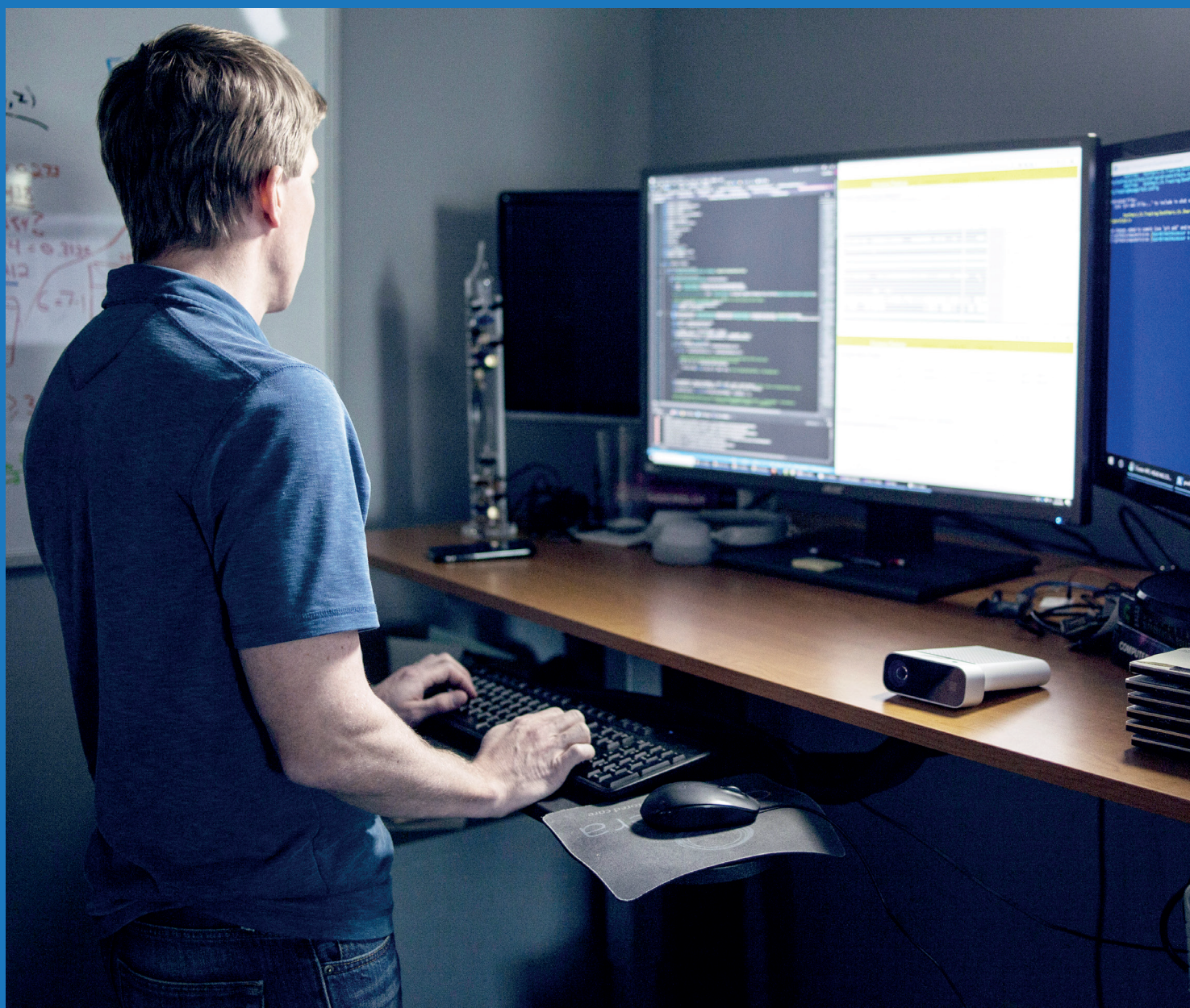


# A Digital Research Environment



## Executive summary

anDREa is an innovative Digital Research Environment for organizations that employ researchers, or participate in research programs. anDREa fulfills the needs and obligations at the intersection of organizational duties towards their researchers, the needs of researchers and the requirements of legislation, regulations and society at large.

