three, ACR, RSNA or AAPM. It also includes 23 institutions from academia, community practices and FDA. And so we span the medical imaging community, as well as the nation. And I invite you to go to our website.
In this multi-group, two-year, NIBIB-funded project, we have two intake portals, and you’ll hear more about these today.

One is the RSNA RICORD system, and the other one is the ACR CIRR system.

On our website, you check back and check where we are in developing this commons, looking at the total ingested to MIDRC, those undergoing quality and harmonization, and those that are released.
We also will have a public access portal, and this is developed by Gen3, at the University of Chicago.

Gen3 is a data ecosystem, including various data commons you may heard of; the BioData Catalyst and the N3C, NCI Cancer Research Data Commons. And MIDRC is now hosted by Gen3, you can find it up there.

Visit our website, www.midrc.org. This is open for researchers, data contributors and the public, as well as our data commons, which is www.data.midrc.org, where you can go to browse and search the data, as well as download the data.
We’re working with various organizations, and as they come aboard we’re including them on our website. Here’s our current MIDRC partners and we’re constantly increasing these.

For more updates on MIDRC please subscribe to our newsletter, as shown by the email there. Or, of course, you can get to it by our www.midrc.org site.

So I’m going to stop here and get to the point of this Town Hall, which is contributing data and how to do that. But, remember, for a given image a patient has already benefitted through medical care. The hospital and medical system has already benefitted through reimbursement. So now let the public benefit from MIDRC’s second use of the images. You can help change the culture of medical imaging. Thank you and let’s work together............I will stop there and we’ll now send this over to ‘Why Contribute Data to MIDRC’ to Drs. Langlotz and Pisano.
Thank you, Maryellen. It’s a pleasure to be here with all of you today. I wanted to spend some time consider why an organization would want to consider contributing data to MIDRC.

And I want to start with a consensus conference that was held in 2018 to look at a roadmap for both foundational and translational research in artificial intelligence. And the top issue that came up in both of these areas, both foundational and translational, is the need for good data to train these artificial intelligence models. So that’s one of the prime motivations for the MIDRC project.
And just to reinforce what Maryellen just described, David Larson’s work with a group of us, including a medical ethicist, examined the status of clinical data and the ethical underpinnings of sharing clinical imaging data. And I think this quote is important to consider; “After clinical data are used to provide care” — and this really reinforces what Maryellen just described — “the primary purpose for acquiring the data is fulfilled. At that point, the data should be treated as a form of public good, to be used for the benefit of future patients. The data are not ‘owned’ in the traditional sense, but rather all who interact with or control the data have an obligation to ensure that it is used for the benefit of future patients and of society.” So I think that’s an important principle for us to follow.

Currently there are some datasets available for machine learning, none of these are for COVID. But there was Ron Summer’s lab at NIH released 30,000 chest X-rays and 32,000 CT images. Our lab at Stanford released, with the MIMIC group over at MIT, over 600,000 labeled chest X-rays. The NYU group has released knee MRI reconstruction data. There’s a Dartmouth histology dataset that’s available.
And then our center at Stanford has now released over 10 AI-ready datasets. But it’s really not enough, if you think about it. These are coming from specific institutions and I’ll have more information on that in just a moment.

When data are released, they can be very powerful in driving change and in the development of very capable AI algorithms. For example, these are some of the data science Challenges that have been held by the RSNA over the past several years; a Pediatric Bone Age Challenge, a Pneumonia Detection Challenge, and Intracranial Hemorrhage Detection Challenge and a Pulmonary Embolism Challenge just this past year. And RSNA’s planning two additional Challenges for the coming year. To give you a sense of the degree of engagement – on the Hemorrhage Detection Challenge we collected over 27,000 head CTs, which constitutes over a million images, from four sites across three continents; sixty volunteer neuroradiologists labeled those images with one of 6 labels; over 1,300 data science teams submitted AI algorithms to solve that problem; over 22,000 entries were made.

So this can be a real catalyst to the development of outstanding AI tools, the availability of data.
But, as I said, the data that’s being used today tends to come from very narrow sources. So this is a study we published in *JAMA* (the *Journal of the American Medical Association*) earlier this year, where we looked at all of the published work on AI and medical imaging and where the datasets arose from, and found that the majority of datasets come from just three states, California, New York and Massachusetts.

And, of course, if you think about it, there’s a wide diversity of both health conditions and demographic conditions across our states. This is just two examples, age and household income, which correlate to health needs. So we really need a much greater diversity of data, and that’s one of the prime motivations for a project like MIDRC, which can draw data from across the country, can develop diverse datasets to train these algorithms, that then can answer these critical COVID questions for any patient in any state, or across the world.

