Gradient Health

Data Partner Manual

A world
without data
is a world
without progress.

Build a world **with data.**

In partnership with
Gradient Health

Guiding you through the proccess of **data partnering.**

Table of contents

Introduction to data partnerships:	**8-9**
What are data partnerships?	8
How we work with you	9
The return on investment (ROI):	**10-11**
ROI overview	10
Financial rewards	11
How we maintain security and privacy of data:	**12-13**
Our data protection approach	12
Security and privacy FAQs	13
How we transfer data:	**14-23**
Data transfer methods	14
Additional information:	**24-25**
Data partnering FAQs	24
Contact	25

What are data partnerships?

Data partnerships are mutually beneficial arrangements that facilitate the responsible sharing of medical data, supporting researchers around the world in the development of successful healthcare innovation.

A data partner is an organization that shares medical data with Gradient Health.

Our data partners include:
- Entire health systems
- Individual hospitals
- Practice groups
- Outpatient clinics

Data partners receive financial rewards for their contibutions, and help ensure that their patient populations are included in the development of new diagnostic and therapeutic tools.

Our team works closely with each of our data partners to make data sharing as seamless as possible.

Continue reading to find out more about how we operate our data partnerships, the return on investment that our data partners receive, and how we maintain the highest standards of data security and privacy.

How we work with you.

1. **Connect and collaborate**
Data partners connect with Gradient to explore collaboration opportunities, align on goals, and complete a simple contracting process. Gradient manages all technical and administrative details, making it easy to get started.
(See pages 14–23: Data Transfer)

2. **Secure data access**
Gradient sets up secure access to imaging data, typically via the partner's PACS, VNA, or EMR system using proven, compliant transfer protocols. Minimal IT effort is required from the partner.
(See pages 12–13: Data Protection)

3. **Data protection and de-identification**
Gradient de-identifies all data in the cloud using validated security and privacy controls. Our processes are designed to meet international data protection standards.

4. **Quality assurance review**
Every dataset undergoes a thorough quality assurance process to ensure responsible and compliant data use.

5. **Publish to Atlas**
Once approved, datasets are published to the Atlas platform, where they become discoverable and available to trusted AI developers.

6. **Developer order and final validation**
When an AI developer orders data, Gradient performs final technical and ethical validations before fulfilling the request.

7. **Revenue share and ongoing partnership**
The data partner receives their share of revenue from the developer's order. Gradient continues to manage operations, compliance, and future opportunities.
(See pages 10–11: ROI)

ROI overview.

Lead the way in responsible data sharing. Partner with us and benefit from a meaningful return on your investment, in more ways than one.

Unlock new revenue streams

When your data is used, we ensure you get a fair return. This enables you to invest new revenue into your health system and expand your services, for your community.

Focus on your healthcare delivery

We handle the heavy lifting. Our team ensures a smooth, low-effort setup and maintenance process, enabling you to focus your efforts on your healthcare delivery.

Benefit from a scalable, sustainable model

Our platform delivers long-term value. After a one-time setup, you'll have ongoing opportunities to contribute more data, at your own pace, receiving continuous benefits.

Make a measurable difference: not only for your health system but for the broader, global healthcare landscape.

Through sharing diverse, representative data, your organization plays a crucial role in reducing bias in medical AI and fostering equitable healthcare solutions to serve people across the globe.

Support communities to get the most out of their medical data, safely and fairly. Responsible data sharing is at our core.

We ensure everyone involved with data sharing is able to benefit from a fair return on investment - from our data partners, AI developers using our platform, to the very communities the data has originated from.

Financial rewards for our data partners.

Data partners receive a revenue share for their data every time it is used on the Atlas platform. As a data partner, you will receive a payment at the end of the month your data is used.

Data partner financial rewards, in action:

- Gradient closes a $3M deal for 10M patient journeys.
- One data partner provides 4M of these.
- The data partner receives $240k from this deal alone.

  Calculation:
  $3M (total deal revenue)
  x 40% (share of data)
  x 20% (typical revenue share for data partners)
  = $240k
- All future uses of the same data would generate additional revenue for the data partner.

To make data sharing worthwhile for all parties, we typically look to partner with health systems that can contribute a minimum of 1M datasets. However, we understand every partnership is unique. So if your threshold is below the 1M mark, please get in touch so we can explore more bespoke options.