### anDREa empowers the management of research institutions to:

- Work in the public cloud with data that is subject to European privacy laws.
- Be compliant to the requirements of the GDPR legislation, information security and ISO27001.
- Proactively respond to the increasing demands regarding information security.
- Take full advantage of the pay-as-you-go model for scalable computing and storage of cloud technology.
- Take full advantage of the innovative power of cloud suppliers.
- Fulfill their duty of care by providing their employees with fast, accurate and secure research workplaces.

### anDREa empowers researchers to:

- Create new workspaces in less than one hour
- Deploy their data-driven research programs in the public cloud.
- Engage in (inter)national collaboratives.
- Freely choose the compute and storage resources that are available in the cloud.
- Work without limitations in the choice of data, programming and modeling environments.

*"We really needed to understand the researchers. We asked them and what they asked for was freedom and we [the organization] wanted to be in control. They said: 'we want any kind of analysis, at any time, sometimes with the newest version, sometimes with a very old version because we have to reproduce something. Also, we want to invite anyone in the world when we are ready for it. Even if it is at night, we don't want to ask their IT department whether this is possible or not'. So, that was the only way of getting together and having a platform approach on one hand and on the other hand taking the researchers serious."*

**Arnoud van der Maas**  
CSIO at Radboudumc

# What is anDREa?

anDREa is a Software-as-a-Service provider focussed on delivering digital tools for researchers and research organisations while meeting legal and societal requirements for data privacy and security. anDREa provides:

---

## Infrastructure:

anDREa provides a pay-for-use infrastructure in the form of Software-as-a-Service that is completely data- and model-agnostic. Because anDREa is agnostic to the type of data and the modeling environments, anDREa supports any type of research in every academic field.

---

## Tailor-made research workspaces:

Researchers are people who, by nature, want to push established boundaries. A standard workplace is a worst-case common denominator: one size fits none. Precisely because of their urge to explore the latest technology, they quickly start organizing their personal infrastructure, often lacking adequate knowledge of security, integration and financial purchasing benefits. Using research workspaces in anDREa, each project can be offered a tailor-made research workspace within one hour, opening up all the latest innovation of cloud technology.

---

## Ecosystem:

anDREa easily integrates into an existing ecosystem of data sources, data generating equipment, external data partners, license servers, research templates, financial reporting tools and FAIR datapoints. anDREa considers itself as part of an ecosystem, not as a closed-box solution.

---

## (Inter)national cooperation:

This is a standard feature of anDREa: setting up fine-grained partnerships within and between organizations. In addition, it is possible to invite outside specialists both nationally and internationally to join a research program. Currently, researchers from all continents are participating in research projects in anDREa.

---

## The increasing demands of privacy legislation:

The requirements of privacy legislation across Europe are becoming increasingly stringent. This involves not only legal accountability, avoiding ever-higher fines and the personal liability of directors and research leaders, but also accountability to society for careful governance of the personal data of research participants. The European privacy legislation GDPR and its national implementations form the blueprint of the architecture of anDREa.

---

## The increasing threat of cybersecurity attacks:

Over the past 10 years, the threat of ransomware attacks has increased, both in number and in damage incurred. For organizations that work in the public domain, the economic, and especially the reputational damage, is enormous. Therefore, ISO27001 certified organizations and grant providers are setting increasingly stringent requirements for demonstrably being in control of information security. Even for organizations as large as a university, it is not possible to deal with all the new threats. Certainly not if collaboration via the Internet is an absolute precondition for research and education at the same time.

The anDREa initiative has been developed by a consortium of the university medical centers: RadboudUMC, UMC Utrecht, and Erasmus MC in the Netherlands. anDREa is now available to all universities, universities of applied sciences and commercial and non-commercial research institutes. Explicitly included are institutions that work on a smaller scale, such as hospitals and (inter)national research associations, and for-profit researchers. Today, anDREa has provided 1.298 workspaces that are being used by 3.515 researchers, both nationally and internationally.

## anDREa architecture

The GDPR compliance of anDREa stems from its architecture. It is compliance by design. The architecture of anDREa consists of a number of mutually reinforcing principles:

---

### **Principle 1: an organization provides its own subscription in Microsoft Azure.**

This is a fundamental principle in the entire compliance architecture of anDREa. A subscription is a defined environment in Microsoft Azure in which a specific research organization works. Everything that happens within that subscription is under the control of that organization, all data and IP are the legal property of the organization and that organization is accountable for all costs for storage and compute related to that subscription.