**COVID Use Case Examples**

- Identify infection
- Diagnose disease
- Assess extent
- Monitor therapy
- Detect complications
- Predict outcome
What are some of the use cases that can be addressed if you contribute data to COVID? Really identifying the infection, diagnosing or distinguishing it from other causes of pneumonia, assessing the extent of the disease, monitoring various therapies, detecting complications and predicting outcomes, like ‘will this patient be admitted to the hospital?’, ‘will the patient need ICU stay?’, mortality rates. Those can be very important things to predict.

So why contribute data to MIDRC? You can contribute your data, you can contribute your expertise in labeling data or other, contributing to AI development specifically. You can participate in the overall resource center that we’ve developed, you can learn something about AI development. You can certainly piggyback on our efforts, we’re in discussions with other large repositories of COVID data, such as N3C which is collecting from now over 100 different healthcare organizations across the country. And something that may be of interest to the folks on this call is that there will be a small number of data contribution grants and contracts that RSNA plans to make. I wouldn’t want this to cause you to hesitate, they will be available later and if you begin the process of contributing and preparing to contribute data now, that’s going to give you an advantage when those grants become available. They’re small grants, but they can help support you as you contribute data to MIDRC.

So I hope you’ll consider contributing data. We think it’s a project that will allow us to answer questions that would not otherwise be answerable, aggregating data from a broad variety of sources and pulling together the scientists needed to develop these advanced algorithms. Thank you very much, at this point I will turn it over to Dr. Etta Pisano.

[Presenter, Dr. Etta Pisano] 13:29 Thanks, Curt. I’m gonna share my screen here, it’ll take me a second to get it up. So I can’t tell if you can see it but I’m assuming you can. So I’m going to expand a little bit on the points that Curt just made. And talk specifically about some of the scientific questions we might address. And elaborate a little on the –

[Host, Dr. Maryellen Giger] We can’t see the screen.
[Presenter, Dr. Etta Pisano] You can’t see the screen? Ok. That’s strange, because it should be – let me just go back to the – I push ‘share screen’. Can you not see my screen at all? ‘Share screen’. Can you see it now?

[Host] No. I think once you hit ‘share screen’, can you – oh –

[Presenter, Dr. Etta Pisano] There. Now can you see it, do you see it now?

[Host] Yes, we do.

---

Slide 1
(14:22)

**Why Radiologists Should Participate in MIDRC**

Etta D. Pisano, MD FACR
Chief Research Officer
American College of Radiology
March 12, 2021

---

[Presenter, Dr. Etta Pisano] I don’t know what happened, it just – anyway, so as I said, I’m going to expand on what Curt said. It’s not letting me advance my slides – there you go.

---

Slide 2
(14:36)

### Registries versus Clinical Trials

<table>
<thead>
<tr>
<th>Clinical Trial</th>
<th>Registry</th>
</tr>
</thead>
</table>
| • Subjects prospectively assigned to treatment or placebo  
• Evaluates pre-defined health outcomes | • Observational, real-world data collected during course of clinical care  
• Could include EHR data, genetics data, clinical data  
• Allows for inclusion of more diverse populations |

Ok, I’m going to first talk a little bit about the kinds of data you can collect in big studies like this. One of the ways I – maybe some of you know about some of my work and the ACR and the world of radiology’s work in clinical trials, which has highly curated data and patients are consented and prospectively contribute information on specific questions. Clinical trials are extremely expensive to set up and they requires an infrastructure to talk to patients directly and get their consent. And they only usually study a dozen or so health outcomes, at most. Registries, on the other hand, use real world observational data, including EHR data, clinical imaging data as well as other kinds of data streams, like genetics. What’s good about these kinds of studies, and what’s really robust about these kinds of studies, is they allow a more
diverse population to participate. Because, you know, the infrastructure to support a clinical trial is quite expensive and takes time to setup over many years. With a registry, most institutions can provide access to their data, if they so desire. Their infrastructure already exists to collect the data.

**Medical Registries**

- Data taken from patient encounters that allow study of clinical questions using real-world data.
- ACR has about 13 other research registries already launched – on pediatrics, rare diseases, thyroid nodules, Contrast mammography, TAVR, others
- ACR has numerous Quality Improvement Registries under the National Radiology Data Registry (NRDR) program
  - E.g., National Mammography Database (NMD), Dose Index Registry (DIR), Lung Cancer Screening Registry (LCSR)
- For peer review

Essentially, the main difference is that this is not highly curated data. As you know, different doctors enter data differently. We’ve gotten a little better with EHR, but we’re still at times a little messy. So it does take some cleaning of the data on the backend, much more than you would expect in a clinical trial. The ACR has a long history – I’m sure many of you were already involved in our registries. Most of them are setup for quality and safety. The National Mammography Database, the Dose Index Registry, the Lung Cancer Screening Registry etc – those are for peer-review, not for research primarily, although some research is being done in it. The CIRR, which is feeding into MIDRC, is one of the 13 registries the ACR has launched, and some of you may already have been involved in some of these other registries.