Our data protection approach.

Our responsibly sourced data is rigorously de-identified before it ever gets to the end user, empowering innovators without the worry of added compliance risk.

Patient data protection is our highest priority. In addition to having extremely high security standards, all data usage is reviewed to ensure ethical usage and patient benefit.

By the time medical data reaches our developer-accessible databases, it has gone through multiple layers of checks to ensure that all patient data is removed from metadata. Our metadata redaction practices adhere to "DICOM PS3.15 2023a - Security and System Management Profiles" as outlined by the DICOM Standards Committee.

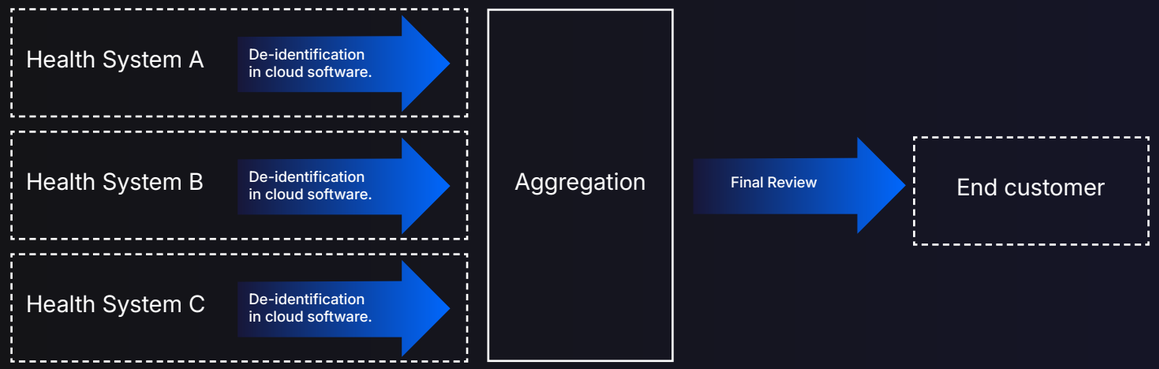
All data is de-identified using our thorough de-identification software.

Every image is scanned pixel by pixel for potential Personally Identifiable Information (PII) and redacted.

We take care of the data redaction and de-identification, so that you don't have to.

Our multistep approach to data security ensures no PII ever reaches an end customer.

Once you share your DICOM data with us, it is securely transferred and de-identified by our team in our cloud. That data is then uploaded to an aggregation system, where it is combined with other provider's data and further de-identified.



Download our security & privacy whitepaper to find out more.

Security and privacy FAQs.

When is data transferred?

No data will transfer to an external party until it is de-identified. In many cases, Gradient acts as an honest broker and de-identifies the data on behalf of the provider[1].

Who is liable for data?

Gradient Health carries an insurance policy for accidental extraction of PII. Reach out to our team for a certificate of insurance. As of 2022, our policy is through Clear Blue Insurance Company.

Who owns the data?

A facility that generates data maintains "ownership" rights. Patients have a right to inspect their data and obtain copies but they may not have the data or reports modified or stricken. Facilities may use data not only for treatment purposes but also have rights to use data with respect to educational training, quality control, and research, subject to HIPAA requirements[2].

Does Gradient own the data?

No. Ownership of the data remains with the provider. Gradient only owns any additional information that may be layered on top, including organizational data and annotations.

What is a trusted 3rd party broker?

A trusted third party broker is a group, company, individual, or institution that does not perform research, but is allowed to coordinate data usage between providers and researchers. They must be secure as well as ethically sound.

Why should Gradient be my trusted 3rd party broker?

Gradient takes care to make sure that patient data protection is our highest priority. On top of having extremely high security standards, all data usage is reviewed to ensure ethical usage and patient benefit.

How does this help the world?

Research, first and foremost, must be performed on diverse patient data if the output is expected to work on diverse patients. Gradient is bringing together datasets from around the world, from patients of different ethnicities, ages, and backgrounds to ensure that research benefits everyone equally.

1. https://irb.research.chop.edu/honest-broker
2. https://pubmed.ncbi.nlm.nih.gov/24503046

Our data transfer methods.