---

### **Principle 2: we bring our services to your data.**

anDREa brings its services to the subscription(s) of the research organization. We serve all user organizations from one central core. We accomplish this via the Microsoft Azure Active Directory, which associates a subscription of the user organization with the core of anDREa. We provide the same level of services for everyone, be it a university or a small group of researchers.

---

### **Principle 3: your data never leaves your subscription.**

Because we bring the services to the data, your data never has to be transferred to us. Sounds logical, but the transfer of your data to the platform of the service provider is still the common model. The moment the data, for which you are legally responsible, is transferred to a third party, you risk incompliance with the principles of privacy legislation. This is explicitly not the case with anDREa: your data remains in your own subscription, under your own management. You decide on the effectiveness and lawfulness of its use.



---

### **Principle 4: workspaces for fine-grained access management.**

Within your own subscription, workspaces can be deployed at the request of the researchers. Workspaces contain Virtual Machines, data storage and Microsoft Azure services, in all possible combinations. These services include serverless SQL databases, Azure Machine Learning, Kubernetes and many other Microsoft Azure services. A workspace can therefore support an entire research project by housing all data and models, but a workspace can also contain a (subset of) data from another workspace. This last variant makes it possible to easily set up a living lab for training purposes from a research project.

Each workspace has its own workspace-owner, often the principal investigator. The workspace owner has full control and full responsibility over their own workspace. They can grant and revoke access rights, and keep control over data egress. Participants can be invited from their own organization, from other organizations or as individuals via a regular e-mail address. This makes it possible to give worldwide participants access to a research project.

# anDREa architecture

---

## **Principle 5: each subscription is delivered according to the specific GDPR implementation of the country.**

anDREa supports all country-specific GDPR implementations. The core of the matter is principle 1: every organization supplies its own subscription in Microsoft Azure. Each organisation can configure their own subscription based on their own needs and their specific compliance requirements. anDREa plays no part in and has no control of the configuration of the organisation's subscription. This model applies to every country in the European Economic Area.

---

## **Principle 6: Foreign researchers work within the legal context of the country of data residence.**

All those who have access to the data are subject to the compliance requirements of the country where the research organisation resides. Suppose a Dutch institute invites a French researcher to participate in a research program. The owner of the workspace judges that it is lawful use and gives the French specialist access to the workspace in question.

The logical consequence of principles 3: your data never leaves your subscription and 5: every subscription is delivered according to the specific GDPR implementation of the country is that the French researcher has access to information that is regulated under the Dutch GDPR. Even if an institute invites a researcher from a country that has completely different legislation (for example from the USA or an African country), this principle remains unabated: every foreign guest researcher works on a Dutch research project in the context of the Dutch GDPR. This model applies to every country in the European Economic Area.

*"We have an increasing need for traceability and auditability of data. If you don't have an environment like anDREa, that means that you would always have to build it all up. You probably won't get any funding for your research if you don't have this in place. And this is not just in healthcare, this is the case in all of the research environments, also for the government or in banking. Regulations are going all over the place in terms of privacy, especially in Europe [...]. It is important that we protect the data and the privacy [of the participants]."*

**Corné Mulders**  
CIO at Utrecht UMC

---

## **Principle 7: clear separation of legal responsibilities.**

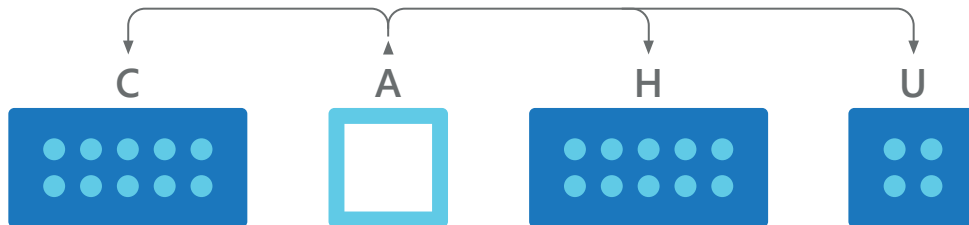
As can be seen from principle 1: each organization provides its own subscription in Microsoft Azure and 2: our services to your data, there is a clear separation of the roles between anDREa as a service provider on the one hand and the research organizations as owners of data on the other.