The point being, that the infrastructure to support medical registries, including clinical data, pretty robustly already exists, which is why MIDRC is so attractive to the ACR and we can help MIDRC succeed in collecting not just imaging data, but the clinical data as well.

**Why Registries are Valuable**

- “Real world data” — longitudinal information on conditions — individual and aggregate
- Allows for:
  - Education
  - Research and development
  - Assessment of Novel and more effective therapies and disease management
  - Using AI and machine learning, can discover new factors that impact patient management
- Multi-institutional registries allow for diverse patients, institutions, and treatments to be represented in the data
Of course, this is hugely valuable, not just for research but for education, and for answering specific questions about COVID. We believe that the combination of the imaging data with the clinical data will allow us to figure out which therapies are most effective, how to manage disease more effectively, a lot of the things that Curt discussed. And, you know, candidly, we don’t need AI just to do what we can already do. We radiologists are already pretty good at finding pneumonias, we’re already pretty good at finding the gross things, the big things. We need to be able to make conclusions from things that are not so obvious from the images. And so we hope, with large datasets, as MIDRC will amass, we’ll be able to do those kinds of discoveries and hypothesis generation.

**Benefits to research/clinical community**

- **Research**
  - Provide data for hypothesis generation
  - Supplement clinical trials and other kinds of data
  - Aggregated data especially helpful for:
    - Learning more about rare diseases
    - Discovering unusual outcomes and complications in large populations
    - AI/machine learning/radiomics
    - Conducting virtual clinical trials
- **Education**
- **Public health surveillance**
- **Reimbursement**
  - Provide evidence for payment justification

In addition, it’s really possible to do pragmatic clinical trials within registries. So that you can, especially for rare diseases, not COVID per se because COVID is unfortunately not rare, but the rare complications of COVID, might be more accessible to interrogation in this registry. So we know post-COVID brain fog is happening, even for some people who weren’t really sick. We know a lot of the things that are kind of uncommon in COVID may be more discoverable with these big registries. And, you know, candidly, COVID is likely to become a chronic illness. I hate to say that, we’re all hoping for a nice summer coming up, after vaccinating ourselves and our families. You know, candidly, it might be like a flu in the long run, where we have to get annual vaccines, and there may a role for following the disease over time.

And finally we have applied for, for the whole MIDRC registry, because we have not heard from Medicare. If you participate in this there may be additional payment from Medicare, and we hope that comes through, but we don’t know yet if it’s going to come through.
So what is the data that goes beyond the images that, you know, there’s limited dataset, which means there is some PHI, in order to determine that the patients are diverse, we’re going to need to know which zip codes they’re from, for example, and what dates are their exams. Both of those things are considered personal health information. There is a limited dataset that will be included in the submission. The minimal dataset to contribute to MIDRC is very, very lightweight. But if you want to do everything, if you want to put in more than the minimum, you can put in the exams plus the demographics, plus the lab tests plus the medical history etc. And as Curt mentioned, we are going to be linking with the N3C, which is a much broader dataset. But these kinds of things are protected health - many of them, the zip code for example, are protected health information. So there is protection for that. We are planning to protect it, not put it out in the public domain, but it does allow us to understand the value in terms of representing a diverse population if we have that information.

So I just want to elaborate on a few of the clinical questions. You already heard predicting outcomes is one of the things. Well we actually already have a project funded to do that within the existing MIDRC, led by Despina Kontos, of the University of Pennsylvania, and Joel Saltz, from Northwell Health. The two of them are looking at the radiomics and machine learning. They’re going to use radiomics and machine learning to try to predict patient outcome in patients with COVID pneumonia who get admitted to the ICU. So the way we do that is that we’re collecting cases with COVID and a control group are people who had COVID tests but were negative who ended up in the ICU, and we’re going to see if we can predict which patients can progress to more progressive disease. And then, of course, which factors about those
patients might correlate. Not just the imaging factor, but the other factors as well – demographic characteristic, other diseases, and of course, the imaging findings.

**Other Clinical Questions that Can be Answered Using MIDRC Data**

Slide 8

(21:23)

Question 2: What brain findings might correlate with COVID long-hauler brain fog? (see the article in NYTimes, 3/8/2021 - even some people who had asymptomatic COVID are reporting this long-term problem)

Question 3: As COVID becomes a chronic problem that recurs regularly, like the flu - Are there some constellations of treatments that work best for certain constellations of imaging findings – in the lungs, in the brain, in various age groups?