At Gradient, we prioritize data sharing that is secure, responsible and efficient. To achieve this, we know how important it is to adapt to our partners' needs. Each entity is different, and contexts can vary from provider to provider, so that is why we offer a bespoke approach to our partnerships and data sharing methods.

This section of the data partner brochure serves as a guide to walk you through the different ways in which we operate our data transfers. By selecting the one that is most appropriate for your use case we can help you successfully integrate with our private cloud in the most secure and efficient way.

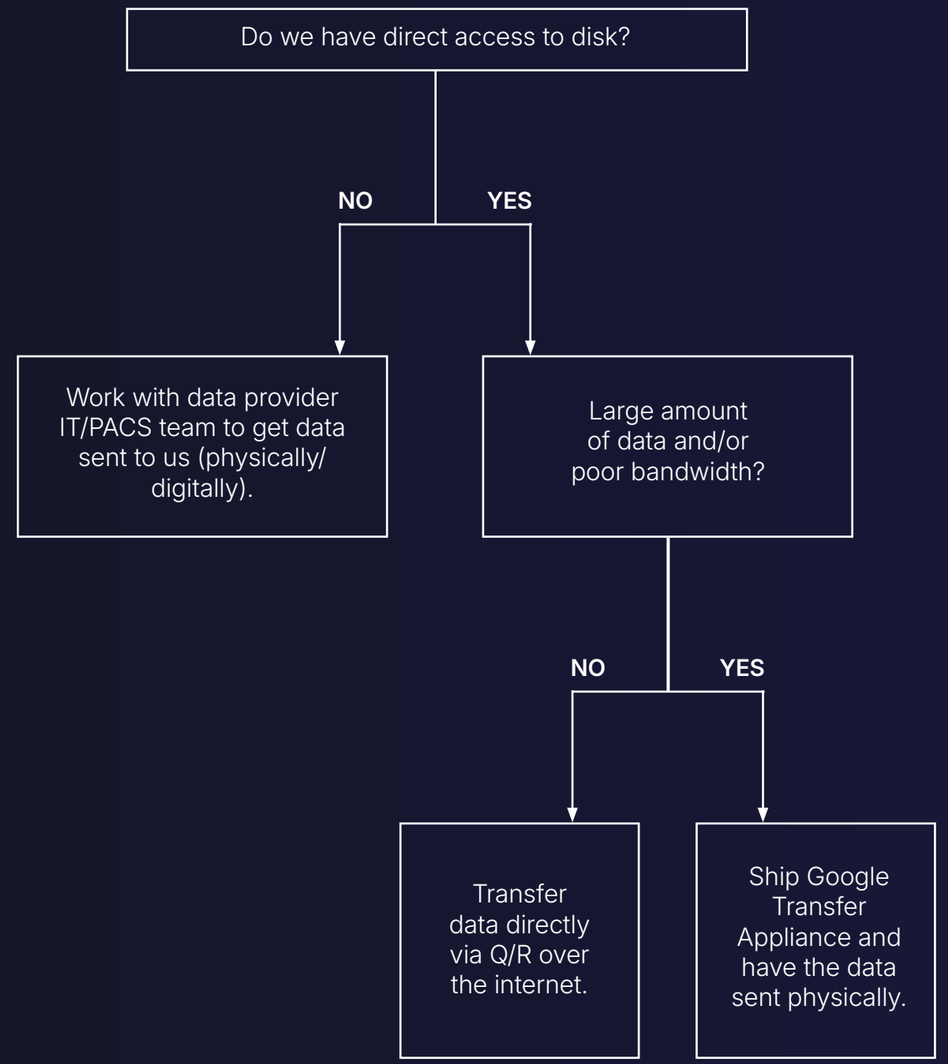
Historical Data Transfer Methods

Gradient Health's infrastructure is hosted on the Google Cloud Platform (GCP).
For each data provider, we create a dedicated project that ensures complete isolation of resources and data, keeping it separate from those of other data providers.

> " Reliable data is the foundation for building, testing, and validating the tools that will shape the next century of healthcare. At Gradient, we help reduce the operational burden while ensuring that health systems remain in control of how their data is used. It's less about outsourcing responsibility and more about enabling participation without creating another IT project for data providers."
>
> - Benji Meltzer, Chief Data Officer.