The research organization is a data controller as per the legal definitions of the GDPR. It determines the purposes for which and the means by which personal data are processed. anDREa is a data processor in the legal definitions of the GDPR. We only process personal data on behalf of and under the direction of the controller.

# How anDREa works

## One core serves multiple research institutions.

Based on these mutually reinforcing principles, we've built a central anDREa core (A) that brings its services to your Microsoft Azure subscription. From one core, we serve multiple research institutions such as companies (C), hospitals (H) or universities (U). This is the central architecture:



## Every institution creates its own research workspaces.

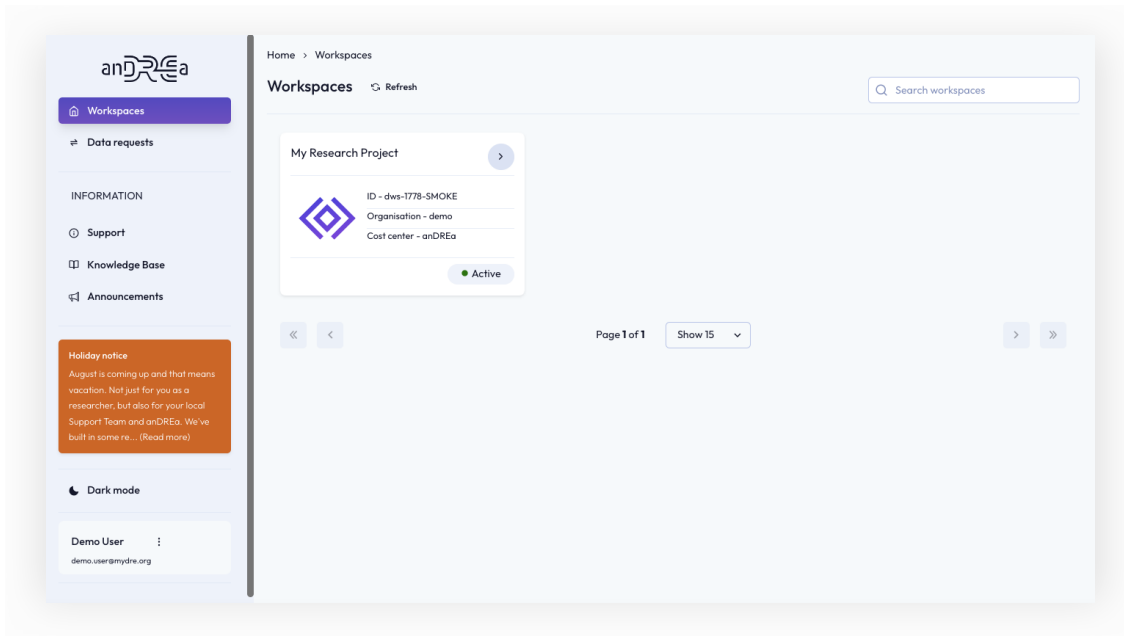
As a subscription owner, you have full control over the number of workspaces within your organization. You can create workspaces, size Virtual Machines and allocate Microsoft Azure services. You can also assign the role of workspace owner to a principal investigator, who in turn can invite other researchers to join him in his research project.

The screenshot shows the 'Create a workspace' interface. At the top left is a back arrow and the title 'Create a workspace'. At the top right is a 'Cancel' button. Below the title is a subscription ID 'ANDREA-Workspaces-DEMO-0001' with a 'Change' link. A section titled 'Workspace info' contains several input fields: 'Workspace Title \*' with the value 'My Research Project', 'Workspace Acronym \*' with 'SMOKE', 'Accountable\*' with 'demo.user@mydre.org', 'Privileged member' with a 'Search user...' field, 'Cost Center \*' with '12345', and 'Study ID' with an empty field. At the bottom left is a moon icon and a 'Next: Budget' button. On the right side, a panel titled 'Creating a workspace' shows 'Target subscription' as 'ANDREA-Workspaces-DEMO-0001'.