There are other things, too. I mentioned the brain fog, there was just an article in the New York Times last week where people were reporting this problem, even people who had asymptomatic COVID. Maybe there are some findings that we can ascertain in their brain imaging for these people. We don’t know yet, we haven’t started this project. But that’s an example of something that we might be able to do and study. In addition, as I mentioned, COVID might become like the flu, become a chronic problem over time. And I don’t mean, chronic COVID, you know, the long-termers. I’m talking about a recurring problem over time, and are there some constellation of treatments that we can determine based on the correlated imaging findings in the lungs, in the brains, in various age groups, that might be useful for the recurring annual event that we’re expecting.

Slide 9

(22:13)

So I put up this last slide just to remind me to say, you know, these are the obvious. This was from a paper in *Radiology*, one of the first papers that was published last summer when COVID was obviously - we recognized it was first affecting the brain. And these are very obvious findings and any radiologist, probably a first-year resident, could say there’s something wrong here. We really are – you know, with AI and machine learning, we really are hoping to be able to find much more subtle things that correlate with poor outcome, brain fog, whatever. So, I
mean, that’s the real promise of AI. And that’s the real reason to contribute to this, because each of us, as an individual, can recognize things, obviously more subtle things that are on this slide. But, you know, AI potentially could recognize even more subtle things that are not really visible to us without a lot more study and analysis, without the big cohorts that will be available in MIDRC. So I’m now going to close and finish up and pass it off to the technical side, where we have Adam Flanders and Mike Tilkin are going to talk about how to get your data and the structure of MIDRC.

[Presenter, Dr. Adam Flanders] Thanks, Etta. Hello, everybody! Can everybody hear me and see my screen?

[Moderator, Dr. Paul Kinahan] Yeah, thanks, Adam.

[Presenter, Dr. Adam Flanders] Ok, wonderful. Hi, I’m Adam Flanders. You heard that ‘what’, you heard the ‘why’, so Mike Tilkin and I are going to talk about the ‘how’, what are the nuts and bolts for getting data into MIDRC. And if you have concerns or fears, we’re here to help allay your fears, and help you realize it’s not as hard as you might think it is.

MIDRC Team is Experienced With...
- Management and curation of DICOM/clinical data through clinical trials, registries and data science challenges.
- De-identification of PHI and HIPAA regs.
- Secure transfer of large volumes of medical image data.
- MIDRC has data published with more in the pipeline.
- Securing appropriate permissions with leadership at donating sites.
  - IRB
  - Compliance, privacy, regulatory officers
  - Preparing appropriate documentation
  - Data Use Agreements
  - Data Sharing Agreements

And part of the reason why we want you to feel comfortable with this is that the MIDRC Team has a lot of experience in this concept, meaning we have a lot of experience with management and curation of DICOM data and clinical data through clinical trials, registries and data science
Challenges. We have a lot of experience with de-identification of PHI and HIPPA regulations. We know all about secure transfer of large volumes of medical imaging data. MIDRC, as you heard, already has data published with a lot more in the pipeline, so it’s not like this is just sitting still. We’re actually actively getting data out there. And we also have a lot of experience with securing appropriate permissions from leadership at donating sites, at your organization, you know, helping with IRB, making sure the proper paperwork is filed with compliance and regulatory officers and so on. And we have boilerplate data use agreements and data sharing agreements that we’re going to share with all of you to make this process relatively straightforward.

What is the Pathway to Contribution?

So the pathway. We hope we’re going to be able to answer some of these questions, although we’re kinda doing an overview, we want to help you understand how to get started, what the regulatory and legal requirements are, what imaging and data might be needed (you’ve heard some of this already), how do you submit these cases, how do we de-identify these cases and when will the data be made accessible and so on. And then what will we do to help you support that entire process.

So, getting started, you know, you’re already getting started – attend the Town Hall, you’re already here! Sign up through the MIDRC portal, if you haven’t already done so, to tell us about yourself and what kind of data you might have. One of the keys is to identify a champion at
your site, if it’s not already you. And to try and identify key personnel at your organization, which can vary by site; who’s your compliance privacy officer, who do you need to talk to at your IRB to get this through. Make sure it passes muster with your department leadership, and see who can help you on the IT side in your own organization to help with the whole process.

But, with all that being said, we’re here to help shepherd you through that process. Mike!

[Presenter, Mike Tilkin] Yeah, thanks, Adam. Yeah, as Adam says, we’ll actually help you get started, the first thing we’ll do is make sure we have a MIDRC liaison to work with you. And this will help with things like standard data transfer agreements. This isn’t stuff you have to come up with from scratch. We actually have agreements that we’ve used across many sites. We’ll help guide you through that process and your tam through that process. Likewise, as you go through the steps of obtaining IRB determination. There’s a global IRB, but your local IRB needs will also vary, we’ll help you navigate that. Likewise, interfacing with your local privacy folks, as well as your legal team. So again, as I think Adam said, we’re here to help, we have a lot of experience doing this. And as the process goes through, everything from dealing with data identification extraction issues that have regulatory implications or any legal components, again, we’ve got very experience folks that can help you work through that process.
As I’ve described, the data starts with really a core, and this begins with imaging information and contextual information, demographics and test results. And that’s really the minimum data set, and we want to make sure we’re collecting really across the board.