Historical DICOM data transfer flow:

To get started, we typically assess whether the following conditions are met, to guide our route forward with the data transfer.



Types of data transfer we facilitate

Direct raw files transfer to Google Cloud Platform (GCP) 17

Google Transfer Appliance (TA) 18-19

Direct upload into GCP 20

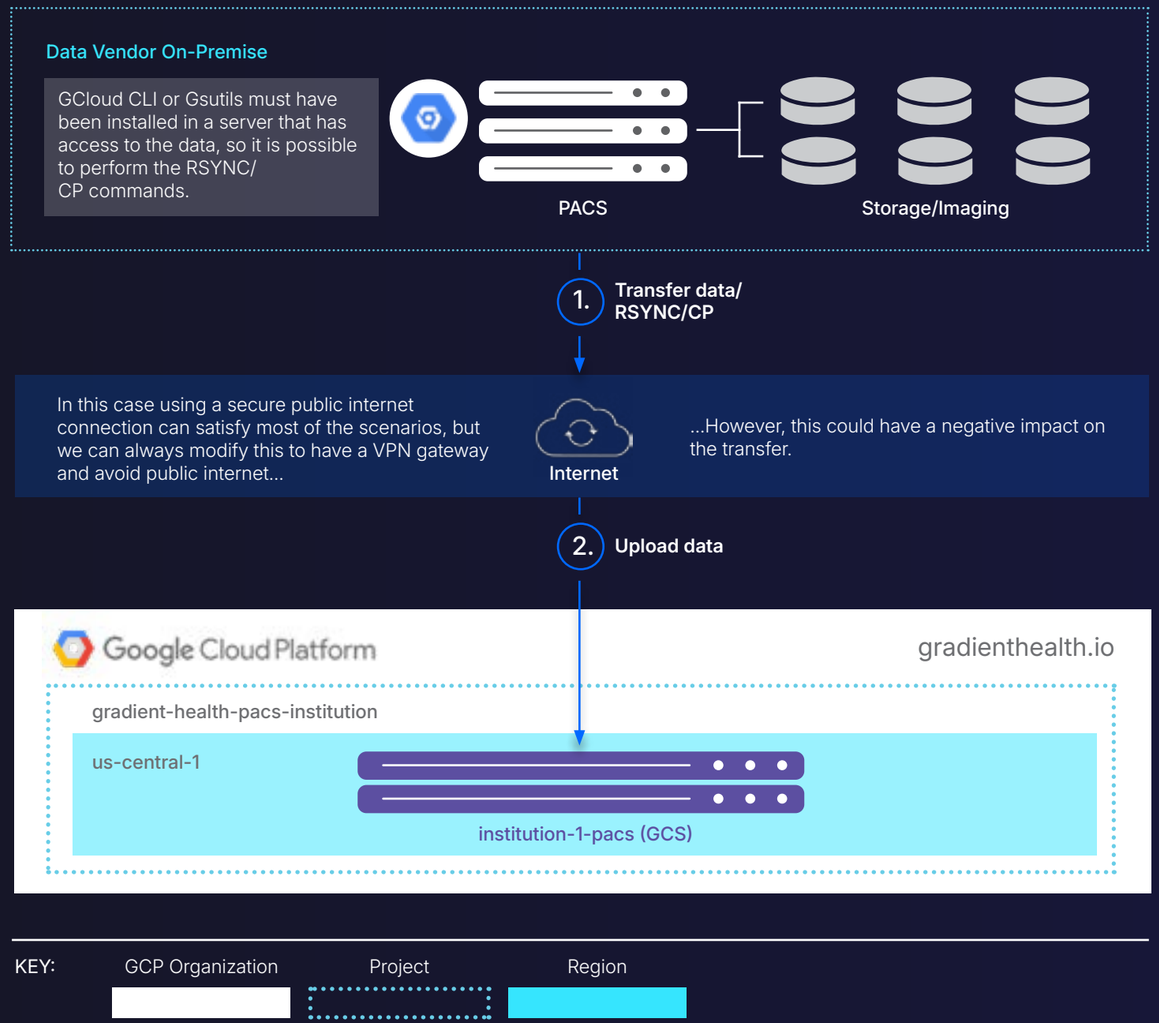
Transfer using a Virtual Private Network (VPN) 21