# How anDREa works

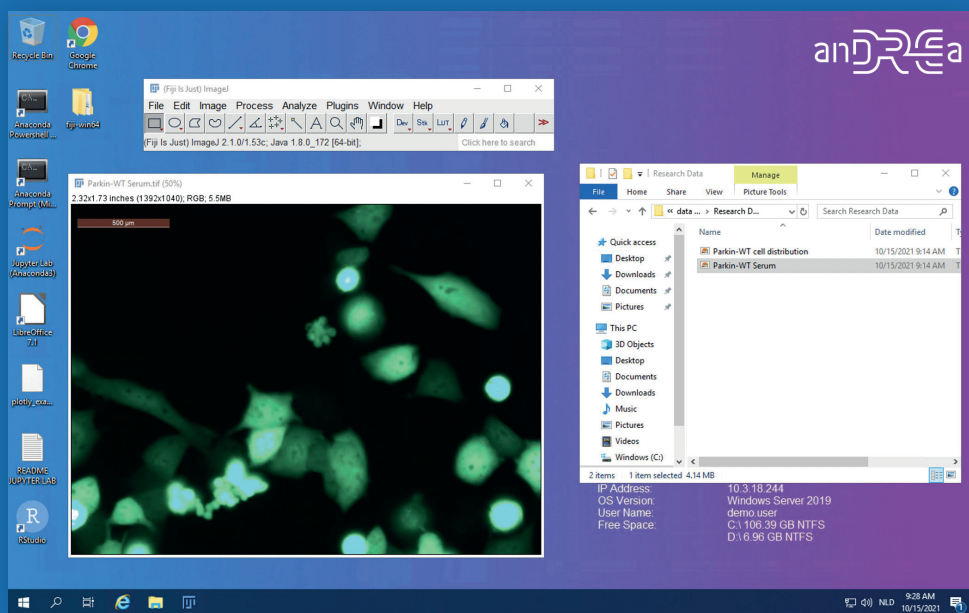
## Every researcher only has access to the workspaces he/she is invited to.

Every researcher with access rights only sees the workspaces he or she is invited to. Even if there are hundreds of projects within your organization, you would only see and have access to workspaces that you're invited to.



## Every institution creates its own research workspaces.

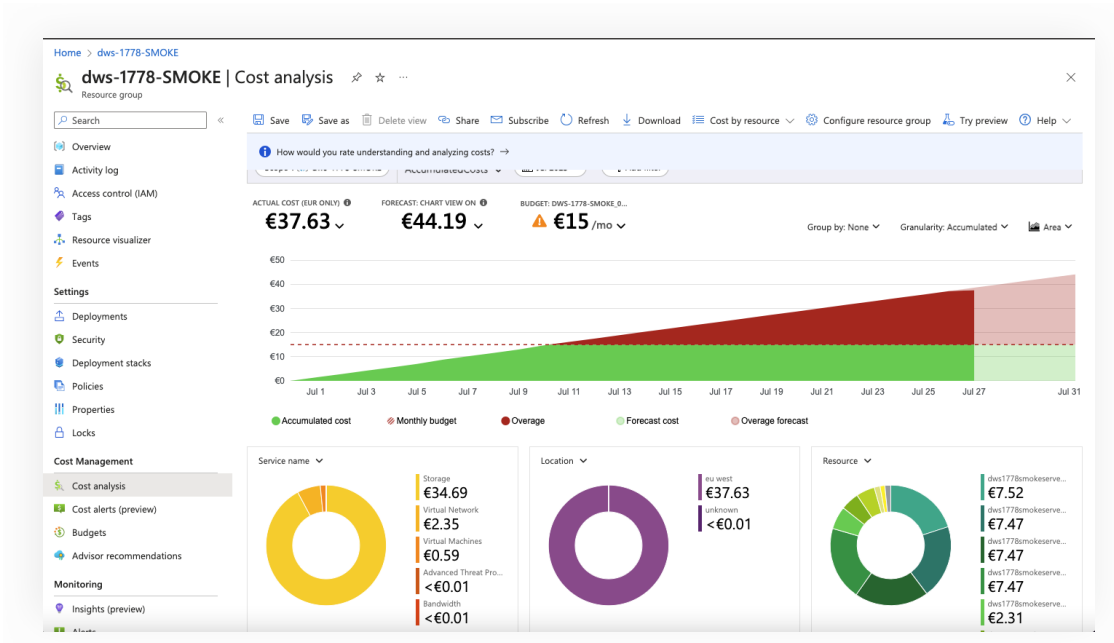
As a subscription owner, you have full control over the number of workspaces within your organization. You can create workspaces, size Virtual Machines and allocate Microsoft Azure services. You can also assign the role of workspace owner to a principal investigator, who in turn can invite other researchers to join him in his research project.