And then as you’ve described, we have an extended dataset, if you will, including more clinical information, socioeconomic information, all the things that will kind of help us over the course of time answer some of the clinical questions that was described earlier and continue to track the progression of disease.

And finally, as was also described, we have a good array of partners. Our goal really is to interface with the other organizations collecting information, partner with whomever we can, frankly, to get as much value out of the data we’re collecting and interface with other data collection efforts as much as possible.

So, the submission process. So first, as you’ve heard, we’ve got really two intake frameworks through the RSNA and the ACR. So as you kind of identify your interest in participating, we’ll help work with you to figure out the most efficient approach to getting data into the system,
both interacting with your local systems, as well as the overlying framework. So these kind of represent the two pathways. As I said, we’ll kind of start the process by helping understand your needs, and by helping understand what kind of makes the most sense, if you will. Because our goal, at the end of the day, is to keep the burden as low on the sites as possible.

So just to give you a feel for what this kind of looks like, kind of at the local site there’s really this initial acquisition process, where we’re getting the imaging and the data in to local software that will help with the de-identification process, the contextualization process, and really get data in a safe way to upload, frankly. So in this example, this is the ACR CIRR registry that was described earlier. So it’s an extensive dataset, it’s geared for a limited data set, HIPPA information, and where quality assurance and other de-identification processes can occur. This gives the ability to kind of do analysis.

The goal here is to get a safe harbor dataset created, that we can then transmit to our global archive. And this is where we really get that kind of advantage of bringing together a vast array of data.
And so really the point I want to make here, you know, this is a robust established pathway. Likewise on the RSNA framework, there’s a similar kind of mature pathway that involves, again, this kind of local data assembly process, a data aggregation process, deidentification to help you get the data secure and transferred up into the central archive with the RSNA. And again, annotations and curation processes. And then ultimate kind of transmission to the MIDRC public data commons.

So in all cases, our goal is to create high quality data that’s been harmonized, that’s gone through a deidentification process, that’s been appropriately mapped to standard expressions so that it can support the research goals that we’ve described. And then again, the deidentification process is robust, protecting the patients and institutions and all of us to ensure really the high quality result on the other end.
Once the data’s in our public archive, and this is hosted and managed by our Gen3 collaborators, we’ll have public access to the data. So this’ll allow downloading the imaging data, stratified clinical data, things of that sort.

And then is finally the support. So this is just a couple of shots of some of our online documentation, so these are the FAQs that we have on the left. On the right, you can kind of submit at ‘contact us’. So we’re going to continually provide both background information to help you through this process, as well as folks you can connect with directly to help you navigate this process. So that’s really the kind of high level. I think next our goal was to take questions and address some of this in a more interactive fashion.
QUESTION AND ANSWER

[Moderator, Dr. Paul Kinahan] Great, thanks, Mike. Thanks, Adam. I’m Paul Kinahan, and we’re going to segue way to the last part of the program, which is just an open format question and answer period, where we take your questions and hopefully provide answers.

Part of the motivation is that this is just the first of what we anticipate to be several meetings over the coming year as we work through the process and get the lines out, as you heard earlier.

And so we also appreciate questions so we can add clarifications for people later, and also so we can add it to our ‘Frequently Asked Questions’ that’s on the website that Mike just referred to.

So feel free to put questions in the Q & A box. I will mention that there was a question earlier about closed captioning. Thank you for that pointer, we’ll address that, see how we can do that, for both this meeting retrospectively and going forward.

So in the Q & A box, the first question I see is

**Q: (33:31) “How do you deal with institutions that are resistant to data sharing?”**

And that’s a great question, and that’s actually one of the reasons that we’re here. And I can think of several hopefully humorous answers, but today we’ll get more serious. Let me start with Curt and Etta on that part, to see you how guys would address this, because you’ve seen this, and then later we’ll open this up to the other panelists to see if you have other thoughts. Curt, do you mind taking a first stab at that?

**A: [Dr. Curtis Langlotz]** Sure, I’d be happy to Paul, and thank you. It’s a great question and we do see that from time to time. And I think the first way to address it is using the kind of call to action that both Maryellen and I referred to. That these data are a resource that can be deployed to benefit us all. And, to the extent that we can make secondary use of them after they’ve been used for patient care, reimbursement and all of those things, that’s a great societal benefit and it’s part of just helping us improve the world, and improving healthcare.