- Using a Picture Archiving and Communication System (PACS) server 21-22
- Raw transfer 23

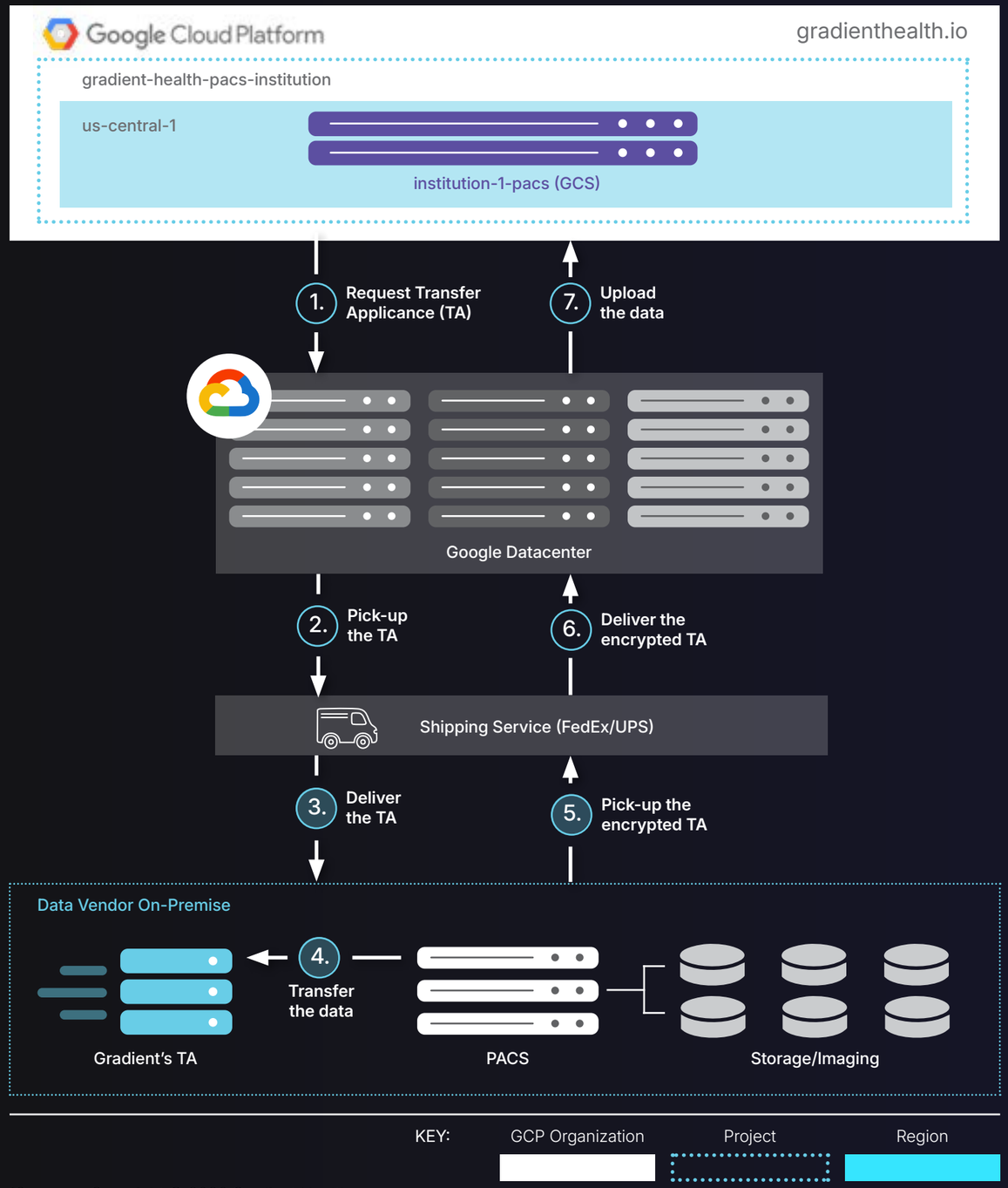
Direct raw files transfer to Google Cloud Platform (GCP)

One of the fastest ways to get your data into our infrastructure is through a direct transfer to our private cloud on GCP.