# How anDREa works

## Full cost management.

As a principal investigator you can set warnings when consumption or storage limits are reached within your research project. As a subscription owner, you can do the same on an organizational level.



## (International) cooperation in research.

Fellow researchers can be invited by the primary investigator that owns a research program. It is not a requirement that a specialist from another organization also uses anDREa, an e-mail address is enough. Currently, researchers from around the world are cooperating on research projects hosted on anDREa.





# The value of anDREa for the management of research organisations

The core of anDREa's proposition is that we help research institutes with the challenge they face regarding GDPR and information security (including ISO 27001) requirements when data is stored and processed in the public cloud. The use of anDREa is an extension of the research institute's duty of care to address questions related to the GDPR and information security.

## Meet researcher needs while retaining control.

A Digital Research Environment must support data processing by a heterogeneous group of researchers and educators. This group is dynamic as researchers and educators need to collaborate with others, including external parties, and need freedom in the use of tooling and resources. anDREa achieves this by putting organizations in control of the research workspaces.

anDREa reverses the model. The information for which a research organization bears legal responsibility remains at all times under the full control of the research organization itself, in its own protected environment in the cloud. anDREa has succeeded in building a solution that brings our services to the data of a research institute. We do this from one central core, with which we serve multiple user organizations at the same time.

*Data is safe in DRE for several reasons, explains Mirjam. "Firstly, data from Epic, our Electronic Health Record (EHR), can be consulted in a DRE workspace by using an application that pseudonyms the data. This safeguards patients' privacy. Secondly, the researcher who uploaded the data regains control of it; they determine who gets permission to use it, and no one can download it without additional authorization. Thirdly, the portal registers activity. The owner can trace who uses what data when.*

**Mirjam Brullemans,**  
Data Steward at Radboudumc

## Protect data and ensure compliance to GDPR.

The unique architecture of anDREa is that it brings its services to the research organization's data.

Typical research environment designs require the research organization to upload data to the service providers platform. With data that is subject to privacy legislation (GDPR in Europe), It presents a difficulty. If data for which a research institute is responsible is transferred to a platform of an external party as it is not clear how and by whom that information is processed under every conceivable circumstance.

## Flexibility to align research environment to your organisation structure.

Every organization that uses the services of anDREa has its own protected environment in the Microsoft Azure cloud in the form of a subscription, a container that defines its own user environment. In the model above, these are the white boxes. One subscription can be used for the entire organization, but for larger organizations several subscriptions can coexist, for example to separate faculties from each other. The organization creates its own subscription(s) and retains full legal ownership of its subscription and full control over anything that happens within it.

# The value of anDREa for the management of research organisations

---

## Elastic storage and compute.

Workspaces can be deployed to the subscription of the organization. In the model above, the workspaces are represented by blue dots. These workspaces can be as large (or small) as desired, both in terms of computing power and information storage. Data is stored within workspaces and researchers and educators have complete freedom to build their own analytics environments to conduct their data-driven research.

---

## Ability to re-use existing licences.

The organization can make its existing licenses available within its Research Workspaces. The license rules must of course be confirmed with the technology supplier e.g. the right to re-use the license in a cloud environment. anDREa is not a party to the licensing, our contribution is limited to support. The typical working model is that the workspace-owner is responsible for all licenses. This individual, mandated by the organization, knows what licenses are needed and what is available. In addition, anDREa provides insight on access to specific Virtual Machines and usage of applications. This Information can be leveraged to monitor licensing compliance.

---

## Auditability and Control.

Because the anDREa service is associated with subscriptions from research organizations, we automatically ensure that legally conclusive accountability is created for the use of (privacy) sensitive data:

- Who had access to the workspace and in which role?
- Who could see the data?
- When did that happen?
- Who gave permission for this?
- What security procedures have been followed?
- Who made a data egress request? Who approved this and when?
- Who opened internet ports and why?
- Who has made which resources available (e.g. Virtual Machines with GPU capacity)? Who turned it on or off?
- Where was the data and where was it processed?
- And most importantly: How do you know this?

This entire registration is built up at the level of a workspace, an individual research project. anDREa does that for each research project within a subscription.