I think also you can point to things like HIPPA, so when data are de-identified they then move into the realm of data for research. There is a broad view, a legal view, in fact that those data can be used for research. And if the data have already been collected and retrospectively used for research, that’s a very typical use that most IRBs find acceptable. So, in the end, we find that most institutions, not all, but many institutions can become comfortable with that. And we’re obviously happy to share many of the security arrangements that we make, to make sure the data stays secure, even as it’s de-identified. But we make great efforts to keep the data secure, private, because we want to make sure the data stays secure and the institutions feel comfortable donating it to us. So Etta, I’ll turn it over to you.
**[Dr. Etta Pisano]** I have one thing to add, which I think is very important, is that we have really benefitted, I’m talking about MIDRC now, has really benefitted from relationships with strong advocates at the institutions themselves. So we don’t have a relationship with your Dean, your Chair, or President of your university, or the head of your practice. You do. And so if you care about this, you can advocate for it. And find the resources. And, as Curt said, I didn’t mention this, the ACR also will have grants, both organizations are going to be helping sites that are really highly motivated to get cases in. And so you are the one who can convince people up the chain that this is worth doing, and we will then put the lawyers etc on it to help sign the documents so everybody feel safe.

**[Dr. Paul Kinahan]** Thanks, Etta. Anybody else? Maryellen, did you want to add anything to that?

**[Dr. Maryellen Giger]** No, I think that’s good. There is a note in the chat about if you do get an NIH grant, you do have an obligation to share your data. And MIDRC can serve as that data sharing portal.

**[Dr. Paul Kinahan]** Ok, great, thanks. That’s great, and this is a central point. And I think advocacy is key word that we heard there, too. Another question came up, and this I think I would defer to Mike and Adam, it’s

(37:27) **Q:** “Are organizations outside of the US welcome to participate in uploading and sharing data?” Adam, would you have an opinion on that?

**A:** **[Dr. Adam Flanders]** I believe, and people can correct me on this, but I think that we’re allowing submissions from anybody.

**[Dr. Curt Langlotz]** Paul, I can also clarify that, that is true. Anyone can contribute data, although I don’t believe we can provide the grants, or the contracts, to support the donation of data from international sites, but we can certainly accept it.

**[Dr. Paul Kinahan]** Right. Thanks, Curt. Thanks, Adam. Next question that was raised, and please put your questions in the Q & A utility, I think this would be for Maryellen perhaps

(38:25) **Q:** “Will the center use the database to evaluate new AI tools?” I think the question is will MIDRC itself be doing some evaluation of the AI tools with the data.

**[Dr. Maryellen Giger]** Yes, good questions. So MIDRC includes – a large portion is the infrastructure to create this registry, but another portion is to create AI algorithms, but also to develop tools for others to help evaluate their methods. Also as data comes into MIDRC, we will sequester a certain percentage of all data, and this sequestered dataset will not be available to anyone. However, until it’s time for methods to be evaluated. And in that sense, let’s say an
investigator, or potentially a company, wants to evaluate their algorithm for a particular clinical task, a particular population, we will be able to run that within MIDRC and give them a score based on their clinical question on the sequestered dataset. They won’t get anything else except the score, and they might come back the next year, we would draw a similar distribution of cases to evaluate it. And we want to do this because 1.) this maintains the integrity of that sequestered dataset for testing its algorithms. And hopefully then we can communicate the performance better to both the public and potentially any regulatory body.

[Dr. Paul Kinahan] Thanks, Maryellen. But this would be, just to clarify too, this would be in addition to the main purpose which is to allow other people to use the data as well, right?

[Dr. Maryellen Giger] Yes, the sequestered is probably going to be 20% or less than the entire database. A very important step, but not the main.

[Dr. Paul Kinahan] Great, thanks, Maryellen. Next question, I think for Bob Grossman, at Gen3. I think probably you’re in the best slot to answer this

(40:33) Q: “Can you expand on plans to pool data with shared data, or expanding standardized access to other publicly available registries?” Now, to be quite honest, I don’t know if by ‘registries’ that’s other imaging registries or other EHR. So maybe, Curt, you can weigh in on that as well. But Bob, do you have an answer?

A: [Dr. Bob Grossman] Yes, the MIDRC commons will expose – we can answer it in several ways, and I might need a little clarification – the MIDRC commons will expose an API (application programming interface) and it can participate in different federated queries across other imaging commons or medical commons or other types of commons, that expose API. There’s a pretty active set of working groups that are working on how to do that, and we’re participating in those. There’s also been a number of conversations that have been very specific that maybe Maryellen could address, on specific other commons that we can integrate clinical with machine data we have in MIDRC. So, before I pass it to someone else, technically the MIDRC commons is designed to sort of support different types of ways that we can federate queries in different ways. We’re also actively doing something called data submission ID, so if data is submitted by a particular time that contains imaging sequencing and clinical data that go to separate repositories, there’s a fairly easy way, as long as they’re submitted at the same time, to bring those back together. There’s a lot more I could say, but maybe I’ll pass to Maryellen for some specific initiatives that we’ve been talking to people about.