The following sections provide a detailed explanation of the different methods we use to transfer direct raw files to GCP, and in which instances we recommend them.



Google Transfer Appliance (TA)



One of the most common ways to transfer large volumes of data is by using TA, a high-capacity storage device that enables our data partners to securely ship their data to a Google upload facility. From there, Google takes responsibility for uploading the data to their private Cloud Storage.

The device is tracked from the moment it is ordered, through shipment to the data partner's premises, and all the way back to Google's data centers. In this scenario, the data provider copies the raw data from their Picture Archiving and Communication System (PACS) server, typically DICOM files and the directory structure, onto TA.

Typically, we recommend that the data partner perform this transfer themselves, and we can provide support, including scripts, to ensure the process is efficient. If needed, we can execute and monitor the transfer directly. (Please note that this option requires access to the private network and PACS servers via SSH or a similar secure protocol).

When do we recommend this method?

If one or more of the following conditions are met, we usually recommend this method:

- Large volume of data: The data volume exceeds several dozen terabytes.
- High number of files/studies: Millions or even billions of files or studies.
- Insufficient internet upload bandwidth: Local network transfer speeds need to be around 10 Gbps. If your internet upload speed is significantly lower, it may not be sufficient for this type of transfer.
- Small average file size: Transfer speed is heavily affected by the average file size. If the average size is between 0 and 50 MB, it can negatively impact the efficiency of a direct transfer.

What are the benefits of using this method?

- This is the fastest option when dealing with large [volumes/file sizes] of data or with internet that has a poor bandwidth.
- It has little to no impact on the performance of the PACS server, as we don't need access to a PACS client.
- The TA requires minimal configuration, and if no direct assistance from Gradient is needed, no sensitive access or additional setup is required.

What are the drawbacks of using this method?

- It requires coordination and time to configure the TA.
- The data partner needs to have the capacity and personnel torun the scripts, at a minimum, and they need to keep on top of the process.
- If support from Gradient is required, additional configuration needs to be done (e.g. VPN access and direct access to the servers).

Direct upload into Google Cloud Platform (GCP)

As stated before, whenever we start a new partnership, we create a dedicated GCP project in our private cloud. A project is the primary organizational unit for using and managing Google Cloud resources. It organizes all GCP resources (VMs, databases, APIs, etc.), isolates them from other projects (each with its own billing, permissions, and data), and provides a secure way to manage sensitive data within its own private cloud environment.

As part of data partner's dedicated project on GCP, the data provider will have access to a bucket where raw data can be uploaded directly from their premises to their private bucket in GCP. The advantage of this option is that it is extremely easy to configure, and by using Google Cloud utilities, transferring the data is both secure and highly efficient.

When do we recommend this method?

If one or more of the following conditions are met, we usually recommend this method:

- Low volume of data: The data volume is in the gigabytes or low terabytes range.
- Good internet bandwidth: Your internet upload speed is higher than 4 Gbps and your internet service provider provides a minimum guaranteed quota.
- Technical personnel: You have a dedicated IT team capable of running and executing scripts to perform the transfer.

What are the benefits of using this method?

- This can be the simplest option when the data partner has good internet bandwidth (higher than 4 Gbps).
- It requires minimal configuration. Gradient will provide you with a service account, which you can use to access Google Cloud CLI tools, with minimal to no interaction from the team at Gradient.
- If the bandwidth is high and the data volume is low, this can be a much faster route than using an alternative method, for example, TA, which involves shipping and also requires time for the actual data transfer.

What are the drawbacks of using this method?