---

## Built by research organisations for research organisations.

anDREa was founded by Dutch University Medical Centers (UMCs). Although we unreservedly support all research disciplines, from engineering to social science research, we can claim we are 'medical-grade'. Today, we support hundreds of studies that incorporate medical information. This medical background means that all functionality that the Digital Research Environment offers, as well as our development process and awareness of information security, guarantee the highest level of quality.

# The value of anDREa for researchers

It is important to note that anDREa is a solution for organizations, but it is built from the perspective - and in close cooperation with - the researchers and educators who use it on a daily basis.

---

## Transparency.

From the perspective of a researcher, anDREa is transparent. Every researcher can log in to the research projects that they are authorized to access. This is where the data and the modeling environment are located to perform their work. anDREa carries out the legally required registration completely in the background without imposing any restrictions on it.

---

## Cooperation with researchers inside and outside of the organization.

The workspace-owner is a user mandated by the organization who can make decisions about, among other things, the costs of a project and who may have access to the workspace and data. The workspace-owner is typically a principal investigator or department head and is responsible for inviting researchers from inside and outside of the organization to access the data of the research project. There are no restrictions on this from anDREa.

Inviting specialists or students from another institute or for-profit organization in another country is done in exactly the same way as inviting specialists and students from your own institute. All options described above also apply to collaboration with people from other institutes in the same country.

---

## Setup Speed.

Because of the duty of care of organizations, anDREa places great emphasis on the speed with which we can provide researchers with infrastructure. Only one hour is required from the moment that the decision is made to start a new research project until the researcher can actually log in to their research workspace. In addition, it is possible to use VM templates that your own Support Team can create itself; these templates contain pre-installed software. A designated local support team member can deploy workspaces for their own organization. No intervention by anDREa is required to start up the research project.

---

## Flexibility on Operating systems.

Within a workspace, several operating systems can be started. Microsoft Azure supports Windows and Linux environments. These are available as virtual machines and Microsoft Azure offers the possibility to elastically resize them. For example, it is possible to develop an analysis model on a virtual machine that works with two cores. Once the model has been adequately tested, the same virtual machine can be scaled up as needed, to 200 or more cores. Computing power is available as needed. In addition to converting cores, Microsoft Azure provides the same elastic capabilities for graphics computing units (GPUs) designed to manipulate images.

---

## Flexibility on Data types.

A workspace contains the data on which the research is performed. That can be any type of data; lists, databases, images, sound, maps, binary objects such as large scans; everything that can be delivered digitally can be stored in a workspace. The size of the storage is in principle unlimited and grows and shrinks according to need.

Data can be manually uploaded into a workspace or downloaded from externally accessible sources. These can be data sources that can be accessed via the Internet, such as Castor and RedCap. A local Network Attached Storage device or Image storage device is also possible, but this requires approval from the Security Officer and Principal Investigator because of the impact on availability, integrity and confidentiality.

---

## Flexibility on Programming languages and AI/ML.

All programming and modeling environments that are available on Linux and/or Windows within the Microsoft Azure cloud will remain available if anDREa is used. Commonly used are the programming languages Python, C, C#, C++, Ruby, R, Julia, Java. The most commonly used modeling software suites are Matlab, SAS, SPSS, Atlas.ti and Jupyter Notebooks. This list is however not complete, everything available in the Microsoft Azure Marketplace can be deployed within a workspace of anDREa.

# The value of anDREa for researchers

---

## Support for Open data and FAIR.

The anDREa service package is currently focused on workspaces to store, process and analyze data. This makes a workspace an ideal environment to prepare data for delivery to Open Data, FAIR data points, registries, etc. Ideal because the data can be processed in an unbounded way, even in a decentralized setting.

The anDREa Digital Research Environment is also used for large long-term studies in which numerous international researchers participate. Where previously researchers had to be physically flown in, which became impossible with COVID-19, researchers can now also work outside Europe on data that can only be retrieved from the Workspace with authorization. The actual storage and processing, even if the user is outside Europe, takes place fully within the Microsoft Region West Europe; via the remote desktop protocol we only transfer pixels, no data that can be tapped.