[Dr. Maryellen Giger] Thank you. And we have been talking to various other COVID registries, for example N3C, and looking at how these two registries will communicate, in a sense, up in the cloud there. So stay tuned for that. MIDRC collects some clinical data, but a lot of it will come potentially in collaboration with these other registries.
[Dr. Curtis Langlotz] Maybe, Paul, I’ll just add briefly, that some on the call may be familiar with the notion of an honest broker? So the idea is that as you submit your data, it will of course be de-identified. So the medical record number in it will be removed, and what will be added will be an identifier, that’s known to this honest broker that also knows the identifier of that same patient who might have their EHR data submitted through N3C. So those two datasets then can be matched up in the cloud, as Maryellen says, so we can analyze these multi-modal datasets on the same patient, the imaging and all the clinical data that’s been collected. I think that N3C now has data from over 44 institutions, and over a billion rows of EHR data. So there’s a lot there to work with.

[Dr. Paul Kinahan] Great, thanks, Curt.

[Dr. Bob Grossman] Can I say one other thing real quick? This is a particular interest of mine and I’m actively working with others to develop technologies to make this easier, so if there’s anyone in the audience who has a particular interest and particular ways of pooling data to make it easier for researchers, please reach out to us, it’s something we’re quite interested in.

[Dr. Paul Kinahan] Great, thanks Bob, thanks Curt. The next question I think I’ll take.

(44:27) Q: “How is it ensured that manual annotations on data from different centers follow the same protocol?”

A: That’s a great question. And I think, in brief, one of the key foundations when we formed MIDRC was data quality and harmonization, so we actually have an active group that is pursuing that, the entire process. Because, as you heard, they’re different streams for data to come in, it’s coming in from multiple sites, there’s different annotators, different software. So we’re in the process of developing SOPs for data quality and harmonization, which would apply to annotations, as well as other things, acquisition protocols, de-identification to name a few. One of the key things is first to estimate whatever the quality is, and annotation falls into that. And then to report on it. And that’s something we intend to deliver along with the images, as well too. But that is a work in progress.

[Dr. Adam Flanders] And, Paul, I just wanted to add that I think the goal is what we want to do is catalogue, because obviously we could have a lot of similar types of annotations, we want to make sure that if we already have a particular set of annotations on a similar set of data that’ll fulfill a particular research need, that folks don’t need to go back and, you know, start all over again. So that we’re going to try and maintain a detailed catalogue of what kinds of annotations we have on the data as we move along, that we either produce internally or get contributed to us.
[Dr. Paul Kinahan] Ok, thanks Adam. The next question is

(46:18) Q: “What imaging modalities will be included?” In particular the question was ultrasound. I think I’ll take a brief answer on that, see if anyone else wants to chip in.

A: The predominant data right now is X-ray radiographs, then the remainder CT. But we are getting other things and the idea is eventually to take in anything. But we’re focusing on the X-ray and CT first because of the relevance to COVID-19. I think we have gotten some ultrasound in? But I would defer to Adam and Mike, in particular, if you have any other information about that.

[Dr. Maryellen Giger] This is Maryellen. But I do want to say we’ve been in discussion with the American Institute of Ultrasound in medicine. And they are very interested in working with MIDRC as a place for their registry of ultrasound cases. So definitely. Keep in mind, we’ve only been live since August, so we’re moving along and appreciate all your suggestions.

[Dr. Paul Kinahan] Thanks, Maryellen. Mike, Adam, anything else to add to that?

[Dr. Adam Flanders] Well, I think one of the other datasets that we’re pursing right now with our collaboration with the American Society of Neuroradiology is to get neuro-related COVID cases, you know, brain and spinal cord now. We don’t have any in the collection right now, but we’re working towards that goal.

[Dr. Paul Kinahan] Ok, great.

[Mike Tilkin] I think we’re geared to take data from really a wide array of sources and modalities and the like. I think the goal is to create as robust a collection as we can over the course of time. Obviously, we’re just getting started but that’s certainly our aim.

[Dr. Paul Kinahan] Thanks. Maryellen, think this next question follows on with something that you were addressing earlier which is

(47:18) Q: “It looks like there are already a list of investigators with defined projects using those data. Will other investigators have the same access to the curated to do their own research, since it’s a national initiative?”