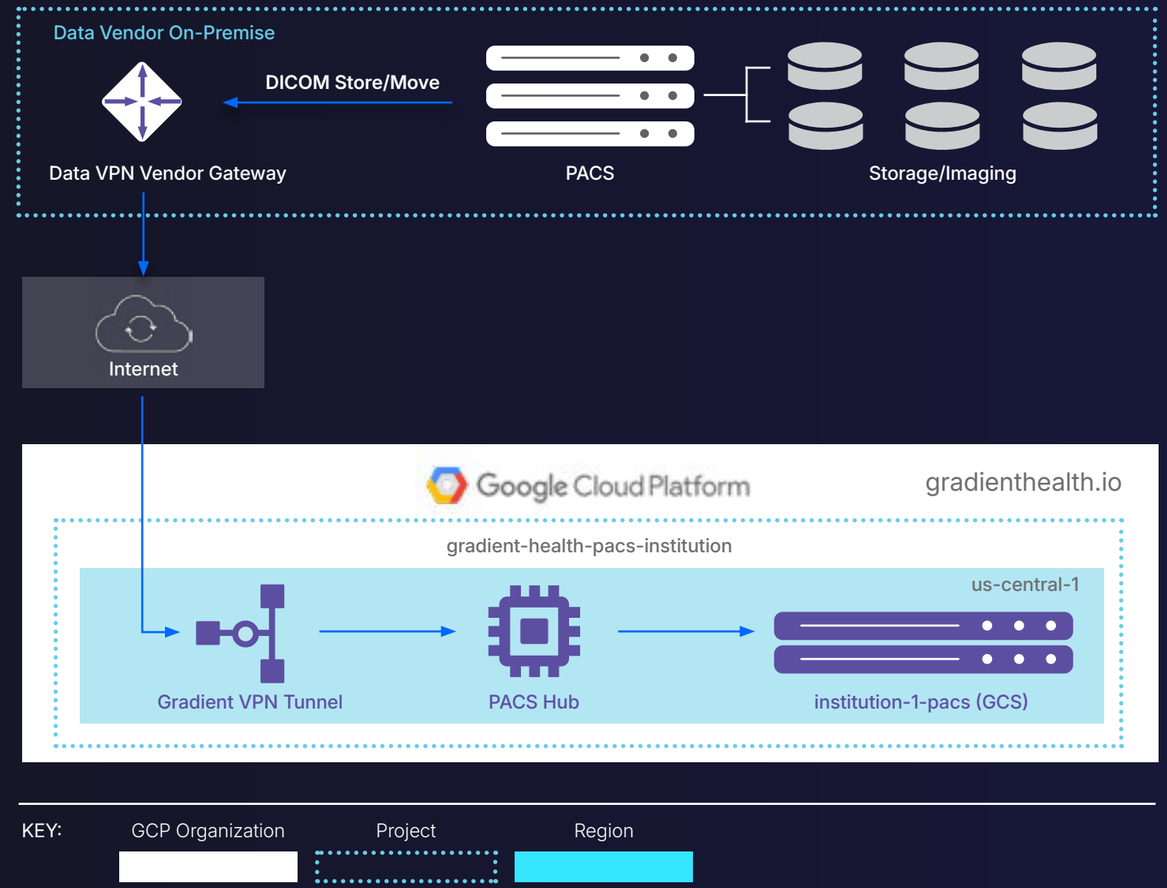
- This approach requires a certain level of technical expertise to install, configure, and run the Google Cloud Command Line Interface (Gcloud CLI).
- It heavily depends on internet bandwidth. For example, if the connection is not stable or guaranteed the transfer speed may be suboptimal.
- Similar to the TA option, the data provider must understand the directory and file structure of their PACS server to efficiently execute the transfer scripts.

Transfer using a Virtual Private Network (VPN)

It is possible to use a VPN connection to the data partner's PACS server to transfer data. In this scenario, both companies configure a VPN connection to transfer studies from the data provider's PACS server.

The following sections provide a detailed explanation of the different methods we use to transfer using a VPN, and in which instances we recommend them.

VPN: Using a Picture Archiving and Communication System (PACS) server



VPN: Using a Picture Archiving and Communication System (PACS) server *(continued)*

By using a standard PACS server and the DICOM protocol, data partners, can share their studies with us using either of these two methods:

1. Live connection: Gradient maintains a direct connection to the data provider's PACS server and request studies on demand, based on the customer's acquisition preferences.
2. One-off integration: Similar to the live connection, but instead of on-demand requests, Gradient uses reports previously provided by the data partner to identify and request all relevant studies. Alternatively, a discovery job can be run to identify studies based on a set of filters (such as date, modality, etc.).

When do we recommend these methods?