---

## Access to cutting-edge cloud services.

The cloud environment of Microsoft Azure also offers advanced technology to build new applications. Think of Machine Learning, Databricks, Kubernetes or serverless SQL. These services can also be made available within a workspace. It is necessary to check whether the desired service is available within the Microsoft datacenter region where the subscription is located as not every service is available in every Microsoft region. However, if a service is available, it is almost always possible to make it available in an anDREa workspace. Microsoft Corporation invests \$19.3 billion yearly on innovation in Microsoft Azure. That budget means that new technologies, from speech recognition to machine learning, are rapidly becoming available to end users. anDREa works closely with Microsoft Corporation to open up new capabilities for projects managed by anDREa, as well as having intensive contact with the relevant Engineering and Industry groups within Microsoft Corporation to discuss developments that are of importance to researchers and educators.

---

## Education.

In addition to research, anDREa is also used for educational purposes. Recently a number of research groups have used the anDREa platform to organize hands-on courses. For example, a genetics course supervisor prepared a workspace with a single Linux virtual machine, where the sample data and necessary software were pre-configured. Each student could log in from their own Virtual Machine and communicate with the main Virtual Machine within a workspace. Genetic analyses use a lot of computing power and not every organization can afford this. With anDREa this is easy to set up because of the 'on-demand' scaling of computing power and working memory. After the course, you could choose to download the results for your own archive, after which the workspaces were cleaned up and reused for the next course. Other examples involved students learning to train models or collect questionnaires based on real world data samples.

*"There's a lot of red-tape with storing and sharing data. I understand that the rules are important: it guarantees higher quality research and protects patients. But as a researcher, I don't want to focus on this. I want to do research. DRE takes care of the rules and regulations for me."*

**Jan van den Brand**  
Postdoc at Radboudumc,  
currently data steward at Erasmus MC

# anDREa and Cybersecurity

Confidential information used for research purposes is by definition a business model for malicious parties. anDREa has a strong set of cybersecurity measures that protect the anDREa core and all workspaces served by it from malicious access.

- 1 All users get an @mydre.org user account that enforces its own security policy regardless of the local security policy or the workplace used. Trusted devices are not allowed under the same scheme.
- 2 A @mydre.org user account does not yet give access to a Workspace. One must be explicitly invited to a workspace.
- 3 Each Workspace is isolated from all other Workspaces.
- 4 A professional penetration test is performed each year.
- 5 In the default Workspace, data can only be egressed after a request is granted.
- 6 The default Workspace has no internet access. It is therefore in no way possible to export information without authorization, for example via an e-mail to yourself.
- 7 Virtual Machines are used for viewing, editing and analyzing data.
- 8 Copy-paste between local machine and Virtual Machine is disabled. Also, access to USB sticks, printers and other peripherals other than sound is not possible within the Virtual Machine.
- 9 The data in a Workspace is stored in a storage account (Share) accessible to that Workspace with a 30-day rolling 24h snapshot that can be used self-service. All Virtual Machines of a Workspace have access to this Share.
- 10 If necessary, a mandated user can open one or more ports to fixed IP addresses, for example to retrieve R or Python packages or data from external data sources or to supply data to a registry.
- 11 A Workspace cannot be used for hosting because only inbound traffic is allowed.
- 12 Configuration changes to Workspaces performed by anDREa on the instructions of the customer are reviewed by anDREa and if there are no risks to the Core and other Workspaces, anDREa ensures that this change is sanctioned separately by the customer's Security Officer and PI/Department Head.
- 13 anDREa has 24x7x365 platform-wide monitoring that keeps an eye on suspicious behavior.
- 14 There is a separate OTAP street which makes it impossible to get unauthorized code into production.

To get a good idea of the conceptual framework that underlies the Digital Research Environment of anDREa, we refer to the Security Manifesto. All anDREa cybersecurity policies are on top of the strong cybersecurity measures that Microsoft Corporation provides for Microsoft Azure.



Interested in a discussion  
about the possibilities or a  
demonstration of anDREa?

Please feel free to contact:

---

**Elo Bosma**

Chief Executive Officer anDREa B.V.  
[elo.bosma@mydre.org](mailto:elo.bosma@mydre.org)

---

**Pascalie Broer**

Business Manager anDREa B.V.  
[pascalie.broer@mydre.org](mailto:pascalie.broer@mydre.org)

For help articles and the  
policies and procedure,  
please visit:

---

**anDREa's Knowledge Base**

<https://support.mydre.org/portal/en/kb>