A: [Dr. Maryellen Giger] Yes. The 12 collaborative research projects that were funded within the initial funding of MIDRC, was to jumpstart development of machine intelligence, maybe two thirds of them are looking at AI methods for diagnosis, prognosis, assessing response to therapy, integration with other clinical markers and biomarkers. Other ones involve coming up with ways to do task-based distribution, how to host Challenges and how to do virtual clinical trials.
However, the non-sequestered data, the bulk of the data that will be out there, is open to everyone, not just those doing those projects. In fact, we have worked with investigators who are submitting grants on new COVID projects to help, even though it’s all open, to help inform grant reviewers and all about MIDRC, and that this data would be accessible for their research grants.

[Dr. Paul Kinahan] Great, thanks Maryellen. And if I could add, actually that’s the primary goal is we want to make this available as a national resource for everybody to access to do their own research on. Bob, I think the next question might be best addressed to you.

(50:00) Q: “Is the data portal available to people using assistive technologies?”

A: [Dr. Bob Grossman] We’ll work to make it – I’ll dig into it, I don’t really know. I do know that we’ve tried hard to make it 508 compliant. We have an API that’s open, that presumably assistive technologies can use. But I don’t know the details and I’ll have to get back to you.

[Dr. Paul Kinahan] Ok, thanks, Bob. But that was a great question to bring up. Thank you for asking that question. So Etta, I think I’m going to ask you to tackle this next one, because it’s an interesting perhaps challenging issue.

(50:55) Q: “Can people request for their own data not to be included?”

A: [Dr. Etta Pisano] So MIDRC was designed to be a public archive, for withholding of some data for use by the FDA and others to validate AI algorithms. I think it would be challenging to participate in MIDRC, per se, if you wanted to withhold your data from the public domain. I think, I can’t say zero sites that would withhold their data, I guess it’s possible we could find a way to work with you, but really the purpose of this project is to put the data in the public domain. I don’t know if Maryellen wants to add anything to that?

[Dr. Maryellen Giger] Well I’ll just say that that is something that the patient with their hospital or medical center, that’s a decision there. When data is being submitted through RSNA or ACR, that is with the understanding that there has not been any such objection. So I think that if a patient has a concern like that, this goes through HIPPA, everything’s be de-identified and it’s helping others, that’s where I think you should ask the question. But good question.

[Dr. Paul Kinahan] Thanks, Maryellen. And, by the way, going back to the previous question that we had about types of imaging data, did you want to expand on that?

[Dr. Maryellen Giger] Yes, I just want to make it really clear, that while we’re starting out with lung data, chest X-rays and thoracic CTs, our plans are to include all modalities, not just of the
lung, but of the brain and cardiac, for example. Since COVID is affecting various vascular processes. And also, we’re just getting started, but all of those will be included. And each week practically we’re talking with more and more organizations and associations on that.

And I also wanted to clarify an earlier question – there were a couple of questions about the sequestered database and how we might evaluate. I did answer one by typing in a question that companies, someone asked about could companies get an evaluation of their method. And my answer was “yes, on the sequestered dataset.” And then there was another question asking “does this mean that datasets will be licensed to companies?” And the answer there is no, because the sequestered dataset will not leave MIDRC, it will not go to the company. However, and the bulk of the data is being used for development and research. However with the sequestered dataset think of it as a service. Within MIDRC, an algorithm from an investigator or a company could be run and they will get a score. They will not touch the data or anything. However this will expedite, hopefully, getting a promising AI algorithm to the public. Because we will already have a sequestered dataset that regulatory agencies and others could come see that it’s an independent test set. So I think Justin asked that in the questions. So we’re not letting data out for commercial use to industry like that. But we will offer a service. And also it’s a way of expediting it to the public, as well as sustaining MIDRC.

[Dr. Etta Pisano] Can I just add one point to that, if you don’t mind, Paul? I think it’s important for all of us to remember this is how our patient care improves. So, you know, if we don’t have products that grow from our research, it may not have that big of an impact on patient care. You know, so this is a good thing. You know, I’m not saying we’re making billions of dollars off of these products. But what I am saying is that the patients benefit more broadly the more broadly we can disseminate these tools that are developed as a basis of MIDRC. Honestly, it’s not about making money. It’s about helping improve patients’ lives and care.

[Dr. Paul Kinahan] Absolutely. Yes, that’s why we’re here, thank you, Etta. We’re at the end of the hour. This has been a huge success. We have almost a dozen unanswered questions, what we will do is go through these and post answers to the questions and update the FAQ as appropriate. We really want to thank everyone for joining. I’ll let Maryellen close us out.

[Dr. Maryellen Giger] Thank you, Paul. Thank you to all the speakers, thank you to all the attendees, and all the organizers. I think this has been a very informative – both us informing you, and you informing us, us learning from your questions for MIDRC. And we hope to host these regularly. On the website, we do have a question and answer page. And if any of your questions are coming up that are not answered there, we hope to use them to augment that page. And never hesitate to contact us also. And we look forward to working all together for the public good. Thank you!