- Live Connection: This is a convenient option when your data is changing a lot on a daily basis, and you don't want to provide direct access to your physical server to Gradient Health or a third party to extract the raw files.
- One-off integration: If your data is not changing a lot on a daily basis, this option is suitable if you're only happy to grant the Gradient team direct access to your PACS application server. This integration requires you to configure your VPN, and sometimes provide the reports.

What are the benefits of using these methods?

- This option potentially involves the least amount of configuration from the data provider.
- It reuses the data provider's current infrastructure and still provides a secure connection.

What are the drawbacks of using these methods?

- It heavily depends on internet bandwidth. If the connection is not stable or guaranteed, the transfer speed may be suboptimal.
- It uses the production PACS server, which means the transfer could negatively impact the server's performance.
- To avoid issues with the server's performance, it may be necessary to define maintenance windows. During this, transfer jobs can be executed and it would limit transfers to specific periods of the day (potentially making the overall process significantly longer than with other options).
- The PACS server and DICOM protocol may introduce some additional latency.
- The VPN gateway may also introduce additional latency.

VPN: Raw transfer

We can directly copy your raw files into our infrastructure using a VPN for secure communication, ensuring an end-to-end secure connection. This method is the least common, as it offers few advantages over direct upload to GCP, and the use of a VPN may introduce additional transfer latency. However, there are some specific rare instances where we would recommend this method as the most suitable approach.

When do we recommend this method?

When all the following conditions are met, we typically recommend this method:

- Low volume of data: When the data volume is in the gigabyte or low terabyte range.
- Good internet bandwidth: Your internet upload speed exceeds 4 Gbps, and your internet service provider provides a guaranteed minimum bandwidth.
- Technical personnel: You have a dedicated IT team capable of running and executing scripts to perform the transfer.
- High security requirements: Certain standards and privacy laws may impose strict policies on data transfer.

What are the benefits of using this method?

- It is a suitable method for when the data partner has good internet bandwidth and needs to comply with strict security policies.

What are the drawbacks of using this method?

- It heavily depends on internet bandwidth. For example, if the connection is not stable or guaranteed, the transfer speed may be suboptimal.
- The VPN gateway may introduce additional latency.
- The data provider must understand the directory and file structure of their PACS server to efficiently execute the transfer scripts.
- The data partner needs to have the capacity and personnel to run the scripts, at a minimum, and they need to keep on top of the process.
- It might require additional servers and infrastructure to be set up by the Gradient team.

Data partnering FAQs.

What is a data partner?
A data partner is an organization or entity that shares medical data with Gradient Health to help advance medical research and accelerate innovation in healthcare. Our data partners include entire health systems, individual hospitals, practice groups, and outpatient clinics.

Why should I become a data partner?
By becoming a data partner, you help ensure that your patient population is included in the development of new diagnostic and therapeutic tools. This means your community benefits from more equitable, representative medical advances. Additionally, data partners receive financial rewards for contributing valuable data that supports global research and AI development.

Is it legal and secure to share data?
Yes. Laws vary by jurisdiction, but in the regions where we operate, sharing de-identified medical data for research purposes is legal and widely supported. Gradient Health uses industry-leading de-identification and security protocols to protect patient privacy and ensure compliance. We also maintain a variety of externally validated certifications to give our partners confidence and peace of mind.

What sort of data do you need?
We work with all types of medical data, including imaging (X-ray, CT, MRI, mammography, and ultrasound studies), EHR, pathology, labs, demographics, ophthalmology, and ECG. We also request the accompanying medical context, such as radiology reports and relevant metadata, to provide developers and researchers with a complete and clinically meaningful dataset.

How much could my data be worth, and how do I get paid?
The value of your data depends on several factors, including the type of data, its rarity, and current research demand. As a partner, you receive payments every time your data is used by one of our developer customers. This creates an ongoing revenue stream while supporting important advancements in medical AI and research.

Practically, how do I share my data?
Our team works closely with each partner to make data sharing as seamless as possible. We integrate with your existing systems to securely extract and transfer both the dicom and report data into our platform, where the data is de-identified. We provide technical support throughout the process.

Contact us to begin your data partnership.

We are proud to offer support to all of our customers and providers.

For questions about data partnering, email us at datapartners@gradienthealth.io.

Build a world **with data.**

Gradient Health

Data Partner Manual